

On a 10 dimensions (entre parenthèses) : voitures/bus/velo/piétons/nbr de gens dans les voitures, dans les bus, sur les vélos, dans la rue, la température, la force du vent)

Leurs écarts se calculent avec 2^n et la diamètre (la distance la plus grande dans une dimension) se calcule avec √n.

Features extraction engineering : Processus de création et calcul de critères à mesurer (ex : mesurer la taille d’une personne, mesurer la circonférence de la taille, mesurer la couleur des cheveux.)

– Brainstorm

– Pick some

– Make them

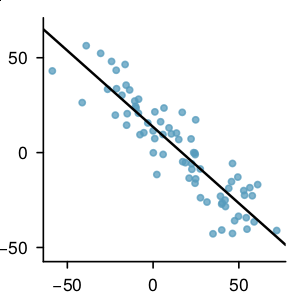
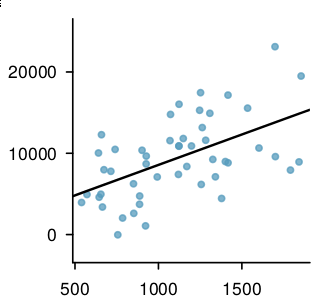
– Evaluate

– Repeat

Modèle Linéaire :

Problème 🡪 prévoir les Y en ne connaissant que les X

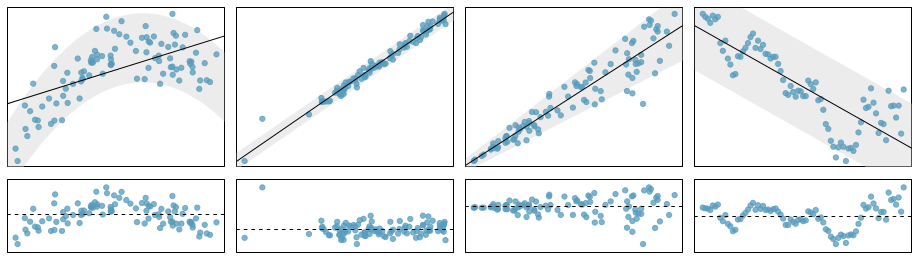
Exemple :



Résidus : ce qui reste du model 🡪 c’est-à-dire les données – le modèle (en gros ce qui reste du filtre)

* yi = (îgrec)i + ei

DONC 🡪 (îgrec)I - yi = résidut = ei



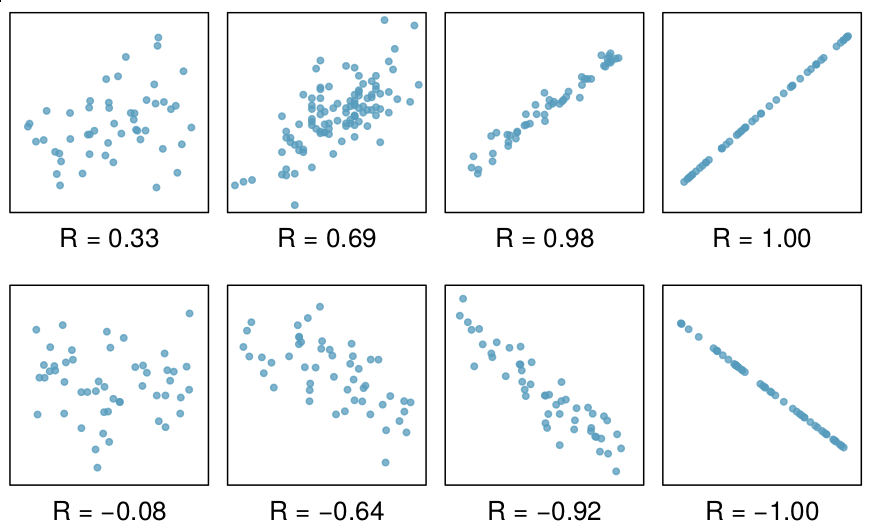
🡨 Modèle

🡨 Résidus

Les résidus doivent être les plus petits possibles.

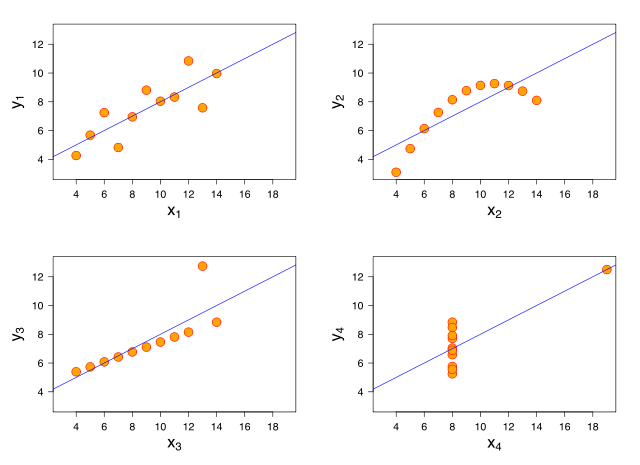
Faire attention aux point aberrants : essayer de trouver l’hypothèse de leur existence dans les données.

Corrélation Positive ou Négative :

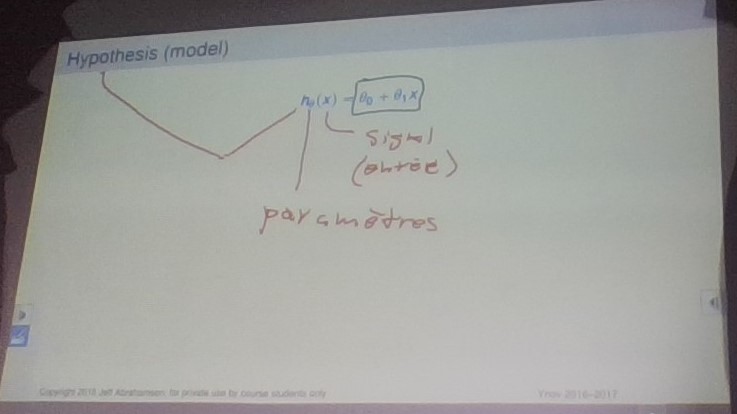


Exemple de l’anneau et le tube -> s’il réchauffe, il est plus grand, c’est positif ; s’il refroidit il est plus petit, c’est une corrélation négative.

Anscombe’s Quartet : Observation de pattern de courbe de modèle :



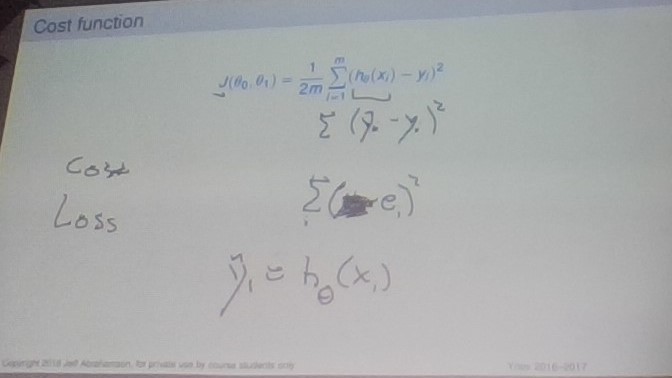
ATTENTION : les corrélations ne sont pas des causes ; il ne faut pas prendre ça en variable confondante.



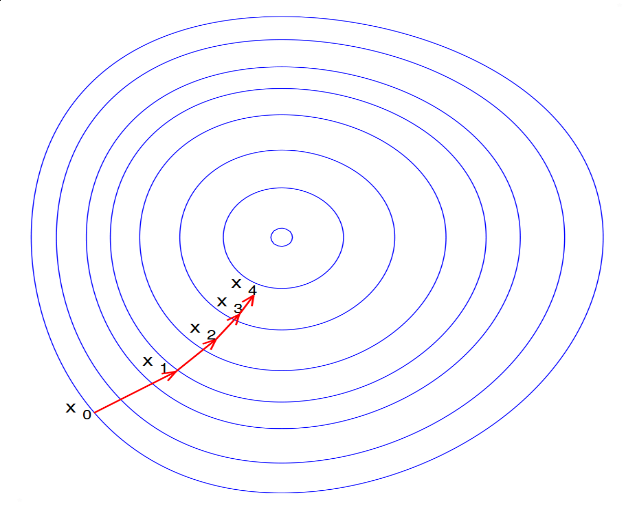
* h têta est la courbe théorique « parfaite » que l’on souhaite obtenir
* Modèle paramétré par têta

Têta 0, têta 1 sont en faites la transposition du vecteur xi.

Loss Function / Cost Function :

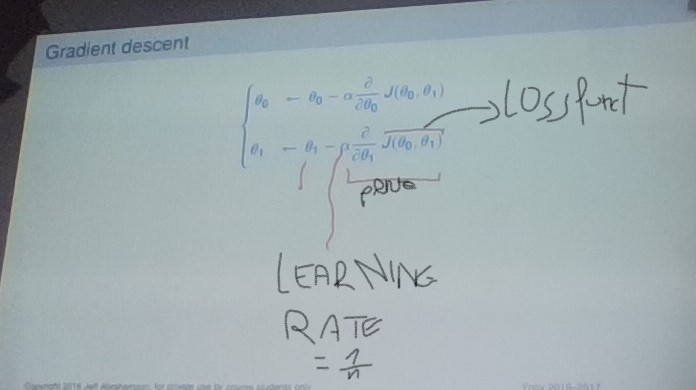


m est le nombre de points total.

Donc (îgrec)I 🡪 est l’hypothèse

Chaque point est un modèle.

Il faut calculer la fonction de perte pour chaque point :

Cela donne une courbe qui ressemble à un bol 🡪 on applique l’algorithme **Gradient descent** qui permet de connaitre la valeur du bas du bol, qui sera la valeur la plus crédible. (Il faut diminuer le pas de changement de valeur dans le bol au fur et à mesure pour être précis). 

La fonction predict -> c’est le (îgrec)i