

Exercice-Ethème 1

Thomas Laurent

14/01/2018

Questions

Question 1

L'apprentissage supervisé fait référence à des systèmes apprenant à partir d'observations étiquetées et s'intéresse à la découverte de relations à partir d'un échantillon de taille limitée. Enfin, l'apprentissage supervisé consiste à déterminer une fonction permettant de prédire automatiquement la classe à laquelle l'observation appartient. Si les étiquettes (sorties) sont continues alors il s'agit d'une régression. Autrement, si les étiquettes sont discrètes, il s'agit d'une classification.

Question 2

La fonction de perte correspond à une fonction évaluant l'erreur globale commise par une hypothèse h de l'espace des hypothèses et est notée $l(y, h(x))$ et mesure l'écart et le coût de l'écart entre la prédiction $h(x)$ et la valeur y donnée par le superviseur. La perte binaire est un exemple de fonction de perte et est définie de la manière suivante:

$l(h(x), y) = 0$ si $h(x) = y$ et égale à 1 sinon

Le risque fonctionnel correspond à l'espérance de la fonction de perte:

$$R(h) = \mathbb{E}(l(h(x), y)) = \int_{X,Y} l(h(x), y) dP(x, y)$$

Question 3

Le risque empirique se définit comme la somme des fonctions de perte sur les n éléments de $X \times Y$ supposés indépendants et identiquement distribués:

$$R_n(h) = \frac{1}{n} \sum_{i=1}^n l(y_i, h(x_i))$$

La minimisation du risque empirique consiste à ce que l'algorithme choisisse l'hypothèse optimale à partir de $\operatorname{argmin}_{h \in H} R_n(h)$, le minimiseur du risque empirique. L'hypothèse retenue est ainsi l'hypothèse de risque empirique minimal, h_n^* . On peut noter que cette approche repose sur la loi faible des grands nombres.

Question 4

La généralisation consiste à appliquer le prédicteur sur des entrées inconnues (nouvelles observations).

Question 5

Le sur-apprentissage se définit par une trop forte cardinalité de H ce qui conduit à une réduction du risque empirique et en contrepartie à une généralisation de qualité médiocre.

Compromis optimisation-approximation-estimation

Le tableau de variation de ϵ_{app} , ϵ_{est} , ϵ_{opt} et T en fonction de H , n et ρ est donné ci-dessous:

Paramètres	H	n	ρ
ϵ_{app}	\downarrow	X	X
ϵ_{est}	\downarrow	\downarrow	X
ϵ_{opt}	X	X	\uparrow
T	\uparrow	\uparrow	\downarrow

L'erreur d'apprentissage dépend seulement de H . Si H augmente alors $R(h_H^*)$ ne peut que diminuer et $R(h^*)$ étant indépendant de H , l'erreur d'apprentissage ne peut que diminuer.

L'erreur d'estimation dépend du nombre d'observations et de H l'ensemble des hypothèses. ρ n'influence donc pas l'erreur d'estimation par définition. Plus l'ensemble H et plus le nombre d'observations sont grands, meilleure sera l'erreur d'estimation.

L'erreur d'optimisation sera d'autant plus grande que ρ augmente par définition. La dimension de H et le nombre d'observations n'influence pas l'erreur d'optimisation et dépend seulement de l'algorithme utilisé.

Pour le temps de calcul T , plus la dimension de l'ensemble des hypothèses sera grande, plus le nombre d'hypothèses à tester sera important, et cela nécessitera un temps de traitement plus important. Le temps de calcul augmente également en fonction du nombre d'observations. L'augmentation de ρ revient à augmenter la tolérance du processus d'apprentissage et ainsi, un minimiseur de la fonction objective sera trouvé en testant moins d'hypothèses. Cela nécessite donc un temps de calcul moindre.