

# Biais/Variance

On veut prédire  $Y$  à partir de  $X$ . Pour cela, on cherche à estimer  $f(X)$  la fonction qui relie  $X$  à  $Y$  par

$$Y = f(X) + \epsilon$$

On peut mesurer deux fonctions qui vont nous aider :

- Fonction de perte ( $L$ ) : c'est la mesure de l'écart par rapport à ce qu'on souhaite mesurer, par exemple :

$$L(Y, f(X)) = (Y - f(X))^2.$$

- Fonction de risque : c'est la quantité que l'on cherche à minimiser. Il s'agit de l'espérance de la fonction de perte.

Comment trouver  $f$  :

Objectif : Trouver une fonction  $\hat{f}$  qui minimise le risque.

Comment : Supposer une certaine forme pour  $f(X)$  et minimiser la fonction de perte de façon analytique ou numérique.

- Paramétrique : On donne une forme explicite à  $f(X)$  qui dépend de paramètres. On cherche une méthode d'estimation des paramètres.
- Non-paramétrique : aucune forme particulière de  $f$ , on estime une courbe ou fonction.

Exemple : la régression linéaire simple

On suppose que  $f(X) = \beta_0 + \beta_1 X$ , on cherche donc les valeurs  $\hat{\beta}_0$  et  $\hat{\beta}_1$  qui minimisent :

$$\mathbb{E}(Y - (\beta_0 + \beta_1 X))^2.$$

Que faire dans le cas discret ?

On cherche à prédire  $G$  (un groupe ou facteur) à partir de  $X$ . Supposons que nous avons  $\widehat{G}$  qui prédise le classe des observations sachant  $X$ , alors on peut définir la fonction de perte (0-1) comme le nombre d'erreur que l'on a effectué :

$$L(G, \widehat{G}) = \mathbb{1}_{G \neq \widehat{G}}.$$

Avec cette fonction de perte 0-1,  $\widehat{G}(x)$  est la classe  $g$  qui maximise  $\mathbb{P}(g|X = x)$ .