

Big Data Analytics

Réduction de la dimensionalité

Comment appréhender des données en grande dimension ?

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,D} \\ X_{2,1} & X_{2,2} & \dots & X_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N,1} & X_{N,2} & \dots & X_{N,D} \end{bmatrix}$$

La malédiction des grandes dimensions !
(Curse of dimensionality)

- Sélection de dimensions
- Projections linéaires (ACP, LDA, ...)
- Projections non-linéaires (kernels, neural network embeddings, ...)

Sélection de dimensions :

- Random forest
- SVM
- ...