

Apprentissage Automatique en Langues

Examen 2019/20 - M2 ATAL - 1h30

Documents non autorisés

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

Exercice 1

- Quel risque y a-t-il à utiliser le corpus de test pour l'optimisation des hyper-paramètres d'un réseau de neurones ?
 - Expliquer pourquoi on peut considérer le *dropout* comme une technique de régularisation d'un réseau de neurones.
 - Quel problème peut-on avoir avec un *learning rate* très grand ? Comment peut-on détecter cela ?
 - Les modèles de langue à base de réseaux de neurones sont des modèles paramétriques, contrairement aux modèles n-gram à repli. Qu'est-ce que cela signifie ?
 - 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 deux itérations de l'algorithme de descente du gradient avec

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

Exercice 2

Supposons que nous voulions créer un réseau de neurones avec une seule couche cachée et la fonction d'activation *logistique*. Le réseau est utilisé pour résoudre un problème de classification multi-classes.

$$h = \sigma(W^{(1)}x + b^{(1)})$$

$$\hat{y} = \text{softmax}(W^{(2)}h + b^{(2)}).$$

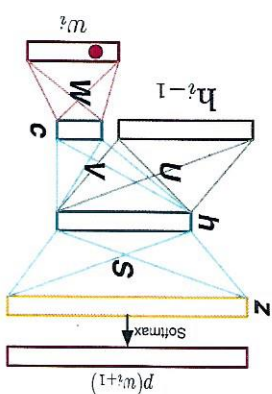
Le réseau est entraîné en utilisant la *cross entropie* comme fonction d'erreur.

$$C(y, \hat{y}) = - \sum_i y_i \log \hat{y}_i$$

- 3.1. Calculez le nombre de paramètres du modèle dans le cas où l'entrée est de dimension D , le nombre de classe est égale à k et le nombre de couche caché H .

- 3.2. Dessiner le réseau de neurones décrit dans la question 3.1 (placer D , H et k sur la figure).

Exercice 3: Réseaux récurrents



- 3.1. Décrivez la figure en détail. Expliquez particulièrement le rôle de h_{t-1} .

- 3.2. Expliquez très simplement la différence entre un réseau récurrent simple et une version plus élaborée comme un LSTM ou un GRU.

- 3.3. Quel est l'intérêt d'un bi-LSTM par rapport à un LSTM?

- 3.4. décrivez le fonctionnement d'un mécanisme d'attention appliqué à un bi-LSTM (n'hésitez pas à faire un schéma).

Exercice 4 : Réseaux convolutionnels

- 3.1. Faites un schéma pour décrire un réseau convolutionnel 2D.
- 3.2. Qu'est-ce que le "padding" ?
- 3.3. À quoi servent les couches de *pooling* ?

Exercice 5 : GANs

- 3.1. Expliquez à l'aide d'un schéma et d'une description, le fonctionnement d'un GAN.
- 3.2. Quelle amélioration apporte les GANs conditionnels.

Exercice 6 : HAC

On considère 6 observations A, B, C, D, E, F décrites par 2 attributs quantitatives appelées V1 et V2 donnés dans le tableau suivant :

	V1	V2
A	0	0
B	-1	2
C	3	2
D	6	0
E	4	1
F	5	3

- 3.1. Représenter ces observations sur un dessin.
- 3.2. Calculer la distance euclidienne au carré (afin de ne pas avoir la racine carrée à calculer).
- 3.3. En utilisant le saut maximum, effectuer une classification hiérarchique ascendante de ces 6 observations.
- 3.4. Donner l'arbre hiérarchique.

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$