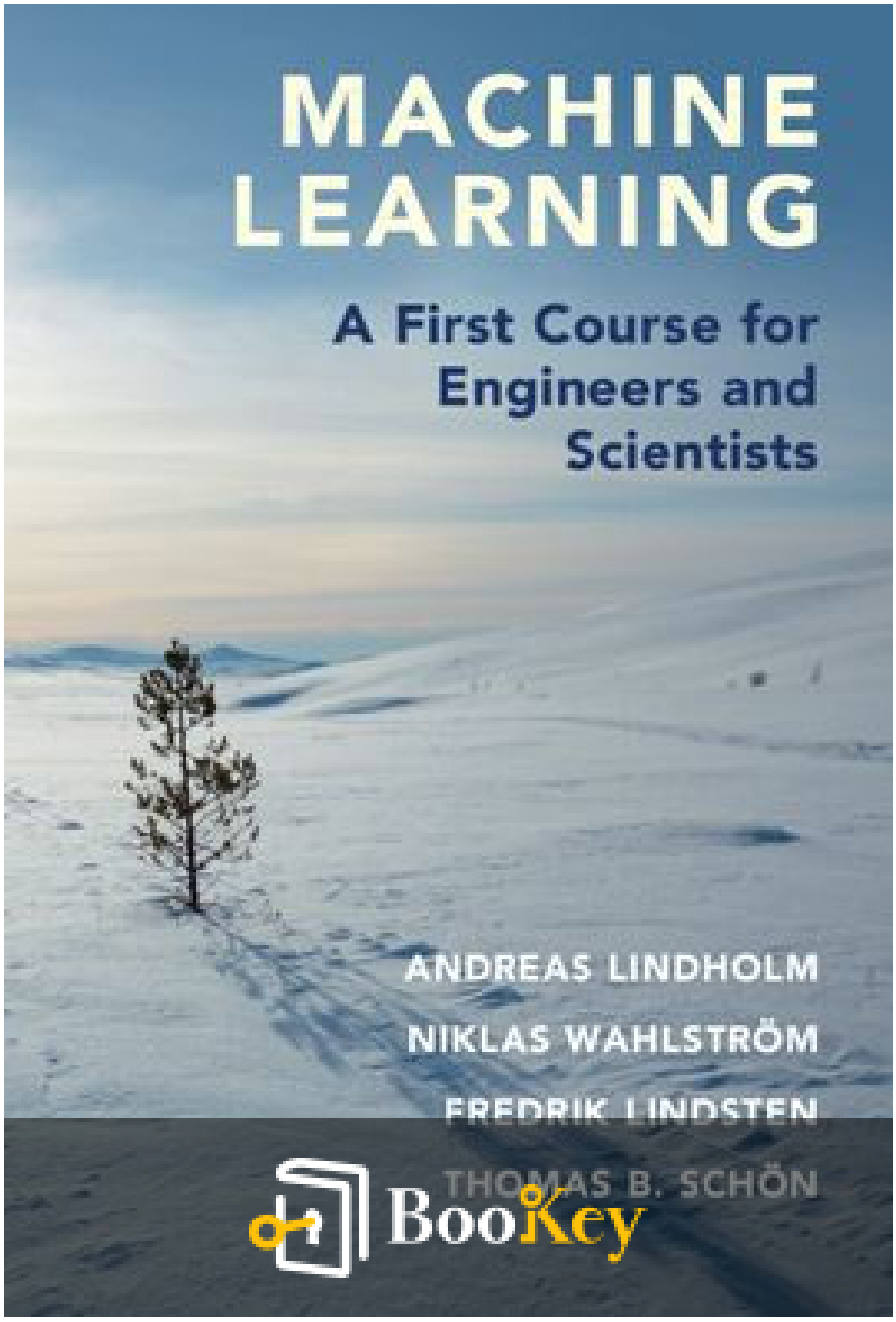
Apprentissage Automatique PDF (Copie limitée)

Andreas Lindholm



Apprentissage Automatique Résumé

Des analyses de données aux solutions prédictives

Écrit par Books1

À propos du livre

À une époque où les machines ne se contentent plus d’observer en silence, mais deviennent des acteurs à part entière façonnant les récits de notre monde, "Machine Learning" d'Andreas Lindholm constitue une porte d’entrée fascinante dans le domaine transformateur des ordinateurs qui apprennent. Lindholm démystifie les algorithmes complexes qui sous-tendent notre ère numérique, naviguant avec élégance entre théories fondamentales et pratiques de pointe. Avec une clarté rafraîchissante, ce livre réussit à allier un charme simple à une sophistication complexe, rendant l'ésotérique accessible tant aux novices qu'aux passionnés expérimentés. Il invite les lecteurs à franchir le voile conventionnel de la programmation, promettant un voyage à travers des paysages où les données deviennent à la fois la muse et le médium d'une innovation sans précédent. Que vous soyez sur le point de créer votre premier réseau de neurones ou désireux d'explorer plus profondément les profondeurs de l'intelligence machine, "Machine Learning" est à la fois une carte et une boussole — guidant, informant et inspirant à chaque étape.

À propos de l'auteur

Andreas Lindholm est un expert aguerri dans le domaine de l'apprentissage automatique, reconnu pour ses contributions novatrices et sa compréhension approfondie de l'intelligence artificielle. Avec plus de dix ans d'expérience, Lindholm a joué un rôle clé dans la fusion des concepts théoriques avec des applications pratiques, réalisant des avancées significatives dans l'amélioration des capacités technologiques. Il détient des diplômes avancés en informatique et en mathématiques appliquées, qui ont posé les fondements de sa passion pour l'apprentissage automatique et les solutions basées sur les données. En tant qu'auteur prolifique et conférencier recherché lors de conférences internationales, Lindholm a influencé à la fois le milieu académique et les praticiens de l'industrie, repoussant constamment les limites de ce qui est possible avec l'apprentissage automatique. Son style d'écriture accessible mais perspicace a fait de lui un auteur apprécié, aidant à démystifier des théories complexes pour les amateurs comme pour les professionnels. À travers son travail, Andreas Lindholm continue d'inspirer la prochaine vague d'innovation dans le monde de la technologie.

Essayez l'appli Bookey pour lire plus de 1000 résumés des meilleurs livres du monde

Débloquez 1000+ titres, 80+ sujets

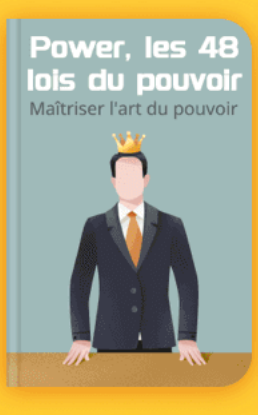
Nouveaux titres ajoutés chaque semaine

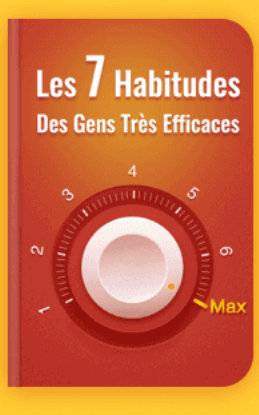
Brand | Leadership & collaboration | Gestion du temps | Relations & communication | Know

tégie d'entreprise | Créativité | Mémoires | Argent & investissements | Positive Psychology

Entrepreneuriat | Histoire du monde | Communication parent-enfant | Soins Personnels

Aperçus des meilleurs livres du monde



Essai gratuit avec Bookey

Liste de Contenu du Résumé

Chapitre 1 Résumé: Apprentissage supervisé : une première approche

Chapitre 2 : Apprentissage supervisé – Une première approche

Dans ce chapitre, le concept d'apprentissage supervisé en apprentissage automatique est présenté, ainsi que deux méthodes fondamentales : les k-Plus Proches Voisins (k-NN) et les arbres de décision. Ces méthodes sont intuitives et simples, mais elles constituent une base solide pour comprendre des techniques plus sophistiquées.

2.1 Apprentissage automatique supervisé

L'apprentissage supervisé traite des données qui comprennent des variables d'entrée (souvent appelées caractéristiques) et des variables de sortie correspondantes (étiquettes). L'objectif est d'apprendre à établir une correspondance entre l'entrée et la sortie à l'aide d'un modèle mathématique, qui pourra ensuite prédire la sortie pour de nouvelles données non observées. Les données d'entraînement consistent en plusieurs exemples indépendants, où chaque entrée est associée à une sortie connue.

La distinction entre les variables numériques et catégorielles est essentielle.

Les variables numériques ont un ordre inhérent (comme la vitesse ou la température), tandis que les variables catégorielles n'en ont pas (comme les couleurs ou les noms). Cette différence conduit à deux types de problèmes : la régression (output numérique) et la classification (output catégorielle).

Exemples pratiques :

1. **Classification de chansons** : Imaginez une application qui identifie si une chanson est des Beatles, de Kiss ou de Bob Dylan en fonction de caractéristiques audio. Ici, la sortie est catégorielle, ce qui en fait un problème de classification.

2. **Distances de freinage des voitures** : Étant donné des données sur les vitesses des voitures et les distances de freinage, le défi consiste à prédire les distances de freinage pour des vitesses non mesurées. Comme la sortie (distance) est numérique, il s'agit d'un problème de régression.

Les modèles d'apprentissage automatique visent à se généraliser au-delà des données d'entraînement, c'est-à-dire qu'ils doivent faire des prédictions fiables pour de nouvelles entrées. Un modèle qui s'ajuste parfaitement aux données d'entraînement mais qui échoue sur des données non vues est dit surajusté. La généralisation est essentielle pour l'application pratique d'un modèle.

2.2 Une méthode basée sur la distance : k-NN

La méthode k-NN prédit la sortie en se basant sur les exemples les plus proches des données d'entraînement, en utilisant la distance euclidienne entre les variables d'entrée. Pour la classification, elle utilise un vote majoritaire parmi les voisins les plus proches, tandis que pour la régression, elle utilise leur moyenne.

Le choix du bon nombre de voisins (k) implique un compromis entre flexibilité et précision. Trop peu de voisins peuvent entraîner un surajustement (capturant le bruit plutôt que le schéma sous-jacent), tandis que trop de voisins peuvent simplifier à outrance, manquant des schémas complexes.

Une étape importante consiste à normaliser les données d'entrée pour garantir que chaque variable contribue également au calcul de distance. La normalisation peut être réalisée en redimensionnant les entrées à une plage commune ou en les standardisant pour qu'elles aient une moyenne de zéro et un écart-type de un.

2.3 Une méthode basée sur des règles : Arbres de décision

Les arbres de décision classifient les données en posant une série de questions par oui ou par non, menant à un ensemble de règles qui segmentent les données en régions distinctes avec des prédictions uniformes. Structurellement, ce sont des arbres binaires avec des conditions (fentes) aux nœuds et des prédictions (nœuds feuilles) formant une surface de sortie constante par morceaux.

Pour la régression, les fentes minimisent l'erreur de prédiction, tandis que pour la classification, elles visent à augmenter la pureté au sein des nœuds en utilisant des mesures comme le taux de mauvaise classification, l'indice de Gini ou l'entropie. Ces mesures évaluent à quel point chaque fente sépare bien les classes.

Décider de la profondeur d'un arbre est crucial : une plus grande profondeur augmente la complexité et le risque de surajustement, tandis qu'une profondeur moins importante peut manquer des schémas. Des techniques comme la définition préalable d'une profondeur maximale ou l'élagage (suppression de parties de l'arbre) aident à contrôler la taille de l'arbre.

Résumé et lectures complémentaires

Ce chapitre présente les k-NN et les arbres de décision comme des méthodes

simples mais explicatives pour les tâches d'apprentissage supervisé. Bien qu'intuitives, elles servent de tremplin pour comprendre des modèles plus profonds et complexes. Parmi les lectures complémentaires, vous trouverez des travaux de Cover et Hart sur les k-NN et celui de Breiman et al. sur les arbres de décision, qui offrent une compréhension plus complète de ces méthodes fondamentales.

Chapitre 2 Résumé: Modèles paramétriques de base et une approche statistique de l'apprentissage

Résumé du Chapitre 3 : Modèles Paramétriques de Base et Perspective Statistique sur l'Apprentissage

Introduction à la Modélisation Paramétrique

Ce chapitre explore la modélisation paramétrique, une méthode utilisée dans l'apprentissage supervisé pour résoudre des problèmes en apprenant les paramètres du modèle à partir des données d'entraînement. L'accent est mis sur la régression linéaire et la régression logistique, deux exemples classiques de modèles paramétriques, où un ensemble de paramètres, noté Ø5ß , est appris et appliqué directement à de nouvelle modèle finalisé, sans avoir besoin des données d'entraînement.

3.1 Régression Linéaire

La régression, l'une des tâches fondamentales de l'apprentissage supervisé, consiste à prédire une sortie numérique Ø5Üf à partir régression linéaire suppose que cette sortie peut être modélisée comme une combinaison affines (essentiellement une combinaison linéaire plus une constante) des variables d'entrée, plus un terme de b

de l'erreur aléatoire. Le modèle est représenté par :

\[Ø5Üf = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ld
\epsilon \]

Cette section détaille comment les paramètres Ø5ß , y
l'origine Ø5ß €, sont déterminés à partir des données
biais de méthodes telles que l'approche des moindres carrés. Les "équations normales" constituent la base pour déterminer les valeurs des paramètres, et lorsque la matrice \(X^T X \) est inversible, une so
est disponible. Des illustrations, comme le modèle de distance d'arrêt d'une voiture, démontrent l'application de la régression linéaire.

Fonctions de Perte et Fonctions de Coût

L'optimisation de la régression linéaire implique généralement la minimisation d'une fonction de coût dérivée d'une fonction de perte, la perte d'erreur quadratique étant courante. Cela capture la différence entre les valeurs prédites et les valeurs réelles, motivant l'approche de coût des moindres carrés.

3.2 Classification et Régression Logistique

La classification se distingue de la régression par la prédiction de résultats

catégoriques. La régression logistique modifie le modèle de régression linéaire pour s'adapter à la classification en introduisant la fonction logistique, qui transforme les sorties linéaires en probabilités. Pour les problèmes de classification binaire, la régression logistique modélise la probabilité d'une classe en utilisant :

\[g(x) = \frac{e^{\theta^T x}}{1 + e^{\theta^T x}} \]

Les paramètres de la régression logistique sont appris par estimation du maximum de vraisemblance, plutôt qu'en utilisant les équations normales. Ce processus implique l'optimisation d'une fonction de coût basée sur la perte d'entropie croisée, particulièrement adaptée aux tâches de classification.

Prédictions et Frontières de Décision

Les modèles logistiques prédisent des catégories en sélectionnant la classe la plus probable, en utilisant communément un seuil (par exemple, 0,5 pour les résultats binaires) pour déterminer les frontières de décision, qui sont linéaires dans ce cas.

Extensions aux Situations Multiclasse

Pour la classification multiclasse, le modèle logistique est généralisé à l'aide

de la fonction softmax, permettant au modèle de produire une distribution de probabilité sur plusieurs classes. La fonction de coût employée est la perte d'entropie croisée multiclasse.

3.3 Régression Polynomiale et Régularisation

Pour améliorer la flexibilité du modèle, des transformations non linéaires telles que la régression polynomiale peuvent être ap
est remplacée par ses puissances supérieures (par exemple, \(x^2, x^3 \)). Cependant, cela augmente le risque de surapprentissage, où les modèles capturent le bruit plutôt que les motifs. Des techniques de régularisation comme la régularisation Ø5Ü?2 (Régression de Ridge) lutter contre le surapprentissage en pénalisant les grandes valeurs des paramètres dans la fonction de coût, exprimée comme suit :

\[\text{Fonction de coût avec régularisation} = \frac{1}{n} ||X\theta - y||^2_2 + \lambda ||\theta||^2_2 \]

3.4 Modèles Linéaires Généralisés

En généralisant davantage le concept, les modèles linéaires généralisés (GLM) adaptent les modèles linéaires à divers types de données de sortie en incorporant différentes fonctions de lien et distributions au-delà des domaines binaire et continu, tels que pour les données de comptage utilisant

la régression de Poisson, où la fonction de lien garantit la positivité du paramètre moyen, modélisé comme \(\lambda = \exp(\theta^T x) \).

Conclusion

Les modèles paramétriques offrent une approche structurée pour apprendre des relations à partir des données, permettant à la fois la prédiction et l'interprétation. Ils relient des relations linéaires simples et des distributions de données complexes, établissant une base pour des modélisations plus complexes abordées dans les chapitres suivants.

Chapitre 3 Résumé: Comprendre, évaluer et améliorer la performance

Ce chapitre se penche sur la compréhension, l’évaluation et l’amélioration des performances des modèles d’apprentissage automatique supervisé. Nous avons rencontré différentes méthodologies pour l'apprentissage automatique, avec pour concept central l’adaptation des modèles aux données d’entraînement dans l’espoir qu'ils puissent fournir des prédictions précises sur des données nouvelles et inconnues. Ici, nous allons explorer dans quelle mesure il est réaliste de s'attendre à ce que les modèles maintiennent leur performance lorsqu'ils sont appliqués à des ensembles de données non familiers, tout en offrant des perspectives sur les outils pratiques pour l’évaluation et l’amélioration des modèles.

Erreur des Nouvelles Données Attendues : Performance en Production

Ce chapitre introduit le concept de fonction d'erreur essentiel pour évaluer les tâches de classification ou de régression. Cette fonction mesure dans quelle mesure les prédictions (les résultats réels (Ø5Üf). Les métriques d'erreur cou de classification pour les tâches de classification et l'erreur au carré pour les tâches de régression. Il est à noter que la fonction d'erreur diffère de la fonction de perte utilisée lors de l’entraînement du modèle ; tandis que la fonction de perte optimise l’apprentissage du modèle, la fonction d'erreur

évalue la performance après l’entraînement.

L’apprentissage supervisé vise à affiner les modèles pour gérer efficacement des flux de données continus, nécessitant une attention portée à la performance sur les nouvelles données, décrite mathématiquement par la fonction d’erreur moyenne. Cela implique de considérer les entrées (x) et les sorties (y) de manière probabiliste, ce qui est une tâche délicate étant donné la complexité et la nature peu pratique de la définition de la distribution des données du monde réel, Ø5Ü](Ø5Üe, Ø5Üf).

Les prédictions des modèles basées sur les données d’entraînement (T) sont exprimées par Ø5ÜOØ5Üf(Ø5Üe; T). L'erreur des nouvell Ø5Ü8new, est dérivée en faisant une moyenne sur les inconnus possibles, mettant en lumière la capacité du modèle à généraliser à partir de T. De plus, le concept d'erreur d'entraînem performance du modèle exclusivement sur l'ensemble de données d'entraînement, une mesure généralement peu indicative de la performance sur des données futures.

Stratégies pour Estimer Ø5Ü8new

Estimer avec précision Ø5Ü8new est crucial pour éval Malgré l'impossibilité de calculer Ø5Ü8new directeme distributions de données inconnues, des estimations pratiques telles que

l'erreur de validation par séparation (Ø5Ü8hold-out) c

La validation par séparation implique de diviser les données disponibles en ensembles d'entraînement et de validation, une approche qui équilibre les besoins en données pour l'entraînement et l’évaluation.

La méthode de validation croisée en k plis affine davantage l’estimation de Ø5Ü8new en divisant les données en k blocs, alternant modèles sur tous les blocs sauf un et la validation sur le bloc exclu. Ce processus itératif exploite l'ensemble du jeu de données, produisant une estimation (Ø5Ü8k-fold) de Ø5Ü8new tout en permettan modèles et d'hyperparamètres.

Décomposition de l'Écart d'Erreur d'Entraînement–Généralisation

Décomposer Ø5Ü8new en erreur d'entraînement et écar élargit la compréhension des performances des modèles. Typiquement, la performance d'entraînement d'un modèle dépasse sa capacité à généraliser sur de nouvelles données, créant un écart de généralisation qui varie selon la complexité du modèle : les modèles plus complexes tendent à s’adapter de manière trop étroite aux détails spécifiques des données d’entraînement, risquant ainsi le surapprentissage. À l'inverse, le sous-apprentissage se produit lorsque les modèles manquent de complexité pour capturer efficacement les motifs des données. Cet équilibre nécessite d’atteindre un ajustement équilibré en optimisant la complexité du modèle pour minimiser

Ø5Ü8 n e w .

Décomposition Biais-Variance

La décomposition biais-variance offre une autre perspective pour analyser Ø5Ü8 n e w , e n l e s e g m e n t a n t e n b i a i s , v a r i a n c e e t b r u i t mesure les erreurs systématiques de prédiction, tandis que la variance indique la sensibilité des prédictions aux variations de données. Une flexibilité accrue du modèle réduit le biais mais peut amplifier la variance, mettant en avant l'importance d'établir un compromis optimal entre biais et v a r i a n c e . C e t é q u i l i b r e e s t c r u c i a l p o u r m i n i m i s e r Ø5 généralisation robuste du modèle.

Outils pour Évaluer les Classificateurs Binaires

Pour évaluer les classificateurs binaires de manière complète, ce chapitre présente des outils tels que la matrice de confusion et la courbe ROC, facilitant l'évaluation de performance au-delà des simples taux de classification erronée. Pour les problèmes déséquilibrés ou asymétriques—courants dans des domaines comme la détection de maladies—des métriques telles que la précision, le rappel et les scores F1 sont essentielles. L'AUC PR complète l'AUC ROC pour évaluer de manière exhaustive les compromis entre précision et rappel des classificateurs.

Conclusions

Ce chapitre souligne l'importance de méthodologies d'évaluation robustes dans l'apprentissage supervisé, essentielles pour développer des modèles qui performent de manière fiable sur des données inconnues. Grâce à une exploration détaillée des techniques d'estimation d'erreur, de l'analyse biais-variance et des outils de classification binaire, il permet aux praticiens d’acquérir une compréhension globale de la dynamique de performance des modèles, posant ainsi les bases pour des sujets plus avancés dans les chapitres suivants.

Section	Résumé
Introduction du Chapitre	Ce chapitre se concentre sur la compréhension et l'amélioration des performances des modèles d'apprentissage automatique supervisés, en explorant des outils d'évaluation et d'amélioration.
Erreur de Données Nouvelles Attendue : Performance en Production	Initie la fonction d'erreur pour évaluer les prédictions. Discute de l'importance de la performance sur de nouvelles données et fait la distinction entre les erreurs de formation et les erreurs sur de nouvelles données.
Stratégies pour Estimer Ø5Ü8new données, Ø5Ü8new, essentielle pour le perfectio	Décrit des méthodes telles que la validation par séparation et la validation croisée k-fold pour estimer l'erreur des nouvelles modèle.
Décomposition de l’Écart entre Erreur d’Entraînement et Généralisation	Explique l'écart de généralisation entre la performance sur les données d'entraînement et celles nouvelles, influencé par la complexité du modèle et le risque de surajustement ou de sous-ajustement.
Décomposition	Analyse les erreurs de prédiction en termes de biais, de variance

Section	Résumé
Biais-Variance	et de bruit. Souligne la nécessité d'un équilibre dans le compromis biais-variance pour minimiser Ø5Ü8 ne
Outils pour Évaluer les Classificateurs Binaires	Mette en avant des outils comme la matrice de confusion et la courbe ROC pour évaluer les classificateurs binaires, avec des métriques spécifiques pour des situations de déséquilibre des données.
Conclusions	Soulève l'importance de méthodologies d'évaluation robustes pour assurer des performances fiables des modèles sur des données non vues, en préparation pour des sujets avancés.

Pensée Critique

Point Clé: Erreur de données attendues : Performance en production

Interprétation Critique: Imaginez affronter chaque nouveau jour comme un modèle face à des données inexplorées. Ce chapitre vous invite à aller au-delà des succès initiaux et à prêter attention aux leçons de l'« erreur de nouvelles données attendues ». En avançant dans la vie, il ne s'agit pas seulement d'exceller dans des tâches familières, mais de se préparer à l'inattendu avec résilience. Cultivez un état d'esprit curieux qui embrasse l'incertitude et privilégiez les expériences qui élargissent vos horizons. N'oubliez pas, tout comme les modèles découvrant des données invisibles, vous aussi, vous pouvez prospérer en vous adaptant, en évaluant et en affinant continuellement votre compréhension. Laissez cette clé d'inspiration vous inciter à rechercher la croissance par l'exploration, en veillant à rester polyvalent et perspicace, tout comme un modèle vise à généraliser efficacement au-delà de ses conditions d'entraînement.

Chapitre 4: Apprendre les modèles paramétriques.

Chapitre 5 : Apprentissage des Modèles Paramétriques

Ce chapitre explore le concept de modélisation paramétrique, un élément fondamental de nombreuses méthodes d'apprentissage automatique supervisé. Il aborde trois thèmes centraux : les fonctions de perte, la régularisation et l'optimisation, des éléments incontournables lors du traitement de modèles paramétriques tels que la régression linéaire et la régression logistique. Bien que brièvement abordés dans le chapitre 3, ces concepts sont détaillés ici, soulignant leur applicabilité au-delà des modèles paramétriques.

5.1 Principes de la Modélisation Paramétrique

La modélisation paramétrique, introduite à travers la régression linéaire et logistique, peut être étendue pour inclure des modèles plus complexes. Le chapitre présente un cadre généralisé pour les modèles paramétriques, basé sur l'idée d'apprendre des fonctions à partir de données en utilisant des paramètres adaptables, notés par ¸. Il est souligné q
besoin d'être linéaires, ce qui permet d'explorer des relations non linéaires plus élaborées, comme le modèle cinétique de Michaelis-Menten pour les réactions enzymatiques, qui se caractérise par sa dépendance paramétrique

non linéaire.

Le chapitre examine ensuite comment ces modèles, qui supposent souvent un bruit gaussien additif, peuvent être adaptés grâce à différentes fonctions de perte, offrant ainsi des moyens flexibles pour décrire la relation entrée-sortie. Par exemple, passer de modèles linéaires à des fonctions non linéaires arbitraires permet de décrire les relations de manière plus précise. Les modèles non linéaires, comme la régression logistique pour la classification binaire, peuvent être étendus en modèles multi-classes et au-delà, en utilisant des fonctions softmax et des réseaux de neurones, qui seront abordés dans les chapitres suivants.

Minimisation de la Perte et Généralisation

L'apprentissage d'un modèle paramétrique consiste essentiellement à optimiser les paramètres du modèle de manière à minimiser l'erreur de prédiction ou la fonction de perte sur les données d'entraînement. Cependant, l'objectif véritable est la généralisation : obtenir un modèle qui prédit avec précision des données non vues, ce que l'objectif d'entraînement ne fait qu'approcher. Pour aligner ces objectifs, plusieurs stratégies sont adoptées, y compris la prise en compte de la précision statistique par rapport aux efforts computationnels, l'adoption d'une fonction de perte distincte de la fonction d'erreur pour une généralisation plus robuste, et l'utilisation de techniques telles que l'arrêt précoce et la régularisation explicite.

5.2 Fonctions de Perte et Modèles Basés sur la Probabilité

Le choix d'une fonction de perte appropriée est crucial, car il impacte le processus d'apprentissage du modèle et sa capacité de généralisation. Différentes fonctions de perte mettent en avant diverses caractéristiques dans les modèles. Pour la régression, les choix courants incluent la perte d'erreur quadratique et la perte d'erreur absolue, ayant chacune des implications différentes sur la robustesse du modèle, notamment en présence de valeurs aberrantes. Ce concept est étendu à la classification, où des fonctions telles que l'entropie croisée, la perte de hinge et la perte logistique définissent différentes approches pour modéliser les probabilités de classe, chacune offrant des avantages et des limites uniques.

L'adoption de modèles basés sur la probabilité permet d'intégrer des hypothèses probabilistes concernant les données dans les processus d'apprentissage. Cette approche aide non seulement à choisir des fonctions de perte, mais aussi à dévoiler les connexions mathématiques entre les distributions de données structurées et les paramètres du modèle, conduisant à des structures de modèles plus informées.

5.3 Régularisation

La régularisation est présentée comme une technique pour prévenir le

surapprentissage, surtout avec des modèles complexes. Les régularisations Œ2 (régression de ridge) et Œ1 (LASSO) sont mises en avant, introduisant des termes de pénalité pour maintenir des valeurs de paramètres faibles, sauf si les données indiquent fortement le contraire. La régularisation explicite implique souvent de modifier la fonction de coût, améliorant la simplicité du modèle, ce qui est généralement corrélé à une meilleure généralisation. Le chapitre aborde également des méthodes de régularisation implicite, comme l'arrêt précoce, qui se révèle efficace pour lutter contre le surapprentissage dans des contextes d'optimisation itérative.

5.4 Optimisation des Paramètres

Un aspect crucial de l'apprentissage des modèles paramétriques est de résoudre efficacement les problèmes d'optimisation. Divers algorithmes, tels que la descente de gradient, la méthode de Newton et la descente de coordonnées, sont explorés pour leur applicabilité à différents types de problèmes. Des techniques pour gérer le taux d'apprentissage et l'initialisation sont essentielles, surtout lorsqu'il s'agit de problèmes non convexes, qui nécessitent des méthodes comme la méthode de Newton en région de confiance pour des résultats robustes et convergents.

5.5 Optimisation avec des Grandes Bases de Données

Gérer de grandes bases de données nécessite des stratégies de calcul

efficaces, comme la descente de gradient stochastique (SGD), qui utilise de mini-lots pour approximer les gradients de manière économique. Cette approche accélère non seulement les calculs, mais introduit également des éléments stochastiques, qui, sous des conditions adéquates de décroissance du taux d'apprentissage (conformes aux conditions de Robbins-Monro),

Pourquoi Bookey est une application incontournable pour les amateurs de livres

Contenu de 30min
Plus notre interprétation est profonde et claire, mieux vous saisissez chaque titre.

Format texte et audio
Absorbez des connaissances même dans un temps fragmenté.

Quiz
Vérifiez si vous avez maîtrisé ce que vous venez d'apprendre.

Et plus
Plusieurs voix & polices, Carte mentale, Citations, Clips d'idées...

Essai gratuit avec Bookey

Chapitre 5 Résumé: Réseaux de neurones et apprentissage profond

Résumé du Chapitre : Réseaux de Neurones et Apprentissage Profond

Introduction aux Réseaux de Neurones et à l'Apprentissage Profond

- Les réseaux de neurones étendent la régression linéaire et logistique en empilant plusieurs modèles pour former des structures hiérarchiques, ce qui leur permet de modéliser des relations complexes entre les entrées et les sorties. L'apprentissage profond, une sous-catégorie de l'apprentissage automatique, se concentre sur ces modèles hiérarchiques.
- Ce chapitre introduit les réseaux de neurones en commençant par la généralisation de la régression linéaire à un réseau à deux couches, puis à des réseaux profonds. Il couvre également des réseaux spécifiques pour l'imagerie et des techniques d'entraînement.

6.1 Le Modèle de Réseau de Neurones

- **Régression Linéaire Généralisée :** Améliore la régression linéaire traditionnelle en utilisant des fonctions d'activation non linéaires (par exemple, logistique et ReLU) pour modéliser des relations complexes.

- **Réseau de Neurones à Deux Couches :** Augmente la complexité en utilisant plusieurs modèles de régression linéaire généralisée pour former des couches, agrégeant leurs sorties comme entrée à un autre modèle de régression.
- **Réseau de Neurones Profond :** Prolonge encore le modèle en empilant plusieurs couches, ce qui lui permet de traiter des tâches complexes comme la classification d'images.

6.2 Entraînement d'un Réseau de Neurones

- **Problème d'Optimisation :** Consiste à trouver les meilleurs paramètres via des fonctions de coût adaptées aux tâches de régression ou de classification (en utilisant des fonctions de perte comme l'erreur quadratique et l'entropie croisée).
- **Recherche Basée sur le Gradient :** Met à jour les paramètres de manière itérative par des techniques telles que la descente de gradient stochastique.
- **Algorithme de Rétropropagation :** Calcule efficacement les gradients nécessaires pour les mises à jour de paramètres, ce qui est crucial pour l'entraînement du réseau.

6.3 Réseaux de Neurones Convolutifs (CNN)

- Conçus pour des données d'entrée sous forme de grille, principalement des

images, les CNN utilisent des couches convolutives informées par la structure des images pour améliorer l'efficacité.

- **Concepts des Couches Convolutives :**

- **Interactions Éparses et Partage de Paramètres :** Réduisent la complexité du modèle en se concentrant sur de petites régions d'images et en réutilisant des paramètres à travers le réseau.
- **Pas et Pooling :** Des méthodes comme les convolutions avec pas et les couches de pooling réduisent les dimensions des données tout en conservant les caractéristiques importantes des images.
- **Élargissement des Canaux :** Plusieurs filtres représentent divers aspects des images, organisés sous forme de canaux, produisant des représentations de caractéristiques riches.

6.4 Dropout

- **Technique de Régularisation :** Entraîne un ensemble de sous-réseaux au sein d'un seul réseau en abandonnant aléatoirement des neurones et en créant des variations.
- **Entraînement et Prédiction :** Utilise les gradients des sous-réseaux pour l'entraînement, et lors de la prédiction, l'ensemble du réseau utilise des poids redimensionnés pour approximer les résultats moyens de tous les sous-réseaux.

6.5 Lectures Complémentaires

- Met l'accent sur le renouveau des réseaux de neurones grâce aux avancées matérielles, à la computation parallèle et au développement d'algorithmes, notamment dans les tâches de reconnaissance d'images.

Exemples d'Applications et Réflexions :

- Les exemples incluent la classification de chiffres utilisant des ensembles de données MNIST, illustrant les structures de réseau et les dynamiques d'entraînement.
- Les réflexions incitent à comprendre des concepts plus profonds tels que les implications de la structure du réseau et les effets des couches de pooling.

Ce chapitre fournit un guide complet pour comprendre le cadre des réseaux de neurones, les principes d'entraînement et les applications pratiques, posant les bases d'efforts avancés en apprentissage automatique.

Chapitre 6 Résumé: Méthodes d'ensemble : Bagging et boosting

Chapitre 7 : Méthodes d'Ensemble : Bagging et Boosting

Dans les chapitres précédents, nous avons abordé divers modèles d'apprentissage automatique tels que le k-NN et les réseaux de neurones profonds. Ce chapitre s’intéresse aux méthodes d’ensemble, une technique consistant à combiner plusieurs modèles de base pour améliorer la prédiction grâce à la « sagesse des foules ». En entraînant chaque modèle différemment, leurs prédictions collectives surpassent celles des modèles individuels. Les prédictions d’un ensemble sont obtenues par moyenne ou vote, ce qui entraîne une amélioration de la précision si elles sont bien structurées.

7.1 Bagging

Le bagging, abréviation de bootstrap aggregating, est une technique de réduction de variance. Il consiste à créer plusieurs variations de données d'entraînement via le bootstrap — un échantillonnage aléatoire de sous-ensembles qui se chevauchent — et à entraîner un modèle de base sur chacun. Cela réduit la variance sans augmenter le biais, diminuant ainsi le

risque de surajustement. Le bagging est particulièrement bénéfique pour des modèles comme les arbres de classification profonds ou le k-NN, qui ont un faible biais mais une forte variance. Par exemple, en utilisant des arbres de régression, le bagging maintient un faible biais tout en réduisant la variance, ce qui mène à des prédictions plus stables. La méthode du bootstrap implique un échantillonnage avec remise, permettant à certains points de données d'apparaître plusieurs fois, tandis que d'autres peuvent ne pas apparaître du tout. Cette technique aide à simuler plusieurs ensembles de données à partir d'un seul échantillon, ce qui est crucial lorsque la collecte de nouvelles données est impraticable.

L'un des principaux avantages du bagging est la réduction de variance grâce à la moyenne des prédictions de l'ensemble. En pratique, même avec de nombreux membres dans l'ensemble, le modèle ne souffre pas d'un biais ou d'un surajustement accru. Au contraire, la moyenne des prédictions aléatoires mais identiquement distribuées minimise efficacement la variance. Le nombre de membres de l'ensemble (\(B\)) se rapporte souvent à des contraintes de calcul, plutôt qu'à des considérations de biais ou de variance.

Les modèles de bagging utilisent l'estimation de l'erreur "hors sac" (« out-of-bag »), où environ 63 % des données participent à chaque ensemble de données bootstrap, pour estimer l'erreur sans nécessiter une validation croisée intensive.

7.2 Forêts Aléatoires

Les forêts aléatoires, une extension du bagging, ajoutent du hasard pour décorréler les modèles et réduire davantage la variance. Elles s'appliquent spécifiquement aux modèles basés sur des arbres. Chaque arbre est entraîné avec un sous-ensemble de variables d'entrée sélectionné au hasard, réduisant ainsi les corrélations entre les membres. Bien que cela augmente la variance individuelle des arbres, la prédiction moyenne de l'ensemble bénéficie de cette décorrélation, comme le montre les problèmes de classification binaire. Les forêts aléatoires évitent le surajustement même avec un grand nombre de membres, bien que les exigences en matière de calcul augmentent.

Le réglage du paramètre \(q\), le nombre d'entrées considérées pour les divisions, peut être déterminé par des règles empiriques ou des méthodes d'estimation d'erreur. Les forêts aléatoires conviennent aux mises en œuvre parallélisées, atténuant ainsi leurs coûts de calcul.

7.3 Boosting et AdaBoost

Le boosting, une autre méthode d'ensemble, vise à réduire le biais dans les modèles à fort biais en entraînant séquentiellement des modèles faibles, chacun corrigeant les erreurs de ses prédécesseurs. Contrairement au

bagging, le boosting crée de la flexibilité en transformant des modèles faibles en un ensemble puissant, adapté aux modèles à fort biais comme les arbres de décision peu profonds ou les stumps.

AdaBoost (Adaptive Boosting) est une mise en œuvre pratique du boosting, combinant plusieurs modèles faibles via un vote majoritaire pondéré, ce qui améliore la précision des prédictions collectives. L'entraînement séquentiel met l'accent sur les points de données difficiles, améliorant ainsi la classification. Comprendre AdaBoost implique de saisir l'ajustement stratégique des poids des modèles en fonction des performances de chaque itération, en mettant l'accent sur les données auparavant mal classées lors des tours suivants.

7.4 Boosting par Gradient

Une interprétation statistique du boosting le présente comme l'entraînement d'un modèle additif. Ce cadre remplace la perte exponentielle d’AdaBoost par des alternatives robustes comme la perte logistique, atténuant ainsi la sensibilité au bruit. Chaque étape de boosting consiste à sélectionner de nouveaux membres de l'ensemble en fonction des gradients des composants précédents afin d'optimiser le modèle complet de manière itérative. La différentiabilité d’ordre supérieur de la fonction de perte choisie est cruciale pour le calcul des gradients, ce qui est essentiel pour la méthodologie de

boosting par gradient.

Avec une vision plus large de l'apprentissage par ensemble, le boosting par gradient s'adresse à la fois aux tâches de classification et de régression en s'appuyant sur des modèles faibles — souvent basés sur des arbres — à travers des itérations répétitives et correctrices. Cette section se termine en présentant AdaBoost et le boosting par gradient appliqués à la classification musicale, démontrant ainsi des applications pratiques et des performances comparatives.

7.5 Lectures Complémentaires

Les lectures clés sur le bagging et les forêts aléatoires peuvent être retracées jusqu'aux travaux fondateurs de Breiman et Ho, tandis que l'évolution du boosting, popularisée par AdaBoost de Freund et Schapire, s'est élargie grâce à des contributions significatives, notamment le développement de cadres de boosting par gradient par Friedman. Les mises en œuvre modernes telles que XGBoost et LightGBM illustrent les avancées continues dans ce domaine.

Section	Résumé
Vue d'ensemble du chapitre	Les méthodes d'ensemble améliorent la précision des prédictions en combinant plusieurs modèles. L'averaging ou le vote des résultats conduit à de meilleures prédictions.

Section	Résumé
7.1 Bagging	- Réduit la variance sans augmenter le biais en utilisant des ensembles de données bootstrap. - Utile pour les modèles avec une forte variance (par exemple, les arbres profonds, le k-NN). - L'estimation hors sac permet d'approximativement évaluer l'erreur. - La taille de l'ensemble est limitée par des contraintes de calcul.
7.2 Forêts aléatoires	- Élargit le bagging en introduisant de la randomisation sur les caractéristiques d'entrée pour les modèles d'arbres. - Réduit la corrélation entre les arbres, diminuant ainsi la variance globale. - Exige des ressources de calcul importantes mais supporte le traitement parallèle. - Le paramètre déterminant de la division (q) influence les performances et peut être ajusté.
7.3 Boosting et AdaBoost	- Minimise le biais en entraînant successivement des modèles faibles qui corrigent les erreurs des prédécesseurs. - AdaBoost adapte les poids pour corriger les classifications incorrectes, améliorant ainsi la performance des modèles faibles.
7.4 Gradient Boosting	- Construit un modèle additif en utilisant l'optimisation par gradient. - Utilise des fonctions de perte robustes pour améliorer la stabilité. - Adapté aux tâches de classification et de régression. - Montre des applications comme la classification musicale.
7.5 Lectures complémentaires	- Cite des recherches fondamentales et des pionniers comme Breiman, Ho, Freund, Schapire et Friedman. - Met en avant des outils modernes, XGBoost et LightGBM, en

Section	Résumé
	tant qu'avancées.

Chapitre 7 Résumé: Transformations d'entrée non linéaires et noyaux

Chapitre 8 : Transformations d'entrées non linéaires et méthodes à noyau en apprentissage automatique

Ce chapitre s'intéresse aux transformations non linéaires des entrées et aux méthodes à noyau en apprentissage automatique, approfondissant les idées présentées dans le chapitre 3. Il traite principalement de la manière dont les transformations non linéaires des caractéristiques d'entrée et du truc du noyau peuvent améliorer les modèles traditionnels comme la régression linéaire et le k-NN, les rendant ainsi plus flexibles et puissants.

Concepts clés :

Transformations non linéaires des entrées: Traditionnellement, des modèles comme la régression linéaire se concentrent sur la modélisation des sorties en tant que combinaison linéaire des entrées. Cependant, en transformant les entrées de manière non linéaire (par exemple, en intégrant des caractéristiques polynomiales), nous pouvons rendre ces modèles plus adaptables à des structures de données complexes. Le chapitre explique qu'en transformant une seule entrée \(x \) à l'aide d'une approche polynomiale, on maintient le modèle comme une régression linéaire, les transformations non linéaires servant de nouvelles caractéristiques d'entrée.

Méthodes à noyau : Le truc du noyau permet d'implémenter des modèles en utilisant un nombre infini de caractéristiques sans les calculer explicitement. Les noyaux calculent les produits intérieurs dans un espace de haute dimension de manière implicite, permettant aux modèles de capturer des motifs complexes plus efficacement. Notamment, les méthodes à noyau sont introduites à travers les machines à vecteurs de support (SVM), qui utilisent la régression par vecteurs de support et la classification par vecteurs de support, se caractérisant par leur utilisation d'une fonction de perte de type « hinge » pour induire de la sparsité dans leurs solutions, se concentrant principalement sur les vecteurs de support — des points de données clés qui définissent la frontière de décision du modèle.

Composantes principales :

8.1 Création de caractéristiques par transformations d'entrées non linéaires : Cette section aborde les subtilités de ce qui constitue un modèle "linéaire" et les possibilités offertes par la redéfinition des entrées via des transformations non linéaires. L'exemple illustre comment l'augmentation de l'entrée avec des termes polynomiaux modifie la courbe de réponse du modèle sans perdre sa caractérisation en tant que méthode linéaire.

8.2 Régression en crête avec noyaux : Étend la régression linéaire par le biais du truc du noyau pour gérer une infinité de caractéristiques. La

régularisation devient cruciale pour prévenir le surapprentissage lorsque l'on ajoute de la complexité au modèle. Cette section décrit le processus d'utilisation de noyaux comme le polynomiale ou le gaussien, soulignant comment ils influencent la flexibilité du modèle.

8.3 Régression par vecteurs de support : Introduit une alternative de régression basée sur les noyaux qui utilise la fonction de perte insensible à \(\epsilon\), conduisant à des solutions éparses se concentrant uniquement sur des points de données critiques (vecteurs de support). Cette propriété permet d'économiser le calcul pendant les phases de prédiction, car seule une partie des données d'apprentissage influence les frontières de décision du modèle.

8.4 Théorie des noyaux : Fournit une assise théorique pour les méthodes à noyau, explorant le lien entre les noyaux et les mappings de caractéristiques, ainsi que le concept d'espace de Hilbert à noyau reproduisant. Elle offre également des conseils pratiques sur le choix des noyaux et discute de la formation de noyaux valides, en mettant l'accent sur les noyaux semi-défini positif en raison de leurs propriétés mathématiques garanties.

8.5 Classification par vecteurs de support : Semblable à la régression par vecteurs de support, mais appliquée aux problèmes de classification. Elle utilise une fonction de perte de type « hinge » pour garantir que seuls

les points de données proches de la marge de décision (vecteurs de support) influencent la frontière de classification. Cette section traite également de la formulation mathématique nécessaire à l'implémentation des classifieurs par vecteurs de support en utilisant différents types de noyaux, comme le linéaire ou l'exponentielle au carré.

Conclusion :

Le chapitre souligne que bien que les méthodes à noyau augmentent la flexibilité des modèles et capturent efficacement les relations non linéaires entre les données, elles nécessitent un choix minutieux des noyaux et un réglage des paramètres. Grâce aux noyaux, même des modèles simples comme la régression linéaire ou le k-NN peuvent s'adapter à des scénarios de données complexes du monde réel, faisant d'eux des outils puissants pour les ingénieurs et les scientifiques confrontés aux défis de l'apprentissage automatique.

Annexe - Théorème du représentateur et dérivation de la classification par vecteurs de support :

Le théorème du représentateur affirme que la solution des modèles basés sur les noyaux régularisés peut être exprimée en fonction des données d'apprentissage, soulignant l'élégance des formulations duales employées dans les méthodes par vecteurs de support. L'annexe aborde également la

dérivation mathématique nécessaire pour passer des formes primales aux formes duales, partie intégrante de l'implémentation des classifieurs par vecteurs de support dans la pratique.

Chapitre 8: L'approche bayésienne et les processus gaussiens.

Chapitre 9 : L’approche bayésienne et les processus gaussiens

Ce chapitre explore l'approche bayésienne en apprentissage automatique, en la contrastant avec la méthode paramétrique qui cherche une seule valeur de paramètre optimale pour ajuster des modèles aux données. Ici, l'apprentissage est redéfini comme la détermination d'une distribution de valeurs de paramètres possibles conditionnées par les données observées. Cette perspective probabiliste nous permet de prédire non seulement un résultat unique, mais également une gamme de résultats possibles, chacun avec sa probabilité. Le cadre bayésien est distinct tant sur le plan théorique que philosophique, offrant une méthodologie robuste applicable à un large éventail de scénarios pratiques dans l'apprentissage supervisé.

9.1 L'idée bayésienne

Dans l'approche bayésienne, les paramètres du modèle sont considérés comme des variables aléatoires. L'accent est mis sur la formation d'une distribution conjointe sur toutes les sorties et tous l X). En pratique, cela déplace l'emphase de l'estimation d'un paramètre à l'évaluation de la distribution postérieure des param

à l'aide de lois de probabilité comme le théorème de Bayes. Ce théorème, Ø5Ü](Ø5ß= | y) = Ø5Ü](y | Ø5ß=)Ø5Ü](Ø5ß=) / Ø5Ü](y), compr
- **Vraisemblance Ø5Ü](y | Ø5ß=)** : Distribution des
paramètres.
- **Prior Ø5Ü](Ø5ß=)** : Distribution des paramètres a
données.
- **Postérieur Ø5Ü](Ø5ß= | y)** : Distribution des para
l'observation des données.

Les distributions de probabilité dans ce contexte représentent des croyances sur les paramètres ou des prévisions sous incertitude. Les prédictions d'un modèle bayésien, sa robustesse face au surajustement, notamment avec peu de données, et la mise à jour séquentielle des croyances avec de nouvelles données rendent cette approche particulièrement attrayante.

9.2 Régression linéaire bayésienne

Le cadre bayésien appliqué à la régression linéaire illustre le passage de l'estimation ponctuelle à l'inférence probabiliste. Une distribution gaussienne multivariée joue un rôle crucial. La régression linéaire bayésienne implique :
- Définir des distributions a priori pour les paramètres du modèle.
- Utiliser les données observées pour mettre à jour l'a priori en une distribution postérieure Ø5Ü](Ø5ß= | y).
- Dériver des prédictions à travers des distributions prédictives postérieures.

Ici, les concepts de régularisation s'alignent naturellement avec les principes bayésiens, offrant des perspectives sur leur utilité pour atténuer le surajustement.

9.3 Le processus gaussien

Le Processus Gaussien (PG) étend l'approche bayésienne à des modèles non paramétriques, considérant les fonctions entières comme des entités stochastiques pour inférer des prédictions. Le PG est un processus stochastique continu vu à travers le prisme d'une distribution gaussienne multivariée pour tout ensemble fini d'entrées. Sa fonction de covariance (noyau) joue un rôle central dans la définition des corrélations entre les valeurs fonctionnelles. Le PG permet la régression avec quantification de l'incertitude pour les prédictions, affichant une flexibilité à travers des noyaux qui spécifient des traits inhérents à la fonction, comme la douceur.

9.4 Aspects pratiques du processus gaussien

Le choix du bon noyau et l'ajustement des hyperparamètres (avec des méthodes comme la maximisation de la vraisemblance marginale) sont cruciaux pour mettre en œuvre les processus gaussiens de manière efficace. Ces choix ont un impact significatif sur la performance prédictive et l'efficacité computationnelle.

9.5 Autres méthodes bayésiennes en apprentissage automatique

Le paradigme bayésien s'étend bien au-delà de la régression, englobant une variété de modèles nécessitant des techniques de calcul complexes comme

App Store
Coup de cœur

22k avis 5 étoiles

Retour Positif

Fabienne Moreau

ue résumé de livre ne testent
ion, mais rendent également
musant et engageant.
té la lecture pour moi.

Fantastique!

Giselle Dubois

Je suis émerveillé par la variété de livres et de langues que Bookey supporte. Ce n'est pas juste une application, c'est une porte d'accès au savoir mondial. De plus, gagner des points pour la charité est un grand plus !

Fi
Le
liv
co
pr

é Blanchet

de lecture
eption de
es,
tous.

J'adore !

Isoline Mercier

Bookey m'offre le temps de parcourir les parties importantes d'un livre. Cela me donne aussi une idée suffisante pour savoir si je devrais acheter ou non la version complète du livre ! C'est facile à utiliser !"

Gain de temps !

Bookey est mon applicat
intellectuelle. Les résum
magnifiquement organis
monde de connaissance

ppli géniale !

Joachim Lefevre

adore les livres audio mais je n'ai pas toujours le temps
'écouter le livre entier ! Bookey me permet d'obtenir
n résumé des points forts du livre qui m'intéresse !!!
Quel super concept !!! Hautement recommandé !

Appli magnifique

Cette application est une bouée de sauve
amateurs de livres avec des emplois du te
Les résumés sont précis, et les cartes mer
renforcer ce que j'ai appris. Hautement re

Essai gratuit avec Bookey

Chapitre 9 Résumé: Modèles génératifs et apprentissage à partir de données non étiquetées.

Chapitre 10 : Modèles génératifs et apprentissage à partir de données non étiquetées

Aperçu

Ce chapitre fait la transition entre les modèles discriminants, qui prédisent des résultats basés sur des entrées données, et les modèles génératifs. Contrairement aux modèles discriminants qui ne décrivent que la distribution conditionnelle d’une sortie donné une entrée, les modèles génératifs décrivent la distribution conjointe des entrées et des sorties. Cela offre une compréhension plus riche des données et permet de générer des données synthétiques ou d'identifier des motifs dans les données sans avoir besoin de sorties étiquetées.

Introduction aux Modèles Génératifs

Les modèles génératifs décrivent comment les données d’entrée et de sortie sont générées conjointement. Un modèle notable est le Modèle de Mélange Gaussien (GMM), qui représente les données comme des mélanges de distributions gaussiennes, et qui peut être utilisé pour l’analyse discriminante ainsi que pour l’apprentissage semi-supervisé et non supervisé, comme le clustering.

Modèle de Mélange Gaussien (GMM) et Analyse Discriminante

- **Les Bases du GMM** : Un GMM décrit un modèle de probabilité où les variables d'entrée suivent une distribution gaussienne au sein de chaque classe. Il utilise la probabilité conjointe Ø5Ü](x, y) = une catégorie avec une distribution gaussienne de x pour chaque classe.
- **Apprentissage des Paramètres** : Dans un scénario d'apprentissage supervisé, les paramètres du GMM sont appris en maximisant la log-vraisemblance des données observées. Cela donne lieu à des méthodes telles que l'Analyse Discriminante Linéaire (LDA) et l'Analyse Discriminante Quadratique (QDA) pour la classification.
- **Applications Semi-supervisées et de Clustering** : Les GMM permettent de regrouper des données non étiquetées et d'apprendre à partir de données partiellement étiquetées. L'imputation des étiquettes et l'algorithme d'Expectation-Maximization (EM) jouent des rôles cruciaux pour gérer les défis semi-supervisés en mettant à jour itérativement les paramètres du modèle à l'aide des prédictions sur les données non étiquetées.

Apprentissage Non Supervisé et Analyse de Cluster

- **Prise en Compte des Scénarios Non Supervisés** : Les GMM sont adaptés à l'apprentissage non supervisé en considérant que les données contiennent des variables de cluster latentes. Cela implique d'apprendre un modèle conjoint Ø5Ü](x, y) uniquement à partir des en clusters de données inhérents.

- **Aperçu de l’Algorithme EM** : L’algorithme d’Expectation-Maximization (EM) est utilisé pour affiner itérativement les paramètres du modèle, en alternant entre l’estimation des variables cachées et l'optimisation des paramètres.
- **Clustering par Ø5ÜX-means** : Comme méthode a les Ø5ÜX-means partitionnent durement les données e la variance intra-cluster. Cela offre un cadre plus simple mais moins flexible comparé au clustering probabiliste des GMM.

Modèles Génératifs Avancés
- **Limitations et Extensions au-delà de l’hypothèse Gaussienne** : Bien que les GMM supposent des distributions gaussiennes, les modèles génératifs profonds offrent une plus grande flexibilité en utilisant des transformations complexes de distributions plus simples, notamment à l’aide de réseaux de neurones.
- **Flux Normalisés** : Cette technique modélise les données comme des transformations de distributions plus simples et implique des mappings réversibles avec des Jacobiennes mathématiquement réalisables pour l’apprentissage des paramètres.
- **Réseaux Antagonistes Génératifs (GAN)** : Les GAN apprennent sans vraisemblances explicites en générant des données via un entraînement antagoniste, où un réseau générateur crée des données et un réseau discriminant évalue leur authenticité.

Apprentissage de Représentation et Réduction de Dimensions

- **Auto-Encodeurs** : Ceux-ci apprennent des représentations de faible dimension en imposant un goulot d'étranglement qui comprime les données d'entrée, reconstruisant les entrées originales aussi fidèlement que possible.
- **Analyse en Composantes Principales (PCA)** : La PCA, considérée comme une solution d'auto-encodeur linéaire, trouve une représentation de dimension réduite grâce à des transformations orthogonales qui maximisent la variance des données.

Remarques Finales

Les modèles génératifs offrent des cadres riches pour comprendre et générer synthétiquement des données. Bien qu'ils impliquent plus d'assomptions que les modèles discriminants, ils ouvrent des possibilités dans des contextes où l'étiquetage est coûteux ou impraticable, comblant le fossé entre les paradigmes d'apprentissage supervisé et non supervisé.

Lectures Recommandées

Pour des discussions plus détaillées, consultez les travaux de Bishop (2006), Hastie et al. (2009), et Goodfellow et al. (2016), ainsi que les travaux fondamentaux sur les GAN et les flux normalisés (Goodfellow et al., 2014 ; Kobyzev et al., 2020). Pour des applications et des extensions, consultez la littérature sur les auto-encodeurs variationnels et les méthodes d'apprentissage semi-supervisé.

Chapitre 10 Résumé: Aspects utilisateur de l'apprentissage automatique

Chapitre 11 : L'aspect pratique de l'apprentissage automatique

Ce chapitre se concentre sur les aspects pratiques de l'apprentissage automatique, en particulier dans le contexte de l'apprentissage supervisé. Ce résumé présente les points clés pour développer et affiner des modèles d'apprentissage automatique tout en veillant à une utilisation optimale des ressources limitées, comme les données et le temps.

Points forts du chapitre :

1. **Définir le problème de l'apprentissage automatique :**

- Le développement de l'apprentissage automatique est un processus itératif : entraîner, évaluer et améliorer. L'objectif est d'évaluer les améliorations de manière efficace, sans nécessiter de déploiements fréquents dans des environnements de production.
- La stratégie consiste à utiliser une répartition des données bien définie : des données d'entraînement pour apprendre le modèle, des données de validation pour ajuster les hyperparamètres, et des données de test pour

l'évaluation finale.

- En cas de données limitées, des techniques comme la validation croisée en k-fold peuvent compléter les méthodes de validation par séparation. Il est crucial de s'assurer que les ensembles de validation et de test représentent la distribution attendue en production pour éviter de « tirer à côté de la cible ».

- Le risque de surajustement (overfitting) aux données de validation peut être atténué en élargissant la taille de l'ensemble de validation et en partitionnant soigneusement pour éviter les fuites de groupe, notamment dans des domaines tels que l'imagerie médicale.

2. **Améliorer un modèle d'apprentissage automatique :**

- Le processus d'amélioration des modèles est itératif et peut impliquer l'essai de différents modèles ou le réglage des hyperparamètres.

- Il est recommandé de commencer par des approches simples et de progresser vers des modèles plus complexes. Le débogage et la vérification de la correction du modèle doivent précéder les améliorations élaborées.

- Le chapitre met l'accent sur le maintien d'un équilibre entre la minimisation de l'erreur de formation et la réduction de l'écart de généralisation (la différence entre la performance en formation et en validation).

- L'évaluation des courbes d'apprentissage et l'utilisation de l'analyse des erreurs aident à prioriser les efforts de collecte de données et à identifier les domaines d’amélioration potentielle du modèle.

3. **Stratégies lorsque les données sont limitées :**

- Dans les situations avec peu de données, plusieurs stratégies peuvent être utiles :
  - Augmenter les données d'entraînement par de légères modifications ou tirer parti de jeux de données non liés, mais partageant des similarités.
  - L'utilisation de l'apprentissage par transfert pour transférer des connaissances de modèles formés sur de grands ensembles de données vers de nouvelles tâches corrélées est soulignée.
  - Recourir à l'apprentissage auto-supervisé ou semi-supervisé pour tirer des enseignements d'ensembles de données non étiquetées.

4. **Aborder les problèmes pratiques de données :**

- Les défis pratiques tels que les valeurs aberrantes, les données manquantes et les caractéristiques redondantes doivent être traités pour garantir la robustesse du modèle.
- Les valeurs aberrantes nécessitent une attention particulière—il faut décider si elles constituent du bruit ou un phénomène d'intérêt.

5. **Fiabilité des modèles d'apprentissage automatique :**

- Comprendre et faire confiance aux modèles d'apprentissage supervisé

peut être difficile, surtout pour des modèles complexes comme le deep learning.

- Bien que les modèles plus simples offrent des avantages en matière d'interprétabilité, ils peuvent manquer de la puissance prédictive de systèmes plus complexes. Des efforts sont en cours dans le domaine pour améliorer l'interprétabilité et la robustesse face aux exemples adversariaux et aux scénarios les plus défavorables.

6. **Explorations supplémentaires :**

- Le chapitre fait référence aux ouvrages d'Ng (2019) et de Burkov (2020) pour des aperçus supplémentaires sur les aspects utilisateurs de l'apprentissage automatique, ainsi qu'à la littérature sur les techniques d'augmentation de données et d'interprétation des modèles.

Le chapitre souligne une approche structurée et éclairée dans le développement de solutions d'apprentissage automatique, en se concentrant sur les considérations pratiques et l'amélioration itérative pour résoudre efficacement les défis du monde réel.

Pensée Critique

Point Clé: Apprentissage par Transfert

Interprétation Critique: Imaginez exploiter la puissance d'innombrables informations disponibles grâce à des modèles entraînés dans d'autres domaines et les appliquer à vos besoins spécifiques. L'apprentissage par transfert vous permet de le faire. Lorsque vous tirez parti de modèles préexistants, entraînés sur de grands ensembles de données variés, vous ouvrez la porte à des possibilités infinies même lorsque vous partez avec des données limitées. Pensez-y comme à se tenir sur les épaules de géants, utilisant la sagesse collective intégrée dans des modèles sophistiqués pour résoudre vos problèmes uniques. Cela dépasse le simple raccourci technique ; c'est un changement de paradigme qui accélère votre croissance et vous donne du pouvoir même dans des environnements contraints en ressources. Adoptez cette stratégie pour débloquer un potentiel que vous ne soupçonniez pas, atteignant vos objectifs avec créativité et innovation.

Chapitre 11 Résumé: L'éthique dans l'apprentissage automatique

Résumé du Chapitre 12 : Éthique en Apprentissage Automatique

Dans le Chapitre 12 de "L'Apprentissage Automatique : Un Premier Cours pour Ingénieurs et Scientifiques" par David Sumpter, l'accent est mis sur les défis éthiques qui surgissent dans les applications de l'apprentissage automatique. Ces problèmes découlent souvent de choix de conception apparemment neutres qui entraînent des conséquences inattendues pour les utilisateurs et la société. Le chapitre prône une approche éthique basée sur la prise de conscience, soulignant l'importance de comprendre ces impacts éthiques plutôt que de s'appuyer uniquement sur des solutions techniques.

12.1 Équité et Fonctions d'Erreur

La première section aborde l'équité en apprentissage automatique, en commençant par le choix d'une fonction d'erreur. Sumpter illustre cela par un exemple fictif d'un modèle de recommandation de cours universitaires pour des candidats suédois et non suédois. Malgré des performances globales égales, les taux de faux négatifs révèlent une disparité, les non-Suédois intéressés étant plus susceptibles d'être incorrectement recommandés que les Suédois. Cela met en évidence que l'équité dépend du

contexte : ce qui est juste dans une situation peut être injuste dans une autre. L'algorithme Compas, utilisé pour les condamnations criminelles, illustre l'impossibilité mathématique d'atteindre l'équité selon tous les critères souhaitables simultanément. Cette section souligne que l'équité est subjective et requiert une prise de conscience concernant les décisions de conception et leurs implications.

12.2 Allégations Trompeuses sur la Performance

La croissance rapide de l'apprentissage automatique a conduit à des affirmations exagérées de performance, influencées par des intérêts commerciaux. Par exemple, bien que les réalisations de Google DeepMind en apprentissage par renforcement dans les jeux aient été saluées comme des jalons, leurs capacités ont parfois été surestimées. Un exemple marquant est l'algorithme Compas, utilisé dans les condamnations criminelles. Bien qu'il ait montré de hautes performances sur des critères techniques comme l'AUC, son efficacité dans le monde réel était comparable à celle de prédictions humaines à la légère et n'apportait que peu de nouveauté à la connaissance existante sur la prédiction criminelle.

Le chapitre critique également les affirmations exagérées d'Elon Musk concernant les capacités de l'IA de Tesla, ainsi que l'exactitude des modèles d'apprentissage automatique utilisés dans la prédiction de la personnalité par Cambridge Analytica. Les dangers des affirmations hyperboliques sont

exacerbés par un jargon complexe, qui peut masquer les implications réelles aux yeux des non-experts. Le chapitre encourage à utiliser un langage accessible pour transmettre les performances et les limites des modèles, favorisant une communication honnête avec les parties prenantes.

12.3 Limitations des Données d'Entraînement

Sumpter introduit le concept des modèles d'apprentissage automatique comme des "parrots stochastiques", reflétant l'idée que ces modèles ne font que répéter des motifs observés dans les données d'entraînement sans réelle compréhension. Cette analogie aide à expliquer pourquoi les modèles échouent face à des variations non présentes dans leurs données d'entraînement. Le chapitre discute des biais inhérents aux données d'entraînement, tels que les disparités raciales et de genre dans les systèmes de reconnaissance faciale, qui perpétuent les biais sociaux existants.

Lorsqu'ils sont appliqués à de grands ensembles de données non vérifiés, les modèles linguistiques peuvent par inadvertance produire des résultats biaisés ou incorrects, comme répéter des théories du complot. Cela souligne la nécessité de disposer de jeux de données collectées de manière éthique et équilibrée, et met en lumière les risques liés à l'hypothèse que les modèles comprennent les données sur lesquelles ils sont formés. Sumpter insiste sur l'importance de reconnaître l'influence des facteurs historiques et sociaux sur les données et souligne que la neutralité des modèles est un mythe.

Conclusion

Dans l'ensemble, le Chapitre 12 du livre de Sumpter souligne l'importance de la prise de conscience pour aborder les défis éthiques en apprentissage automatique. Bien que être conscient de ces problèmes ne les résolve pas, c'est un point de départ crucial. Le chapitre encourage les ingénieurs et les scientifiques à communiquer de manière claire, à reconnaître les biais et à envisager activement les implications éthiques dans leur travail. Des lectures supplémentaires, y compris celles de Bender et al., David Sumpter, et Cathy O'Neill, sont recommandées pour des insights plus approfondis sur ces questions complexes.

Pensée Critique

Point Clé: Approche Éthique par la Conscience

Interprétation Critique: Imaginez tenir entre vos mains le pouvoir de l'apprentissage automatique, un outil capable d'avoir un impact monumental, mais rempli de dilemmes éthiques cachés. Adoptez l'approche éthique par la conscience pour naviguer dans ces eaux troubles. Ce chapitre souligne l'importance de comprendre les répercussions éthiques que vos choix de conception entraînent dans la société. Au lieu de faire aveuglément confiance à la façade de la neutralité technologique, reconnaissez que l'intention, la conception et l'application des algorithmes d'apprentissage automatique nécessitent une évaluation consciencieuse de l'équité et de l'impact. En vous engageant dans ce cheminement de conscience, vous découvrez la nature subjective de l'équité, garantissant que vos contributions visent non seulement un succès technique immédiat, mais résonnent également avec une harmonie éthique plus large. Ce principe pourrait révolutionner votre perception de la responsabilité, favorisant un monde où l'innovation est guidée autant par l'empathie que par la compréhension.