

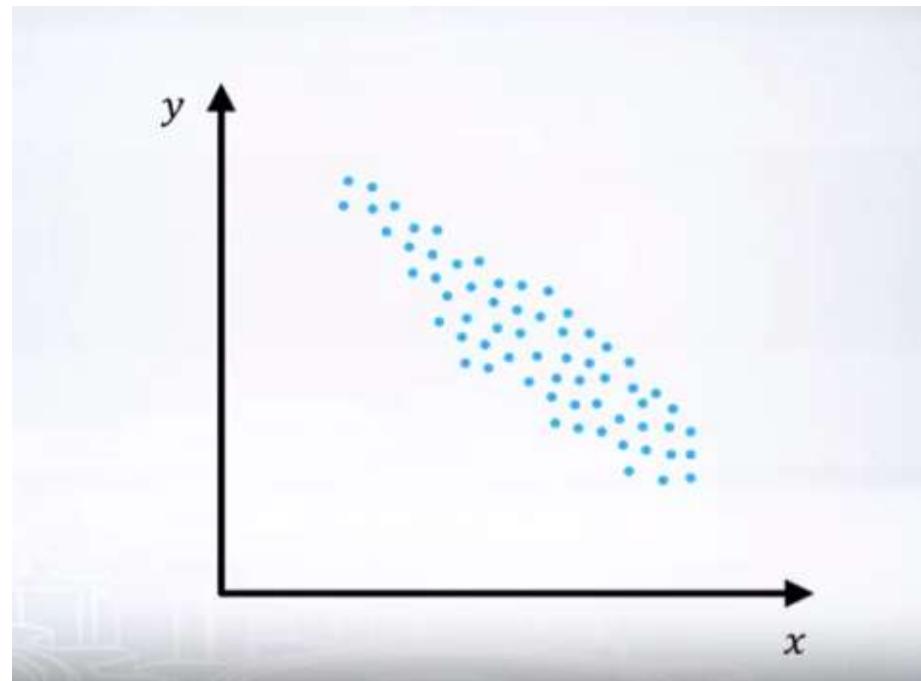
La régression Linéaire

Régression linéaire Définitions

- x_i – exemple
- y_i – target
- $x_i = (x_{i1}, \dots, x_{id})$ – features
- $X = ((x_1, y_1), (x_2, y_2), \dots, (x_l, y_l))$ – training set
- $g(x)$ – modèle, hypothèse
- $x \longrightarrow g(x) \longrightarrow y^{pred}$

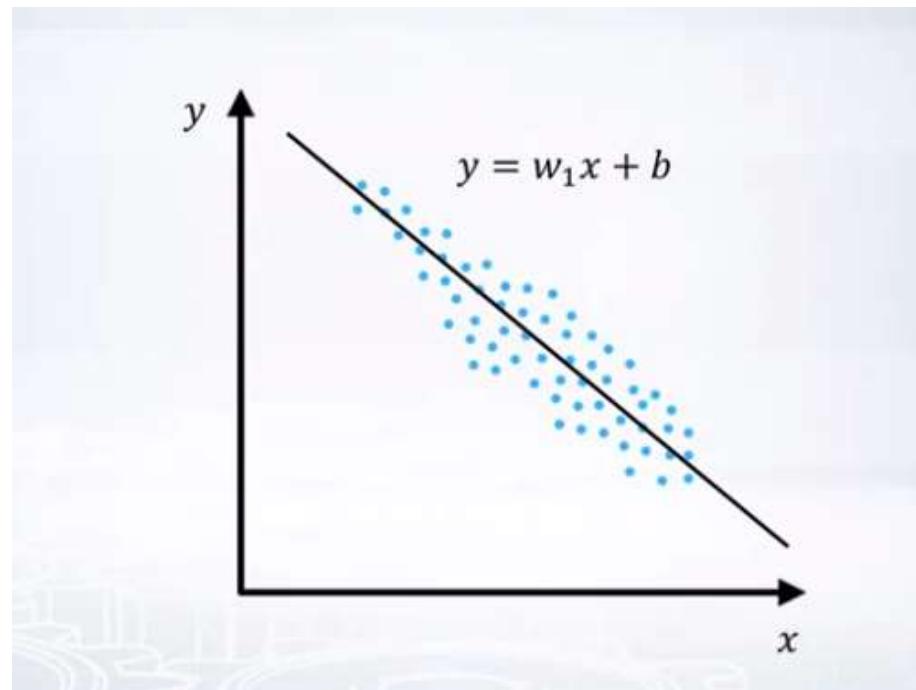
Modèle linéaire pour régression Exemple

Voici le jeu de données, donc nous pouvons voir qu'il y a une tendance linéaire. Si l'entité augmente deux fois, la cible diminue quelque part environ deux fois.



Modèle linéaire pour régression Exemple

Voici le jeu de données, donc nous pouvons voir qu'il y a une tendance linéaire. Si l'entité augmente deux fois, la cible diminue quelque part environ deux fois.



Modèle linéaire pour régression

Des données avec un seul feature

$$g(x) = b + w \cdot x$$

- w – Un seul paramètre (*coefficient*)
- b – *biais*
- En tout, 2 paramètres

Modèle linéaire pour régression

Des données avec plusieurs features: $x = (x_1, x_2, \dots, x_d)$

$$g(x) = b + w_1x_1 + w_2x_2 + \dots + w_dx_d$$

- w_1, \dots, w_d – paramètres (*coefficients*)
- b – *biais*
- $d + 1$ paramètres

Modèle linéaire pour régression

Il serait très pratique d'écrire notre modèle linéaire sous forme vectorielle.

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_d \end{pmatrix}, \boldsymbol{w} = \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_d \end{pmatrix}$$

Notation vectorielle

$$g(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x}$$

Pour un échantillon \mathbf{X}

$$g(\mathbf{X}) = \mathbf{X}\boldsymbol{w}$$

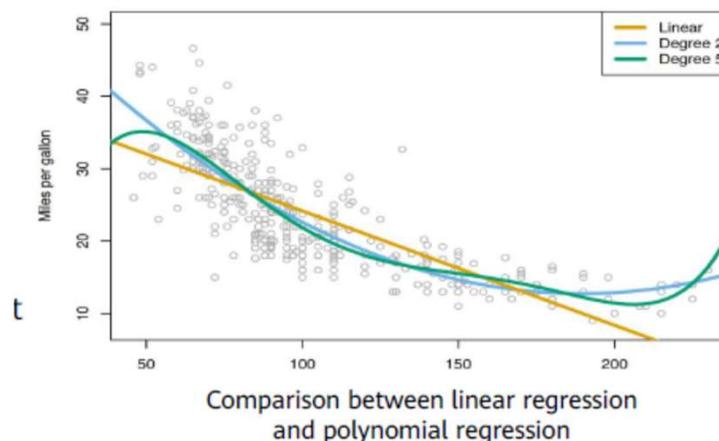
$$\mathbf{X} = \begin{pmatrix} x_{11} & \dots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{\ell 1} & \dots & x_{\ell d} \end{pmatrix}$$

Régression Polynomiale

La régression polynomiale est une extension de la régression linéaire. Généralement, la complexité des données dépasse la possibilité de fitting par une droite. C'est-à-dire qu'un underfitting évident se produit si la régression linéaire est utilisée. La solution est d'utiliser la régression polynomiale.

$$L(w) = w_1x + w_2x^2 + \cdots w_nx^n + b$$

La régression polynomiale est une régression linéaire tant que la relation entre ses paramètres w est toujours linéaire tandis que sa non-linéarité se reflète dans la dimension des features.



Modèle linéaire pour régression: loss function

Loss function MSE

$$L(w) = \frac{1}{l} \sum_{i=1}^l (w^T x_i - y_i)^2$$

Notation matricielle

$$L(w) = \frac{1}{l} \|Xw - y\|^2$$

$$X = \begin{pmatrix} x_{11} & \dots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{\ell 1} & \dots & x_{\ell d} \end{pmatrix}$$

Modèle linéaire pour régression: Question?

- Supposons un échantillon de données X composé de 10 exemples avec 5 features, Combien d'éléments contiendra la matrice X ?
- Combien de d'éléments contiendra le vecteur W ?