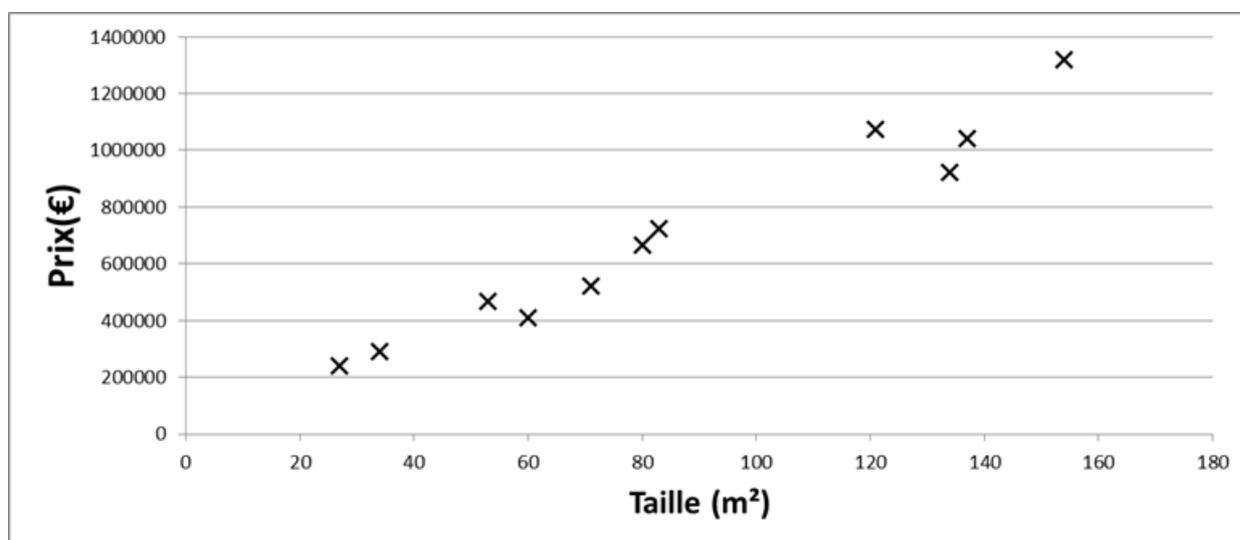


Traitement de données appliqué à la finance TD7 - Analyse de données - régression linéaire

1 Régression linéaire

Considérez le graphique suivant réalisé à partir des données d'une agence immobilière : l'abscisse représente la surface d'un appartement, et l'ordonnée représente le prix de vente. Chaque point représente une vente réalisée dans l'agence.



1. Une personne souhaite vendre son bien d'une surface totale de $120m^2$. Quel prix de vente environ pourrait-elle le vendre ? **Attention** : écrivez la réponse sur votre feuille, sans en parler avec votre voisin, ni même sans lui montrer.
2. Expliquez la façon dont vous avez procédé pour obtenir cette évaluation. Est-ce que cette méthode est sensée donner exactement la même réponse d'un individu à l'autre ? Est-ce une méthode satisfaisante ?
3. Comment peut-on **rigoureusement** définir la meilleure droite qui passe par un ensemble de points non alignés ?
4. Soit un point $P(P_x, P_y)$ et une droite $d : y = ax + b$: quelle est la distance verticale au carré entre la droite d et le point p ?
5. Soit un ensemble de points $E = \{P1, P2, \dots, PN\}$: exprimez la somme des distances verticales au carré de la droite à l'ensemble de points. Une "bonne droite" doit-elle maximiser ou minimiser cette distance ?
Pouvez-vous maintenant définir **rigoureusement** la "meilleure" droite qui passe par un ensemble de points non alignés ? Cette définition repose-t-elle sur un choix personnel de votre part, ou est-elle "universelle" ?
6. Nous appellerons *Score* la fonction qui prend en entrée une droite et produit en sortie la somme des distances verticales au carré de chaque point de E à la droite. Quels sont les paramètres de la fonction *Score* ? Comment exprimer la meilleure droite qui passe par l'ensemble de points E à l'aide de *Score* ?

7. Étant donné une fonction quelconque, comment trouve-t-on en général un extremum local (minimum ou maximum local) de la fonction ?

Que signifie, par rapport à notre problème, un extremum local de la fonction de Score ? Serait-ce plutôt un maximum ou un minimum ?

Calculez l'extremum de la fonction Score. Qu'obtenez-vous ?

8. Pouvez-vous expliquer pourquoi nous avons utilisé des distances au carré et non pas des distances tout court ?