CH1 Intro   Apprentissage machine

couple image

Couple aléatoire $Z = (X, Y) \in \mathcal{X} \times \mathcal{Y}$

existe ✓

obj : trouve 1 fct qui lie X à Y : $\mathbb{P}(Y|X)$

(proba conditionnelle)

peut exister pas

$\exists$ 1 appli $T : \mathcal{X} \to \mathcal{Y}$ $\uparrow$ $Y \approx T(X)$ où T déterminist

obj : trouver $h \approx T$ → peut avoir des erreurs (1 chat 识别 1 chien)

1 fct nous dire chat ou chien

risque moyen   défini par $\mathcal{E}(h) = \int \ell(h(x), y) \, d\mathbb{P}(X, Y)$ (chercher à minimiser

err moy de classificat° qu'on fait

(risque moy) (de prédict° donnée)

où $\mathbb{P}(\cdot,\cdot)$ la loi jointe entre X et Y.

$\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R} \sqcup \{+\infty\}$

est 1 fct de perte. (pour mesure la qualité d'1 fct de prédict° h)

On va chercher $\boxed{h^* = \underset{h : \mathcal{X} \to \mathcal{Y}}{\operatorname{argmin}} \, \mathcal{E}(h)}$ $(\mathcal{P})$

Pour rendre ce pbl attaquable on va utiliser 2 ingrédients :

risque empirique et régularisation.

peut ê calculé en discrétisant Z. On remplace $\mathcal{E}(h)$ par $\mathcal{E}_n(h)$ définie par $\mathcal{E}_n(h) = \sum_{i=1}^{n} \ell(h(x_i), y_i)$

où (x_i, y_i) sont tirés au hasard suivant la loi $\mathbb{P}$.

(On discrétise l'ens : remplacer esp infini → esp fini)

Consiste à choisir un espace de fct $\mathcal{H}$ décrete par 1 nbr fini de paramètres

On remplace ainsi : $(\mathcal{P})$ par $(\mathcal{P}_n)$:

$\boxed{\inf_{h \in \mathcal{H}} \mathcal{E}_n(h)}$ $(\mathcal{P}_n)$

- Quel est l'effet de minimiser $\mathcal{E}_n$ plutôt que $\mathcal{E}$?
-                 de la régularisat° par $\mathcal{H}$?
- Comment minimise $(\mathcal{P}_n)$ numériquement ?
- Quelle
- Quelle est la prédict° de minimisat° mé

Exemple. 1) Classification

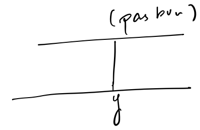
En classificat°, $\mathcal{Y}$ discrèt e.g. $\{0, 1, \cdots, 9\}$

$\{$chat, chien$\}$

la fct perte est typiquement : 0 , 1

- $\ell(\hat{y}, y) = \begin{cases} 0 & \text{si } \hat{y} = y \\ 1 & \text{sinon} \end{cases}$

(pas bon)



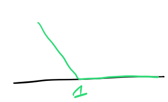
$\hat{y}$: val° prédite

$(\hat{y} = h(x))$

y : vrai val°

- Fct " hinge "

$\ell(\hat{y}, y) = \max(0, 1 - \hat{y}y)$



$\begin{cases} h(x) \leq -1 & \text{si } y = -1 \\ h(x) \geq 1 & \text{si } y = 1 \end{cases}$

- Divergence de Kullbach-Leibler.

$\ell(\hat{y}, y) = D(\hat{y} \| y) = \sum_{i=1}^{I} y_i \log\left(\frac{y_i}{\hat{y}_i}\right)$ I : nbr de classes

2) Regression :

En Regression, l'espace $\mathcal{Y}$ est continu (e.g. $\mathbb{R}^n$)

Par exemple, supposons que $Y = A(X) + B$, où $A : \mathbb{R}^n \to \mathbb{R}^n$

$B \sim \mathcal{N}(0, \sigma^2 I_n)$

image floue — image net ↳ à trouver — bruit

$$\min_{h : \mathbb{R}^n \to \mathbb{R}^n} \iint \ell(h(A(x)+B), \underbrace{x}_{\text{vrai image}}) \, \underbrace{dP_X(x)}_{\text{ttes image poss}} \, \underbrace{dP_B(b)}_{\text{tt réalisation du bruit poss}}$$

Dans ce cas, on prend typiquement :

- $\ell(\hat{y}, y) = \frac{1}{2} \|\hat{y} - y\|_2^2$ (risque quadratique)
- $\ell(\hat{y}, y) = \|\hat{y} - y\|_1$ (robuste)
- $\ell(\hat{y}, y) = NN(\hat{y}, y)$ où NN est un réso de neurones.

Exempl de regularisation : De façon générale, les fct de prédict° sont paramétrees par des poids $w \in \mathbb{R}^d$.

Une fct de la famille $\mathcal{H}$ est définie par : $h : \underbrace{\mathcal{X}}_{\text{esp d'entrée}} \times \mathcal{W} \to \mathcal{Y}$

$(x, w) \mapsto h(x, w)$

Donc $\mathcal{H} = \{ h(\cdot, w), w \in \mathcal{W} \}$ où $\mathcal{W}$ est l'espace des poids admissibles :

- $\mathbb{R}^d$
- $\mathbb{R}_+^d$
- $[0, 1]^d$