

Problème d'apprentissage

1

- Données
 - $D_n = (Z_1, Z_2, \dots, Z_n)$ générées par la "nature"
- IID: "independent and identically distributed"
 - tirés de la même distribution INCONNUE $p(Z)$
 - de manière indépendante

Problème d'apprentissage

2

- Les trois problèmes considérés
 - **classification**: $Z = (X, Y) \in \mathbb{R}^d \times \{1, \dots, N\}$ ($\mathbb{R}^d \times \{-1, 1\}$)
 - **régression**: $Z = (X, Y) \in \mathbb{R}^d \times \mathbb{R}$
 - **estimation de densité**: $Z \in \mathbb{R}^d$
- Ensemble de fonctions \mathcal{F} (solutions possibles), $f \in \mathcal{F}$:
 - **classification**: $f: \mathbb{R}^d \rightarrow \{-1, 1\}$
 - **régression**: $f: \mathbb{R}^d \rightarrow \mathbb{R}$
 - **estimation de densité**: \mathcal{F} contient des fonctions de densité

Problème d'apprentissage

3

- La perte $L(Z, f)$
 - mesurer la qualité d'une solution sur un point de donnée
 - **classification**: $L(Z, f) = L((X, Y), f) = \begin{cases} 0 & \text{si } f(X) = Y, \\ 1 & \text{si } f(X) \neq Y \end{cases}$
 - **régression**: $L(Z, f) = L((X, Y), f) = (f(X) - Y)^2$
 - **estimation de densité**: $L(Z, f) = -\log f(Z)$

Problème d'apprentissage

4

- Le risque
 - perte espérée, erreur de généralisation:
$$R(f) = E[L(Z, f)] = \int L(Z, f) p(Z) dz$$
 - **classification**: probabilité de mauvaise classification
 - **régression**: erreur quadratique espérée
 - **estimation de densité**: "distance" de l'estimation de la vrai densité

Problème d'apprentissage

5

- Le problème de l'induction
 - trouver $f \in \mathcal{F}$ qui minimise le risque $R(f)$
 - $p(Z)$ est INCONNU!!!

Problème d'apprentissage

6

- Erreurs d'estimation et d'approximation, capacité
 - la sortie de notre algorithme d'apprentissage: $\hat{f}_n(D_n) = \hat{f}_n$
 - la meilleure fonction dans \mathcal{F} :
$$f_{\mathcal{F}}^* = \arg \min_{f \in \mathcal{F}} R(f)$$
 - la meilleure fonction possible (la décision/l'erreur de Bayes):
$$f^* = \arg \min R(f)$$
 - $R(\hat{f}_n) - R(f^*) = (R(\hat{f}_n) - R(f_{\mathcal{F}}^*)) + (R(f_{\mathcal{F}}^*) - R(f^*))$

• Le risque empirique

- perte moyenne, erreur empirique, erreur d'entraînement:

$$\widehat{R}(f, D_n) = \widehat{R}(f) = \frac{1}{n} \sum_{i=1}^n L(Z_i, f)$$

- $\widehat{R}(f, D_n)$ est un **estimateur bruité** mais **non-biaisé** de $R(f)$:

$$E_{D_n}[\widehat{R}(f, D_n)] = R(f)$$

- **classification**: la fréquence des points mal classifiés
- **régression**: erreur quadratique moyenne
- **estimation de densité**: la vraisemblance négative pour que les points soient générés par f

• Le principe de minimisation du risque empirique

- trouver $f \in \mathcal{F}$ qui minimise $\widehat{R}(f)$:

$$\widehat{f}_n^*(D_n) = \widehat{f}_n^* = \arg \min_{f \in \mathcal{F}} \widehat{R}(f, D_n)$$

- Lemme (Vapnik & Chervonenkis)

$$R(\widehat{f}_n^*) - R(f_{\mathcal{F}}^*) \leq 2 \max_{f \in \mathcal{F}} |\widehat{R}(f) - R(f)|$$

- question **théorique**: est-ce que $\widehat{R}(f)$ **approche bien** $R(f)$ ($f \in \mathcal{F}$)?
- réponse: oui, si la **capacité (complexité)** de \mathcal{F} est **petite**

• Estimer l'erreur de généralisation

- problème: $\widehat{R}(\widehat{f}_n^*)$ est un **estimateur biaisé** de $R(\widehat{f}_n^*)$

- **ensemble de test**: $D'_m = (Z'_1, Z'_2, \dots, Z'_m)$

- **erreur de test**:

$$\widehat{R}(\widehat{f}_n^*, D'_m) = \frac{1}{m} \sum_{i=1}^m L(Z'_i, \widehat{f}_n^*)$$

- si D'_m est iid et indépendant de D_n , alors $\widehat{R}(\widehat{f}_n^*, D'_m)$ est un **estimateur non-biaisé** de $R(\widehat{f}_n^*)$:

$$\lim_{m \rightarrow \infty} \widehat{R}(\widehat{f}_n^*, D'_m) = R(\widehat{f}_n^*)$$

• Problèmes de l'apprentissage

- **algorithmique**: choisir \mathcal{F} tel que \widehat{f}_n^* est facile à trouver
- **théorique 1**: choisir \mathcal{F} tel que $\widehat{R}(\widehat{f}_n^*)$ est près de $R(\widehat{f}_n^*)$
- **théorique 2**: choisir \mathcal{F} tel que $R(\widehat{f}_n^*)$ est près de $R(f^*)$

• Courbes d'apprentissage

- Le principe de **minimisation du risque structurel**