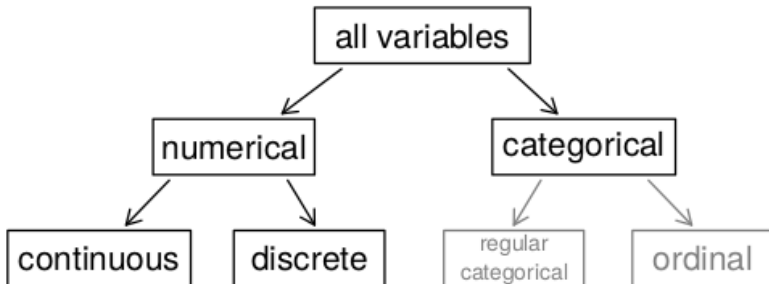


Les différents types de variables :



Moyenne

$$\mu(X) = \frac{1}{N} \sum_N X_i$$

Variance

$$V(X) = \mu \left[(X - \mu[X])^2 \right] = \mu[X^2] - \mu[X]^2$$

Ecart-type

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



High bias, low variance



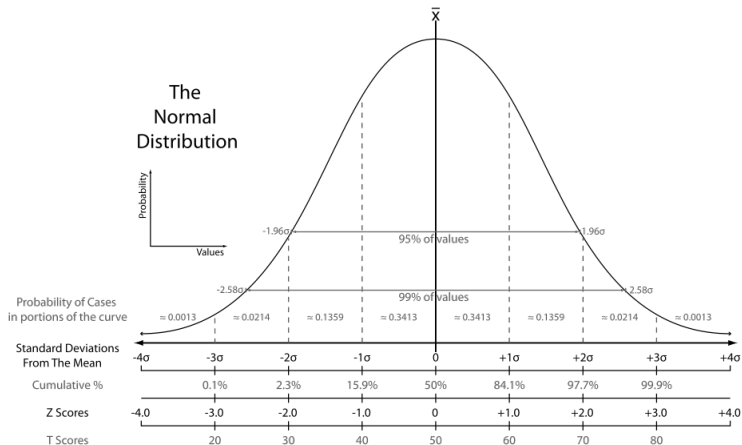
Low bias, high variance

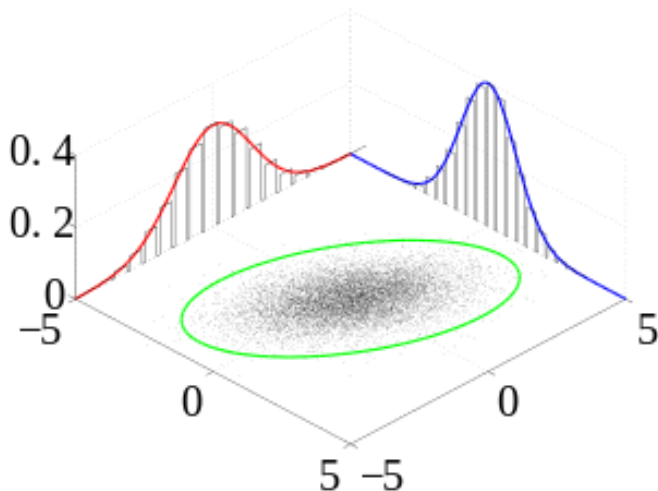


High bias, high variance



Low bias, low variance





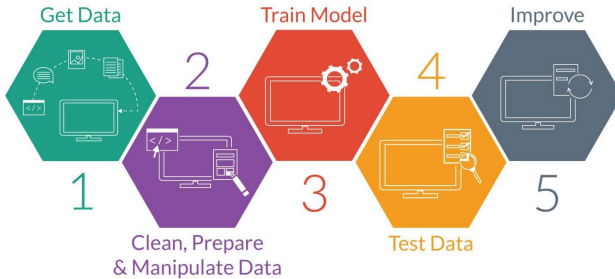
On cherche les lois de probabilités qui génèrent (au mieux) nos données.

4 phases :

- Identifier une question ou un problème
- Collecter des données
- Analyser les données (moyenne, biais, variance, covariance, intervalles de confiance...)
- Conclure

Machine Learning

On cherche un modèle qui approxime (au mieux) le comportement d'un "expert" sur des données.



Les problématiques :

- Les données marginales
- Données biaisées
- Dépendance statistique \neq causalité
- Pas assez de données ?
- ...