

1) Introduction
2) RNN
3) CNN

Apprentissage Automatique en Langues

Examen 2018/19 - M2 ATAL - 1h20

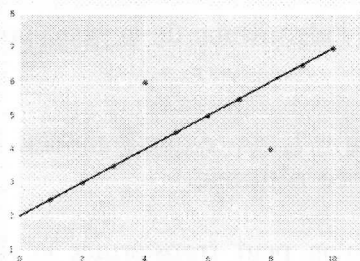
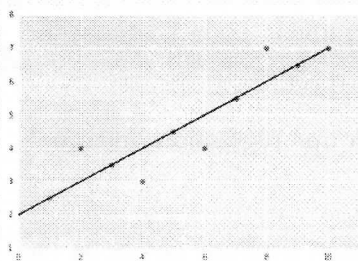
Documents non autorisés

Le barème est donné à titre indicatif et peut subir éventuellement quelques modifications.

Exercice 1 : (6 pts)

1.2 Répondez en quelques lignes aux questions suivantes :

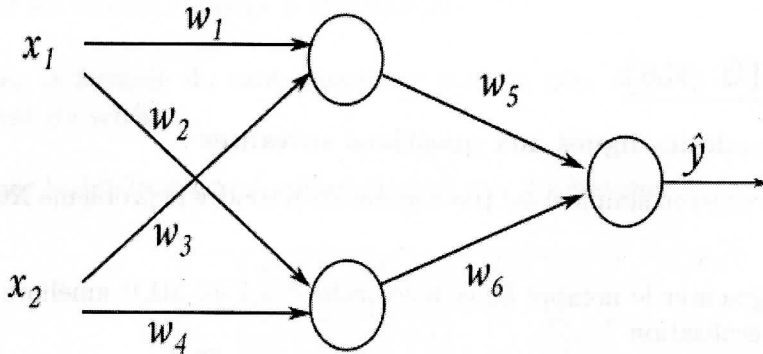
- 1) • Pourquoi un perceptron simple n'est pas capable de résoudre le problème XOR ?
- 1) • Est-ce que l'augmenter le nombre de couche caché dans un MLP améliore sa capacité de généralisation ?
- C'est quoi le risque d'utiliser le corpus de test pour optimiser les hyperparamètres d'un réseau de neurones ?
- Expliquer pourquoi on peut considérer le *dropout* comme un régularisateur dans un réseau de neurones
- C'est quoi le problème qu'on peut avoir avec un *learning rate* très grand ? Comment on peut détecter ça ?
- Quelle est la différence principale entre la régularisation L1 et L2 ? En vous appuyant sur les formules du L1 et L2, expliquez comment elles interviennent dans la modification des poids d'un réseau de neurones ?
- Lequel des deux ensembles de données présentés dans les graphes suivants a l'erreur quadratique moyenne la plus élevée ? Justifiez votre réponse.



- 3) • De quelle taille serait le résultat de la convolution d'une image RVB (3 canaux) de taille 12x12 pixels par un filtre (kernel) de taille 4x4 avec un padding de taille 2 et un pas (stride) de 2 ? — Vous pouvez vous aider d'un schéma explicatif.

Exercice 2 (5 pts)

Soit un ensemble de N triplets $\{(x_{i1}, x_{i2}, y_i)\}$, $1 \leq i \leq N$, l'objectif est de concevoir un modèle pour prédire y_i étant donné les attributs x_{i1} et x_{i2} de chaque entrée x_i . Pour ce faire, nous utilisons le réseau de neurones suivant :



Dans ce modèle w_1, \dots, w_6 sont les paramètres que le modèle doit apprendre. La fonction d'activation (non linéaire) est définie par (q est un hyperparamètre qui à définir avant de lancer l'apprentissage) :

$$f(z) = \begin{cases} (z + |z|)^q & \text{si } z \geq 0 \\ (z - |z|)^q & \text{sinon} \end{cases}$$

L'erreur du réseau est calculée avec :

$$E = \frac{1}{N} \sum_i^N (y - \hat{y})^2$$

3.1 Écrivez le gradient de la fonction d'erreur E par rapport à tous les paramètres w_i en utilisant le théorème de dérivation des fonctions composées.

3.2 Décrivez l'algorithme de la rétro-propagation du gradient pour l'estimation de paramètres.

3.3 Lorsque $q = 1$, est-ce-que le modèle est équivalent à une régression linéaire ? Expliquez votre réponse.

3.4 Est-ce-qu'on peut transformer le réseau de neurones proposé ci-dessus en un réseau équivalent mais plus simple (sans couche cachée) dans le cas où $q = 1$. Justifiez votre réponse.

Exercice 3 (3 pts)

Nous nous intéressons ici à la minimisation de la fonction :

$$f(x, y) = 4x^2 - 4xy + 2y^2$$

a) Calculer

$$\frac{\partial f}{\partial x} \quad (1)$$

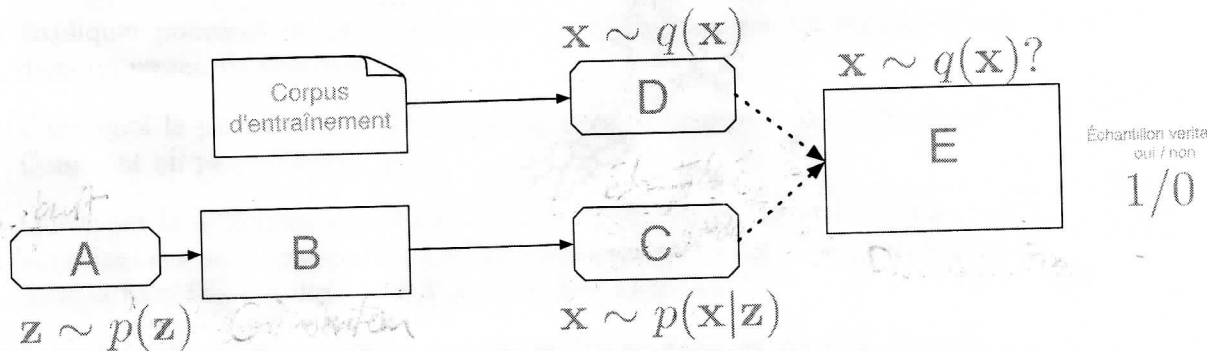
$$\frac{\partial f}{\partial y} \quad (2)$$

b) Appliquer l'algorithme de descente du gradient (deux itérations) avec

$$(x^0, y^0) = (2, 3) \text{ et } \alpha = 0.1$$

Exercice 4 (3 pts)

Voici le schéma de principe d'un GAN



4.1 À quoi correspondent A, B, C, D et E dans le schéma de principe.

4.2 Expliquez le fonctionnement d'un GAN.