

# Statistiques et apprentissage

## Enseignant

Thibaut Le Gouic

## Contact école

Thibaut Le Gouic

## Compétences acquises

Regression

Statistical learning theory

Supervised/unsupervised  
classification

## Parcours donnant accès à ce cours

OMIS

MMEFI (électif)

La **reconnaissance vocale** (Siri, Google Now, S Voice), la reconnaissance de caractère (OCR), les systèmes de recommandation (Netflix, Amazon, Allociné), l'aide au diagnostic médical sont parmi les nombreux problèmes qui peuvent se modéliser mathématiquement par l'apprentissage d'une fonction  $f$  :

**son** → mot/phrase prononcé

**image** → caractère lu

**utilisateur** → recommandation de produit

**image radio** → décision médicale.

On modélisera ces problèmes de manière générique par l'égalité :

$$\mathbf{Y} = \mathbf{f}(\mathbf{x}) + \varepsilon$$

Ce que l'on cherche

Aléa ou imprécision/imperfection du modèle

Observé sur quelques exemples pour « apprendre »  $f$

Observation

où  $\mathbf{X}$ ,  $\mathbf{Y}$  sont des observations qui permettent d'apprendre  $f$ , et  $\varepsilon$  est l'aléa qui représente l'information qu'il manque pour prédire complètement  $\mathbf{Y}$  à partir de  $\mathbf{X}$ . Nous étudierons différentes méthodes pour apprendre  $f$  à partir d'un jeu de données.