Mathématiques pour le machine learning

Module 2

Objectifs

Objectifs

- exprimer des transformations de données grâce à l'algèbre linéaire
- minimiser des fonctions analytiquement
- décrire l'incertain
- décrire des données

Algèbre linéaire

Utilité

- décrire des transformations simples sur un dataset entier avec des mécanismes adaptés
- comprendre les possibilités et les limites de ces transformations simples.

Transformation linéaire

- algèbre linéaire = on se limite aux sommes pondérées des inputs.
- bonne nouvelle : énorme partie des opérations en machine learning

Description des données — échantillon

Python :

```
data = (1, 3)
```

Algèbre linéaire :

$$\mathbf{d} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

Description des données — dataset

Python :

```
data = [(1, 3),
        (2, 2),
        (4, 2)]
```

Algèbre linéaire :

$$\mathbf{D} = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 2 & 2 \end{bmatrix}$$

Description des transformations linéaires

Python :

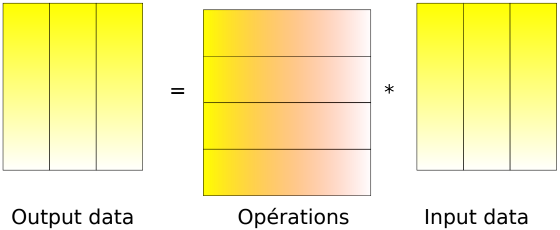
```
def weights(x, y):
    return x * 2 + y / 2
```

Algèbre linéaire :

$$\mathbf{w} = \begin{bmatrix} 2 & \frac{1}{2} \end{bmatrix}$$

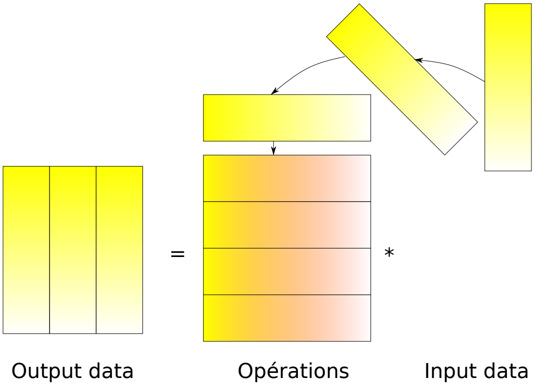
Transformation linéaire = somme pondérée.

Application d'une transformation linéaire à un exemple



Bonne intuition à garder : Verser les colonnes (les exemples du dataset) dans les lignes (les opérations).

Bonne intuition à garder



Bonne intuition à garder : Verser les colonnes (les exemples du dataset) dans les lignes (les opérations).

Application d'une transformation linéaire à un exemple

Python :

```
data = (1, 3)

def weights(x, y):
    return 2 * x + y / 2

res = weights(*data)
```

Algèbre linéaire :

$$f = \begin{bmatrix} 2 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$$= 2 \times 1 + \frac{1}{2} \times 3$$

Application d'une transformation linéaire à un dataset

Python :

```
data = [(1, 3),
        (2, 2),
        (4, 2)]

def f(x, y):
    return x * 2 + y / 2

res = [f(x, y)
       for x, y
       in data]
```

Algèbre linéaire :

$$res = \begin{bmatrix} 2 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 3 & 2 & 2 \end{bmatrix}$$

Application d'une transformation linéaire à un dataset

Python :

```
data = [(1, 3),
        (2, 2),
        (4, 2)]

def f(x, y):
    return x * 2 + y / 2

res = [f(x, y)
       for x, y
       in data]
```

Algèbre linéaire :

$$\text{res} = \begin{bmatrix} 2 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 3 & 2 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 3,5 & \quad & \quad \end{bmatrix}$$

Application d'une transformation linéaire à un dataset

Python :

```
data = [(1, 3),
        (2, 2),
        (4, 2)]

def f(x, y):
    return x * 2 + y / 2

res = [f(x, y)
       for x, y
       in data]
```

Algèbre linéaire :

$$\text{res} = \begin{bmatrix} 2 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 3 & 2 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 3,5 & 5 & \end{bmatrix}$$

Application d'une transformation linéaire à un dataset

Python :

```
data = [(1, 3),
        (2, 2),
        (4, 2)]

def f(x, y):
    return x * 2 + y / 2

res = [f(x, y)
       for x, y
       in data]
```

Algèbre linéaire :

$$\text{res} = \begin{bmatrix} 2 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 3 & 2 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 3,5 & 5 & 9 \end{bmatrix}$$

Application de plusieurs transformations linéaires à un dataset

Python :

```
data = [(1, 3), (2, 2),
        (4, 2)]

def f(x, y):
    return x * 2 + y / 2

def g(x, y):
    return x / 2 + y * 2

res = [[t(x, y) for x, y
                in data]
       for t in [f, g]]
```

Algèbre linéaire :

$$\text{res} = \begin{bmatrix} 2 & \frac{1}{2} \\ \frac{1}{2} & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 3 & 2 & 2 \end{bmatrix}$$

Application de plusieurs transformations linéaires à un dataset

Python :

```
data = [(1, 3), (2, 2),
        (4, 2)]

def f(x, y):
    return x * 2 + y / 2

def g(x, y):
    return x / 2 + y * 2

res = [[t(x, y) for x, y
                in data]
       for t in [f, g]]
```

Algèbre linéaire :

$$\text{res} = \begin{bmatrix} 2 & \frac{1}{2} \\ \frac{1}{2} & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 3 & 2 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 3,5 & & \\ & & \end{bmatrix}$$

Application de plusieurs transformations linéaires à un dataset

Python :

```
data = [(1, 3), (2, 2),
        (4, 2)]

def f(x, y):
    return x * 2 + y / 2

def g(x, y):
    return x / 2 + y * 2

res = [[t(x, y) for x, y
                in data]
       for t in [f, g]]
```

Algèbre linéaire :

$$\text{res} = \begin{bmatrix} 2 & \frac{1}{2} \\ \frac{1}{2} & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 3 & 2 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 3,5 & 5 & \\ & & \end{bmatrix}$$

Application de plusieurs transformations linéaires à un dataset

Python :

```
data = [(1, 3), (2, 2),
        (4, 2)]

def f(x, y):
    return x * 2 + y / 2

def g(x, y):
    return x / 2 + y * 2

res = [[t(x, y) for x, y
                in data]
       for t in [f, g]]
```

Algèbre linéaire :

$$\text{res} = \begin{bmatrix} 2 & \frac{1}{2} \\ \frac{1}{2} & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 3 & 2 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 3,5 & 5 & 9 \\ & & \end{bmatrix}$$

Application de plusieurs transformations linéaires à un dataset

Python :

```
data = [(1, 3), (2, 2),
        (4, 2)]

def f(x, y):
    return x * 2 + y / 2

def g(x, y):
    return x / 2 + y * 2

res = [[t(x, y) for x, y
                in data]
       for t in [f, g]]
```

Algèbre linéaire :

$$\text{res} = \begin{bmatrix} 2 & \frac{1}{2} \\ \frac{1}{2} & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 3 & 2 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 3,5 & 5 & 9 \\ 6,5 & & \end{bmatrix}$$

Application de plusieurs transformations linéaires à un dataset

Python :

```
data = [(1, 3), (2, 2),
        (4, 2)]

def f(x, y):
    return x * 2 + y / 2

def g(x, y):
    return x / 2 + y * 2

res = [[t(x, y) for x, y
                in data]
       for t in [f, g]]
```

Algèbre linéaire :

$$\text{res} = \begin{bmatrix} 2 & \frac{1}{2} \\ \frac{1}{2} & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 3 & 2 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 3,5 & 5 & 9 \\ 6,5 & 5 & \end{bmatrix}$$

Application de plusieurs transformations linéaires à un dataset

Python :

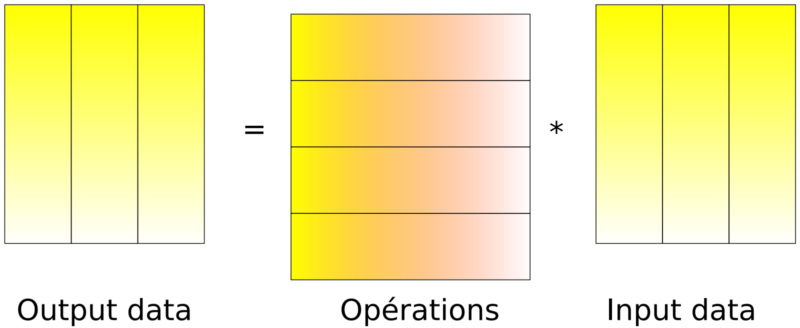
```
data = [(1, 3), (2, 2),
        (4, 2)]

def f(x, y):
    return x * 2 + y / 2

def g(x, y):
    return x / 2 + y * 2

res = [[t(x, y) for x, y
                in data]
       for t in [f, g]]
```

Algèbre linéaire :

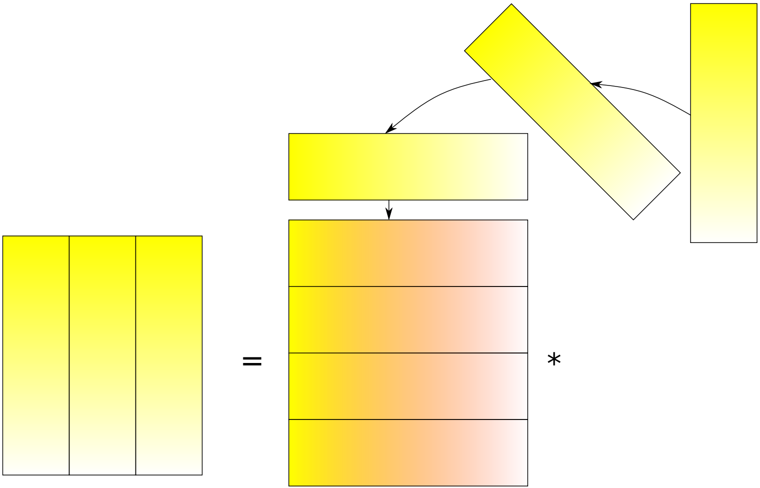
$$\text{res} = \begin{bmatrix} 2 & \frac{1}{2} \\ \frac{1}{2} & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 3 & 2 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 3,5 & 5 & 9 \\ 6,5 & 5 & 6 \end{bmatrix}$$

Bonne intuition à garder



Bonne intuition à garder

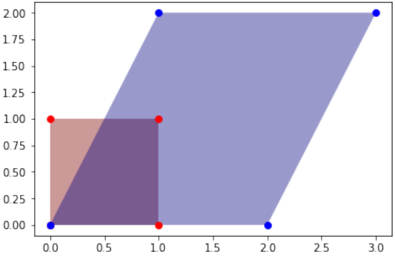


Output data Opérations Input data

Exercice

$$\begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} = ?$$

Exemple de transformation



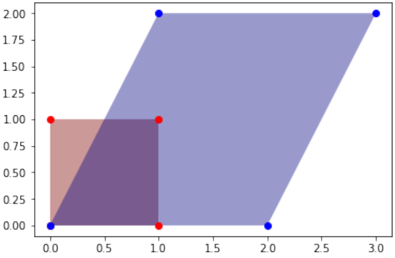
$$\begin{aligned}\text{Bleu} &= \text{Transformation} \times \text{Rouge} \\ &= \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 2 & 3 & 1 \\ 0 & 0 & 2 & 2 \end{bmatrix}\end{aligned}$$

Formes

	Taille
D	(m, n)
W	(o, m)
WD	(o, n)

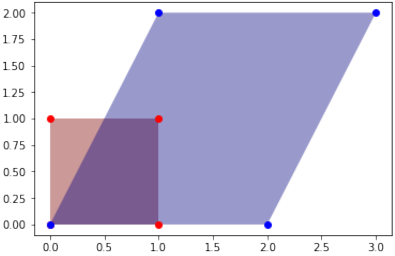
$\rightarrow$ dataset avec m features et n lignes transformé par o opérations donne dataset de o features et n lignes.

Déterminant



Facteur de dilatation de la transformation (ratio aire bleue sur aire rouge).

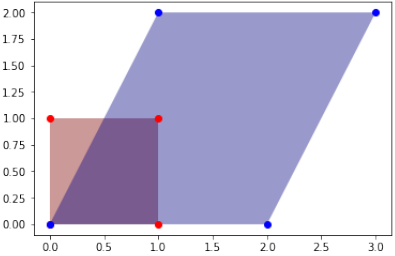
Déterminant



Facteur de dilatation de la transformation (ratio aire bleue sur aire rouge).

Quel est le déterminant de cette transformation ?

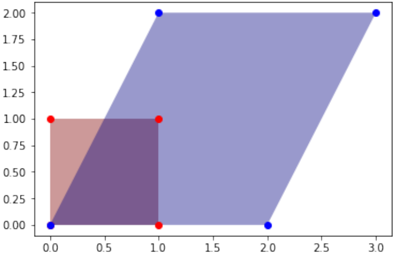
Déterminant



Facteur de dilatation de la transformation (ratio aire bleue sur aire rouge).

Quel est le déterminant de cette transformation ? 4.

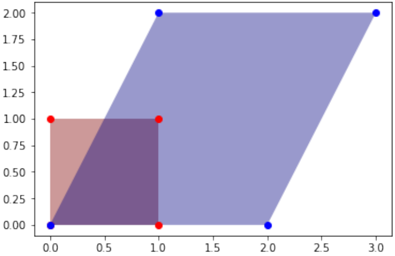
Vecteur propre



Vecteur partant de l'origine qui conserve sa direction malgré la transformation.

Pouvez-vous en trouver un ?

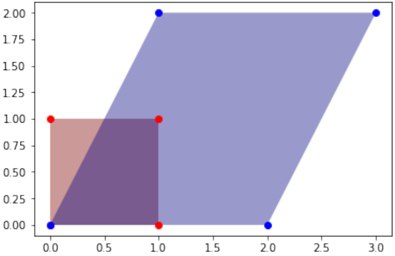
Vecteur propre



Vecteur partant de l'origine qui conserve sa direction malgré la transformation.

Pouvez-vous en trouver un ? $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ par exemple.

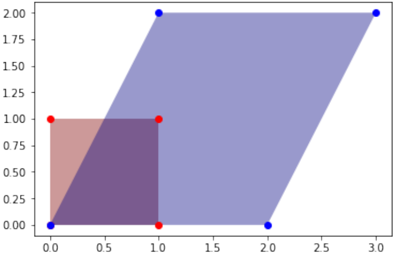
Valeur propre



Facteur par lequel un vecteur propre est redimensionné.

Quelle est la valeur propre de $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$?

Valeur propre



Facteur par lequel un vecteur propre est redimensionné.

Quelle est la valeur propre de $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$? 2.

Analyse

Utilité

Souvent besoin de minimiser une fonction en machine learning.

Utilité

Souvent besoin de minimiser une fonction en machine learning.

(trouver le x pour lequel f est minimale)

Utilité

Souvent besoin de minimiser une fonction en machine learning.

(trouver le x pour lequel f est minimale)

Vous souvenez-vous de comment l'on procède ?

Idée clef

Décider d'un x de départ puis suivre la pente jusqu'au minimum.

Idée clef

Décider d'un x de départ puis suivre la pente jusqu'au minimum.

Pente $=$ dérivée

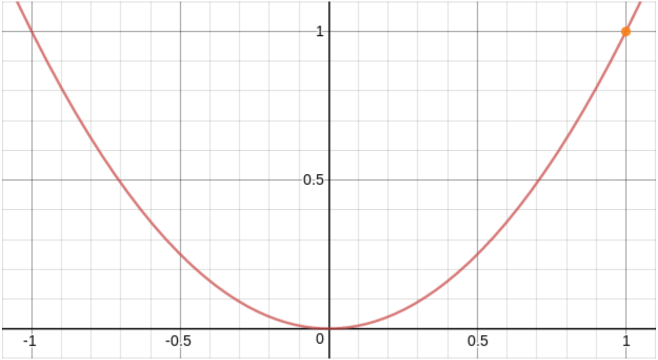
Idée clef

Décider d'un x de départ puis suivre la pente jusqu'au minimum.

Pente = dérivée

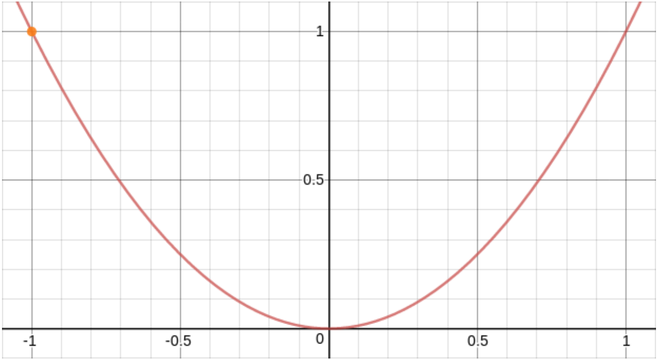
$\rightarrow$ Modifier itérativement x par un pas vers l'opposé de la dérivée.

Pente positive



Opposé de la pente $= -2$. Avec un pas de $0,1$, on passe de 1 à $0,8$.

Pente négative

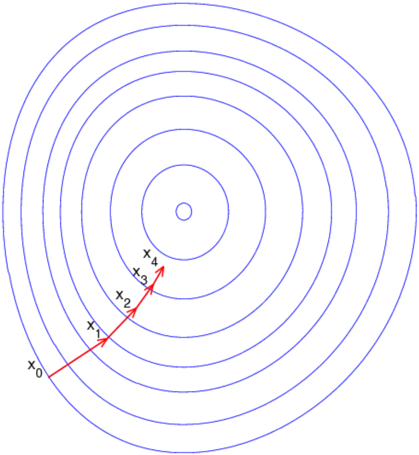


Opposé de la pente = 2. Avec un pas de $0,1$, on passe de -1 à $-0,8$.

Extension à plusieurs dimensions

- dérivée $\rightarrow$ gradient
- identique sinon !

Exemple en 2 dimensions



Probabilités

Utilité

- quantifier l'incertain
- support pour les statistiques

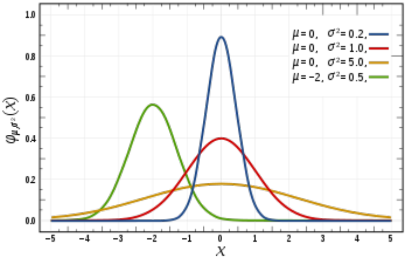
Probabilité

- la probabilité de l'événement X est notée $P(X)$
- $P(X) \in [0, 1]$
- $P(X) = 0 \iff$ X est impossible
- $P(X) = 1 \iff$ X est certain
- $P(\neg X) = 1 - P(X)$

Loi de probabilité

Décrit le comportement aléatoire d'un phénomène dépendant du hasard.

- $\sum_u P(X = u) = 1$ en discret
- $\int P(X)dX = 1$ en continu
- loi uniforme
- loi normale/gaussienne

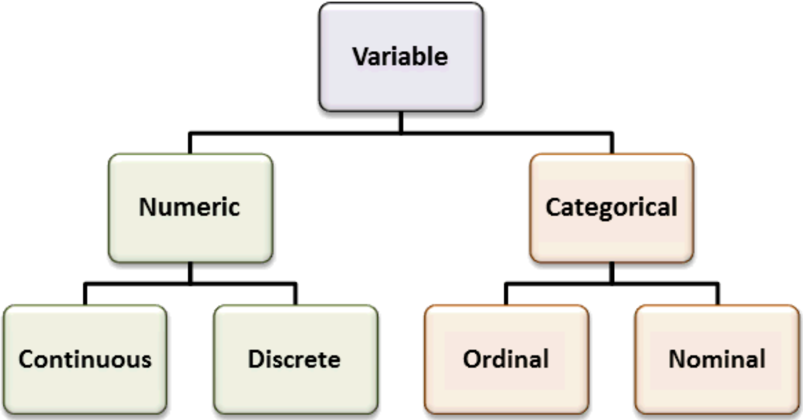


Statistiques

Utilité

- description et compréhension des données
- correction pour faciliter les traitements

Types de variables



Hypothèse

Pré-requis pour les mesures statistiques qui suivent (et la plupart du machine learning) :

- les données **doivent être issues d'une même loi**
- chaque échantillon doit être **indépendant** des autres
- pas évident en pratique !

Hypothèse

Pré-requis pour les mesures statistiques qui suivent (et la plupart du machine learning) :

- les données **doivent être issues d'une même loi**
- chaque échantillon doit être **indépendant** des autres
- pas évident en pratique ! Pourquoi ?

Variance

Mesure la dispersion d'une série statistique (ou d'une variable) :

$$V(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

Pour la calculer :

$$V(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

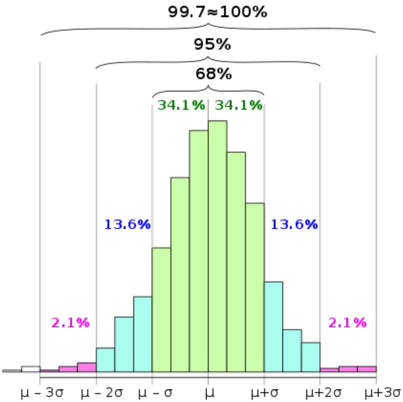
Écart-type

Racine carrée de la variance

$$\sigma(X) = \sqrt{V(X)}$$

Écart-type — règle des 68, 95 et 99,7

Pour les lois normales :



Quartile

Les quartiles (Q_1, Q_2 et Q_3) divisent les données en 4 intervalles contenant le même nombre d'observations.

Déclinable en quantile de taille arbitraire (décile, percentile).

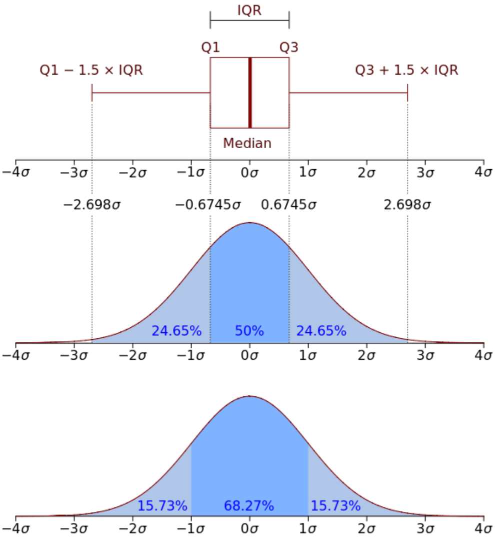
Quartile

Les quartiles (Q_1, Q_2 et Q_3) divisent les données en 4 intervalles contenant le même nombre d'observations.

Déclinable en quantile de taille arbitraire (décile, percentile).

Que veut dire être dans le 95^{e} percentile ?

Boxplot



IQR
Q1
Q3
Q1 − 1.5 × IQR
Q3 + 1.5 × IQR
Median
−4σ −3σ −2σ −1σ 0σ 1σ 2σ 3σ 4σ
−2.698σ
−0.6745σ
0.6745σ
2.698σ
24.65%
50%
24.65%
−4σ −3σ −2σ −1σ 0σ 1σ 2σ 3σ 4σ
15.73%
68.27%
15.73%
−4σ −3σ −2σ −1σ 0σ 1σ 2σ 3σ 4σ

Covariance

Mesure la variabilité jointe de deux variables aléatoires :

$$V(X) = \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])\right]$$

$$\mathrm{cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$$

Pour la calculer :

$$\mathrm{cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Corrélation

Covariance divisée par le produit des écart-types :

$$\mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Intérêt ?

Corrélation

Covariance divisée par le produit des écart-types :

$$\mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}$$

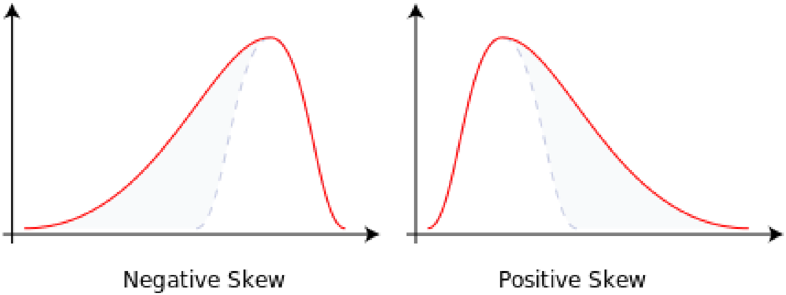
Intérêt ? Pas d'unité.

Test de normalité

Pour tester (et corriger) la normalité d'une distribution, on utilise deux mesures :

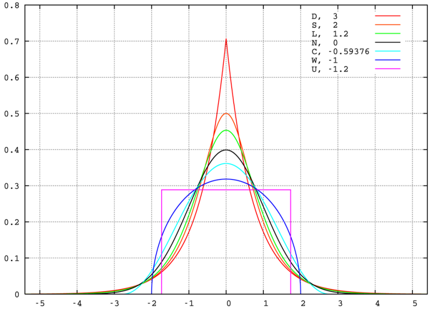
- l'asymétrie (*skew*)
- le kurtosis

Asymétrie



$$\mathrm{asym}(X) = \mathbb{E}\left[\left(\frac{X - \bar{X}}{\sigma}\right)^3\right]$$

Kurtosis



$$\mathrm{kurt}(X) = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$$

Transformation de Box-Cox

Asymétrie et kurtosis peuvent se corriger avec la transformation de Box-Cox ou des transformations log.

Quizz

Algèbre linéaire

- comment décrit-on on enregistrement ?

Algèbre linéaire

- comment décrit-on on enregistrement ?
- comment décrit-on une transformation linéaire ?

Algèbre linéaire

- comment décrit-on on enregistrement ?
- comment décrit-on une transformation linéaire ?
- quel est le sens de la multiplication de matrices dans le contexte dataset/opérations ?

Analyse

- intuitivement, qu'est-ce que nous apprend une dérivée ?

Analyse

- intuitivement, qu'est-ce que nous apprend une dérivée ?
- à quelle valeur de la dérivée d'une fonction atteint-on un minimum ?

Analyse

- intuitivement, qu'est-ce que nous apprend une dérivée ?
- à quelle valeur de la dérivée d'une fonction atteint-on un minimum ?
- si la dérivée est négative, dans quel sens faut-il faire évoluer x ?

Analyse

- intuitivement, qu'est-ce que nous apprend une dérivée ?
- à quelle valeur de la dérivée d'une fonction atteint-on un minimum ?
- si la dérivée est négative, dans quel sens faut-il faire évoluer x ?
- est-ce que la pas d'apprentissage impacte seulement les performances en temps de calcul ?

Probabilités

- quelle est la forme d'une gaussienne ?

Probabilités

- quelle est la forme d'une gaussienne ?
- à combien somme une loi discrète ?

Statistiques

- quelle est l'hypothèse que l'on fait dans la plupart des approches de statistiques / machine learning ?

Statistiques

- quelle est l'hypothèse que l'on fait dans la plupart des approches de statistiques / machine learning ?
- que nous apprennent la variance et l'écart-type ?

Statistiques

- quelle est l'hypothèse que l'on fait dans la plupart des approches de statistiques / machine learning ?
- que nous apprennent la variance et l'écart-type ?
- que nous apprennent la covariance et la corrélation ?

Statistiques

- quelle est l'hypothèse que l'on fait dans la plupart des approches de statistiques / machine learning ?
- que nous apprennent la variance et l'écart-type ?
- que nous apprennent la covariance et la corrélation ?
- comment peut-on savoir si une distribution est normale ?

Conclusion

Conclusion

- algèbre linéaire $\rightarrow$ raisonner sur des opérations simples et les décrire efficacement
- minimiser une fonction continue $\rightarrow$ dérivée
- décrire l'incertain $\rightarrow$ probabilités
- caractériser une série de données $\rightarrow$ statistiques