Examen M2 AMIS
Réseaux de neurones

Jérémie Cabessa, 14 décembre 2023

Durée: 2h. Aucune documentation autorisée. Barème indicatif.

Exercice 1 (5 pts)

Soit $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$ un dataset, où $\mathbf{X}$ est une matrice de features à laquelle on a concaténé une première colonne de 1 et $\mathbf{y}$ est un vecteur de réponses. La *régression linéaire* et la *régression Ridge* correspondent aux problèmes de minimisation (1) et (2) suivants:

$$\min_{\boldsymbol{\beta}} \quad \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|^2 \tag{1}$$

$$\min_{\boldsymbol{\beta}} \quad \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|^2 + \lambda \|\boldsymbol{\beta}\|^2 \tag{2}$$

où λ est un paramètre de régularisation.

1. Donnez le développement qui permet de retrouver la solution $\hat{\boldsymbol{\beta}}$ de la régression linéaire (1).
2. Donnez le développement qui permet de retrouver la solution $\hat{\boldsymbol{\beta}}$ de la régression Ridge (2).

Exercice 2 (8 pt)

1. On considère le dataset formé de trois points rouges, $(0, 2)$, $(-2, 1)$ et $(0, 0)$, et deux points bleus, $(3, 1)$ et $(-2, 1)$, représenté en Figure 1a (page suivante). Donnez la description d'un perceptron capable de classifier correctement les points de ce dataset.

   Indication: *Il est possible de "deviner" les poids de votre perceptron, sans avoir besoin de l'entraîner.*

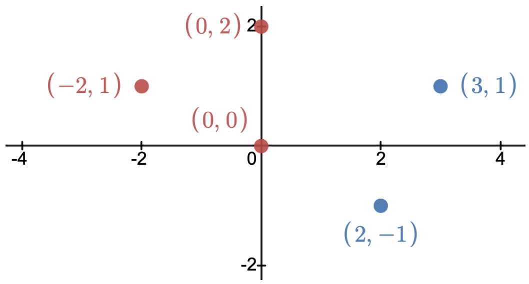
2. On considère maintenant un autre dataset formé avec un point bleu de plus, représenté en Figure 1b. Donnez la description de deux perceptrons tels que:

   - le premier est capable de séparer correctement le point bleu du haut $(-3, 3)$ des trois points rouges $(0, 2)$, $(-2, 1)$ et $(0, 0)$.
   - le second est capable de séparer correctement les deux points bleus du bas $(3, 1)$ et $(-2, 1)$ des trois points rouges $(0, 2)$, $(-2, 1)$ et $(0, 0)$.

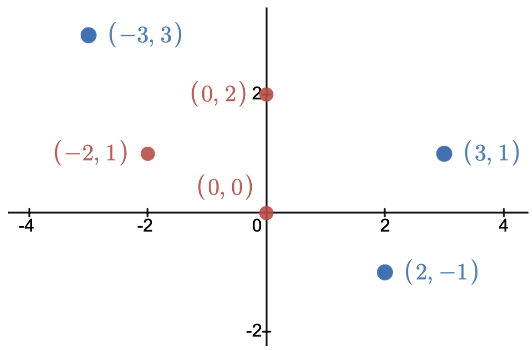
   Indication: *Dans ce cas également, pas besoin de recourir à un entraînement.*

3. Donnez maintenant la description d'un perceptron multicouche (MLP) avec une seule couche cachée, comme illustré en Figure 1c, capable de classifier correctement tous les points de ce dataset. Justifiez.

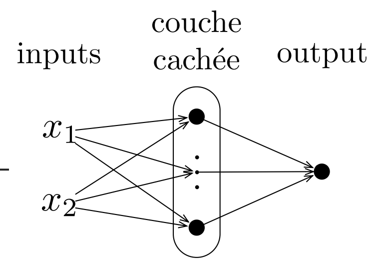
   Indication: *Si vous ne voyez pas l'idée, ne perdez pas trop de temps sur cet exercice...*



(a) Premier dataset.



(b) Second dataset.



(c) MLP à une couche.

Figure 1

Exercice 3 (10 pts)

On considère un réseau de neurones à deux couches dont le graphe computationnel est illustré en Figure 2. Pour ce réseau, il n'y a pas de biais et les fonctions d'activation sont l'identité (i.e., $\sigma = id$), ce qui facilite leurs dérivées.

1. Donnez les équations qui décrivent la dynamique ou "forward pass" de ce réseau.

2. Donnez une seule équation qui exprime $\mathbf{a}^{[2]}$ en fonction de $\mathbf{a}^{[0]}$, $\mathbf{W}^{[1]}$ et $\mathbf{W}^{[2]}$.

3. Soit la fonction de loss suivante

$$\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}; \boldsymbol{\Theta}) = \sum_{i=1}^{p} (\hat{y}_i - y_i)^2 = \|\hat{\mathbf{y}} - \mathbf{y}\|^2 = (\hat{\mathbf{y}} - \mathbf{y})^T (\hat{\mathbf{y}} - \mathbf{y}) = \hat{\mathbf{y}}^T \hat{\mathbf{y}} - 2\hat{\mathbf{y}}^T \mathbf{y} + \mathbf{y}^T \mathbf{y}\,.$$

   Sachant que $\hat{\mathbf{y}} = \mathbf{a}^{[2]}$, donnez une expression pour le gradient $\nabla_{\mathbf{a}^{[2]}} \mathcal{L}\left(\mathbf{a}^{[2]}, \mathbf{y}; \boldsymbol{\Theta}\right)$. Votre expression dépendra de $\mathbf{a}^{[0]}$, $\mathbf{W}^{[1]}$, $\mathbf{W}^{[2]}$ et $\mathbf{y}$. Utilisez votre caractérisation de $\mathbf{a}^{[2]}$ obtenue au point précédent.

4. Abbrégeons $\mathcal{L}\left(\mathbf{a}^{[2]}, \mathbf{y}; \boldsymbol{\Theta}\right)$ par $\mathcal{L}(\boldsymbol{\Theta})$. Calculez maintenant le gradient $\boldsymbol{\delta}^{[2]} := \nabla_{\mathbf{z}^{[2]}} \mathcal{L}(\boldsymbol{\Theta})$. Utilisez l'expression $\nabla_{\mathbf{a}^{[2]}} \mathcal{L}(\boldsymbol{\Theta})$ que vous avez calculée précédemment. Rappelez-vous que σ est l'identité pour ce réseau.

5. Donnez une expression pour le gradient $\nabla_{\mathbf{W}^{[2]}} \mathcal{L}(\boldsymbol{\Theta})$. Utilisez votre expression précédente.

6. Donnez une expression pour le gradient $\nabla_{\mathbf{W}^{[1]}} \mathcal{L}(\boldsymbol{\Theta})$. Si cela vous arrange, vous pouvez supposer que $\boldsymbol{\delta}^{[2]} := \nabla_{\mathbf{z}^{[2]}} \mathcal{L}(\boldsymbol{\Theta})$ a été calculé et utiliser cette expression telle quelle.

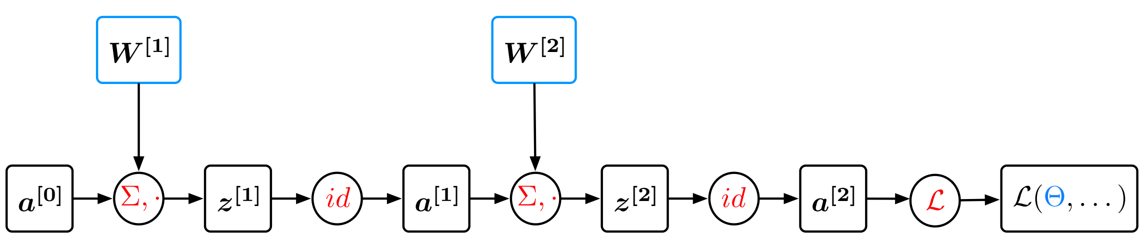


Figure 2: Graphe computationnel d'un réseau de neurones.

Exercice 4 (5 pt)

- Expliquez en quelques mots le fonctionnement d'au autoencodeur (AE).
- Expliquez également en quelques mots le fonctionnement d'au autoencodeur variationnel (VAE).

Exercice 5 (6 pt)

Le mécanisme d'attention s'exprime comme suit

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

et peut être représenté comme dans la Figure 3, où $\mathbf{W^Q}$, $\mathbf{W^K}$ et $\mathbf{W^V}$ sont des matrices de poids de couches linéaires.

- Expliquez ce mécanisme. Faites référence à la Figure 3. Quelle est l'utilité de ce mécanisme?

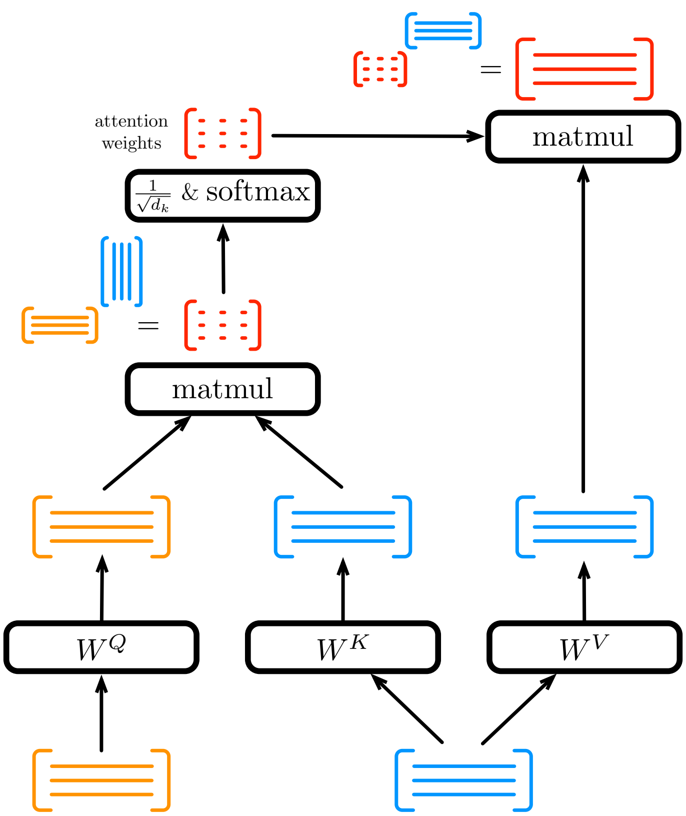


Figure 3: Mécanisme d'attention.