  Énoncés des exercices

#### Contexte

Reprenons notre problématique des loyers abordée dans ce cours.

La question qu'on essaie de résoudre est :

**Étant donné les caractéristiques de mon appartement, combien devrais-je normalement payer mon loyer ?**

Imaginons pour l'instant que la seule caractéristique dont nous disposons est la surface de l'appartement. Notre training set est un ensemble de N = 545 observations de surface et des loyers associés :

**(x,y) = (surface,loyer) .**

1. Commençons par charger et afficher les données d'entraînement
2. D'après la visualisation, quelle hypothèse de modélisation peut-on émettre.

La régression linéaire s’appuie sur l’hypothèse que les données proviennent d’un phénomène qui a la forme d'une droite, c’est-à-dire qu’il existe une relation linéaire entre l’entrée (les observations) et la sortie (les prédictions).

Nous avons donc notre contrainte de modèle sous-jacent qui doit être sous la forme



En général, une observation a plusieurs variables qui la caractérisent (un appartement est par exemple caractérisé par une surface et un nombre de pièces). Donc une observation est caractérisée par un vecteur



  avec D le nombre de variables caractérisant l'observation.

Dans notre cas, puisqu’on est en une dimension,  D=1, on peut écrire pour une observation

### 

### Définissez la fonction loss

Pour effectuer une régression linéaire, on doit ensuite choisir une fonction de perte dont on a parlé dans le chapitre précédent. On ne va pas prendre de risque, on va prendre la distance euclidienne.

**La distance euclidienne**d'une observation yi vis-à-vis de notre modèle ŷ i est :



**Le risque empirique** dont on parlait dans le chapitre précédent est donc, pour N observations :



Et on va donc chercher à trouver le θ qui minimise cette fonction de perte (qu’on note E) :



***Apprentissage : trouvez le θ optimal***



Maintenant qu’on a notre paramètre θ, c’est-à-dire qu’on a trouvé la droite qui colle le mieux à nos données d’entraînement, on peut effectuer des prédictions sur de nouvelles données, c'est-à-dire prédire le loyer en fonction de la surface qu’on nous donne en entrée, en appliquant directement la formule du modèle ci-dessus.

Par exemple, si on l’applique pour une surface de 35 mètres carrés

1. Rreprésentez graphiquement la droite qu'on a trouvée, pour vérifier qu'elle colle bien aux données.
2. Donnez une prédiction du prix pour une surface de 45 mètres carrés
3. Redessiner le graphe et la prédiction avec linear\_model.LinearRegression()