L'entrainement et le test

Exercices

Exercice.1

Considérons l'espace d'entrée $X = \{1\} \times \mathbb{R}^d$, montrez que la dimension VC du perceptron ayant d+1 paramètres $(\omega_0, \omega_1, \dots, \omega_d)$ est exactement égale à d+1, en montrant qu'elle est inférieure ou égale à d+1 et supérieure ou égale à d+1.

(a) Pour montrer que $d_{VC} \geq d + 1$, trouvez d+1 points dans X que le perceptron peut diviser. Construire une matrice non-singulière de taille (d+1) × (d+1) dont les lignes représentent les d+1 points, puis utiliser la non-singularité pour prouver que le perceptron peut diviser ces points.
(b) Pour montrer que $d_{VC} \leq d + 1$, montrer qu'il n'existe aucun ensemble de d+2 points dans X que le perceptron peut diviser.

Exercice.2

Supposons qu'on a un modèle d'apprentissage simple dont la fonction de croissance est $m_H(n) = n + 1$, donc $d_{VC} = 1$. Supposons qu'on possède un échantillon d'entrainement de taille 100. Utiliser la limite VC pour estimer la probabilité que E_{gen} sera dans un intervalle de 0.1 de E_{emp}.

Exercice.3

Supposons qu'on a un ensemble de données contenant 600 exemples. Pour bien tester la performance de l'hypothèse finale, on divise les données en ensemble d'entrainement 400 échantillons et ensemble de test 200 échantillons. On utilise un modèle d'apprentissage contenant 1000 hypothèses et on sélectionne l'hypothèse finale g en se basant sur 400 exemples d'entrainement.

On veut estimer $E_{gen}(g)$. Pour cela on a accès à deux estimations $E_{emp}(g)$ et $E_{test}(g)$.

(a) Soit $\delta = 0.05$, quelle est l'estimation qui possède la plus grande erreur ?
(b) Existe-t-il une raison qui nous empêche de réserver beaucoup plus d'exemples pour le test ?

Exercice.4

Pour les fonctions cibles binaires, montrer que $P[h(x) \neq f(x)]$ peut-être écrite comme l'espérance de l'erreur quadratique moyenne dans les cas suivants :

(a) La fonction cible binaire prend les valeurs {0,1}.
(b) La fonction cible binaire prend les valeurs {-1,1}.

Exercice.5

Prouver par induction que :

$$\sum_{i=0}^{D} \binom{n}{i} \leq n^D + 1 \rightarrow m_H(n) \leq n^{d_{VC}} + 1$$

Exercice.6

Quelles sont les fonctions de croissance $m_H(n)$ possibles parmi les suivants :

(a) $1 + n$
(b) $1 + n + \frac{n(n-1)}{2}$
(c) 2^n
(d) $2^{\sqrt{n}}$
(e) $2^{\frac{n}{2}}$
(f) $1 + n + \frac{n(n-1)(n-2)}{6}$

Exercice.7

Montrer que :

$$m_H(2n) \leq m_H(n)^2$$

Et puis, obtenir une borne de généralisation qui contient uniquement $m_H(n)$.

Exercice.8

On suppose que $m_H(n) = n + 1$, donc $d_{VC} = 1$. On possède $n = 100$ échantillons d'entrainement. Utiliser la borne de généralisation pour donner une limite à E_{gen} avec une confiance égale à 90%.

Reprendre la même chose pour $n = 10.000$ échantillons.

Exercice.9

Soit un ensemble d'hypothèse H dont $d_{VC} = 10$. Quelle est la taille de l'échantillon d'entrainement dont on a besoin pour avoir 95% de confiance ?

Exercice.10

Soit $H_1, H_2, \dots, H_k$, k ensembles d'hypothèses avec une dimension VC finie.

Soit $H = H_1 \cup H_2 \cup \dots \cup H_k$ est l'union de ces modèles.

(a) Montrer que $d_{VC}(H) < k(d_{VC} + 1)$
(b) Supposons que $2^l > 2kl^{d_{VC}}$. Montrer que $d_{VC}(H) \leq l$
(c) Puis, montrer que :

$$d_{VC}(H) \leq min(\mathrm{k}(d_{VC} + 1), 7(d_{VC} + k)log_2(d_{VC}k))$$