Concept Général de l'Apprentissage

Exercices

Exercice.1

Exprimer chacune des tâches suivantes dans le cadre de l'apprentissage des données en précisant l'ensemble des inputs X, l'ensemble des outputs Y, la fonction cible f et spécifier l'ensemble des données D à partir duquel on va apprendre.

(a) Diagnostic médical : vous voulez identifier le problème chez un patient qui a des symptômes ainsi qu'un historique médical.
(b) Reconnaissance des chiffres manuscrits par exemple la reconnaissance des codes postales.
(c) Identification des emails (spam ou non).
(d) Prévision de la variation de la charge électrique avec le prix, la température et les jours de la semaine.
(e) Proposer un problème pour lequel il n'y a pas de solution analytique mais pour lequel il y a des données que vous pouvez utiliser pour extraire une solution empirique.

Exercice.2

Supposons qu'on utilise un perceptron pour détecter les emails spams. Chaque email est représenté par la fréquence d'occurrence de ses mots clés. Et l'output est égal à 1 si l'email est spam.

(a) Donner des mots clés qui peuvent avoir un grand poids négatif dans le perceptron.
(b) Donner des mots clés qui peuvent avoir un poids négatif.
(c) Quel est le paramètre dans le perceptron qui affecte directement le nombre des messages qui sont classés comme spams.

Exercice.3

D'après l'équation de la mise à jour des poids :

$$w(t+1) = w(t) + y(t)x(t)$$

(a) Montrer que $y(t)w(t)^T x(t) < 0$ (sachant que $x(t)$ est mal classé par $w(t)$)
(b) Montrer que $y(t)w(t+1)^T x(t) > y(t)w(t)^T x(t)$
(c) En ce qui concerne la classification de $x(t)$, montrer que le déplacement de $w(t)$ à $w(t+1)$ est un déplacement dans la bonne direction.

Exercice.4

Créons notre fonction cible f est notre espace de données D pour savoir comment l'algorithme d'apprentissage « perceptron » fonctionne. Pour bien visualiser le problème prenez $d = 2$, puis tracez une ligne aléatoire dans le plan qui sera considérée votre fonction cible, cette ligne va diviser le plan en deux parties, donnez à la première partie la classe +1 et à l'autre la classe -1. Choisissez aléatoirement de l'espace des données les inputs x_n et évaluez la fonction cible pour chaque x_n pour avoir la sortie correspondante y_n.

Maintenant générez un espace de données de taille 20. Appliquez l'algorithme d'apprentissage « perceptron » sur votre ensemble de données et montrez combien de temps prend-il pour converger et comment votre hypothèse finale g approxime votre fonction cible f.

Exercice.5

Lequel des problèmes suivants est le plus adapté pour la méthodologie d'apprentissage ainsi que pour la méthodologie de design ?

(a) Identification de l'âge à partir duquel un test médical doit être effectué.
(b) Classification des nombres entre premiers et non-premiers.
(c) Détection des fraudes potentielles dans les cartes de crédits.
(d) Détermination du temps qu'un objet utilise pour tomber par terre.
(e) Détermination du cycle optimal des feux rouges dans un rond-point encombré.

Exercice.6

Pour chacune des tâches suivantes, identifier le type d'apprentissage associé (supervisé, non-supervisé, par renforcement) et les données d'entrainement à utilisées. Si une tâche peut être satisfaite par plus d'un seul type, montrer comment et décrire les données d'entrainement pour chaque type :

(a) Recommander un livre à un utilisateur dans une librairie en ligne.
(b) Jouer le jeu de tic tac toc.
(c) Mettre en catégorie des films de différents types.
(d) Apprendre à jouer de la musique.
(e) Identifier le crédit maximal qu'une banque peut accorder à ses clients.

Exercice.7

Si $\mu = 0.9$, quelle est la probabilité qu'un échantillon de 10 billes aura $\nu \leq 0.1$? (utiliser la distribution binomiale, la réponse est un nombre très petit).

Exercice.8

Considérons un échantillon de 10 billes sélectionnées indépendamment à partir d'un pot contenant des billes rouges et vertes. La probabilité d'une bille rouge est μ.

Pour $\mu = 0.05$, $\mu = 0.5$ et $\mu = 0.8$, calculer la probabilité d'avoir aucune bille rouge ($\nu = 0$) dans les cas suivants :

(a) Si on tire un seul échantillon. Calculer la probabilité que $\nu = 0$.
(b) Si on tire 1000 échantillons indépendants. Calculer la probabilité qu'au moins l'un des échantillons possède $\nu = 0$.
(c) Répéter (b) pour 1.000.000 échantillons indépendants.

Exercice.9

La matrice qui présente le coût des différentes erreurs pour le CIA et le supermarché dans l'exemple cité dans le cours du chapitre 1, est nommée matrice de coût ou de risque.

Pour les deux matrices de risque, donner explicitement l'erreur empirique E_{emp} qu'on doit minimiser pour obtenir g. Cette erreur empirique doit pondérer les différents types d'erreurs basées sur la matrice de risque. Prendre $y_n = +1$ et $y_n = -1$ indépendamment.