

CM4 : La Regression

But de la régression : on veut prédire une valeur numérique au lieu d'une étiquette (toujours supervisée).

Régression quantile (Quantile Regression):

La RQ est plus précise que la régression des moindres carrés, car lieu d'avoir une moyenne on a la moyenne sur un quantile, il permet de bien étudier les extrêmes.

La régression linéaire :

PRINCIPE :

Le but principal de la régression linéaire en machine learning est de **trouver la meilleure ligne droite** (ou hyperplan dans le cas de plusieurs variables) qui représente la relation entre la variable cible et les caractéristiques. Cela est fait en minimisant l'erreur entre les valeurs prédites et les valeurs réelles, généralement à travers une méthode appelée **moindres carrés** (Ordinary Least Squares, OLS).

Utilisation dans le domaine de machine Learning

1. **Prédiction de prix** : Par exemple, la régression linéaire peut être utilisée pour prédire les prix de l'immobilier en fonction de caractéristiques telles que la surface, le nombre de chambres, et l'emplacement.
2. **Analyse de tendances** : Elle peut aussi être employée pour analyser les tendances de vente dans une entreprise en fonction de divers facteurs comme les investissements en marketing.
3. **Modèles prédictifs** : Dans des cas plus avancés, elle peut être utilisée pour prédire des comportements financiers ou économiques basés sur des variables historiques

Exemple pratique

Supposons que vous avez un ensemble de données sur les prix des maisons avec des informations sur la surface, le nombre de chambres, et l'âge de la maison. La régression linéaire permettrait de modéliser la relation entre ces caractéristiques et le prix des maisons. Ainsi, une fois le modèle entraîné, vous pourriez entrer les caractéristiques d'une nouvelle maison et obtenir une estimation du prix.

Limites

- **Relation linéaire** : La régression linéaire fonctionne mieux lorsque les relations entre les variables sont linéaires. Si les relations sont non linéaires, des techniques comme la régression polynomiale ou d'autres algorithmes plus avancés peuvent être nécessaires.
- **Sensibilité aux outliers** : La régression linéaire est sensible aux points aberrants (outliers), qui peuvent fausser les prédictions.

Moindre Carré Non Négatif :

$$\arg \min_{\mathbf{x} \geq 0} \left(\frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{c}^T \mathbf{x} \right),$$

Kernel Ridge Regression (KRR):

Le **Kernel Ridge Regression** (KRR) combine deux concepts principaux : la **régression ridge** et les **méthodes à noyau**.

1. **Régression ridge** : C'est une forme de régression linéaire régularisée qui ajoute une pénalité sur les coefficients du modèle (proportionnelle à leur norme au carré). Cela permet de contrôler le surapprentissage (overfitting) en favorisant des solutions avec des coefficients plus petits.
2. **Méthodes à noyau** : Elles permettent d'étendre les modèles linéaires à des modèles non linéaires en projetant les données d'origine dans un espace de caractéristiques de dimension supérieure (sans avoir à calculer explicitement cette projection). Le noyau (kernel) mesure la similarité entre les points dans cet espace transformé.

En combinant ces deux techniques, le KRR peut effectuer une **régression non linéaire régularisée**, en exploitant la puissance des noyaux pour capturer des relations complexes dans les données tout en contrôlant la complexité du modèle.

[PDF contenant la formule](#) pour le Kernel Ridge Regression.