

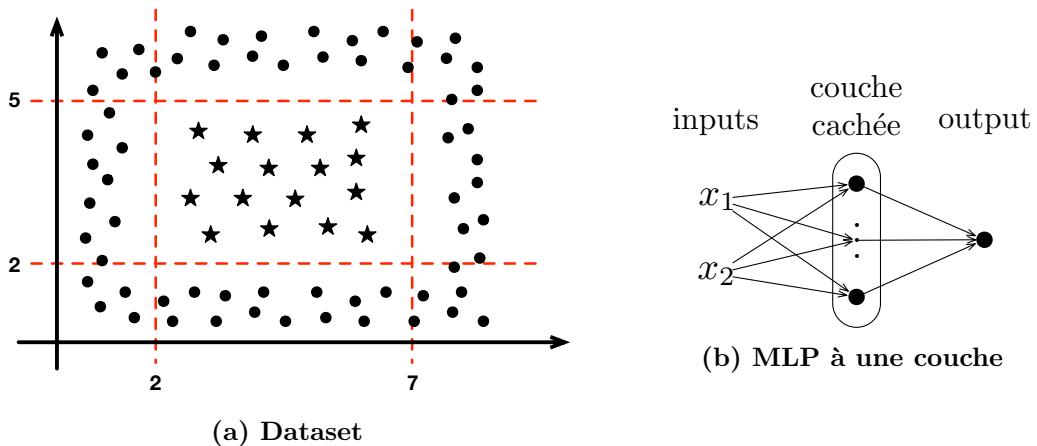
Réseaux de neurones: Examen M2 AMIS

Jérémie Cabessa, 20 décembre 2024

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

Exercice 1 (8 pt)

On considère le dataset formé d'étoiles et de ronds représenté dans la Figure (a) ci-dessous. Donnez la description d'un perceptron multicouche (MLP) avec une seule couche cachée, comme illustré en Figure (b) capable de classifier correctement tous les points de ce dataset.



Exercice 2 (8 pts)

On considère un réseau de neurones à une seule couche dont le graphe computationnel est illustré en Figure 2 (page suivante). Pour ce réseau, la fonction d'activation est l'identité (i.e., $\sigma(x) = x$), ce qui facilite le calcul de sa dérivée (i.e., $\sigma'(x) = 1$).

1. Donnez les équations qui décrivent la dynamique ou “forward pass” de ce réseau.
2. Soit la fonction de loss suivante

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

Sachant que $\hat{\mathbf{y}} = \mathbf{a}^{[1]}$, donnez les expressions qui permettent de calculer les deux gradients suivants:

$$\nabla_{\mathbf{b}^{[1]}} \mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}; \Theta) \quad \text{et} \quad \nabla_{\mathbf{W}^{[1]}} \mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}; \Theta)$$

3. À partir de ces gradients, écrivez en pseudocode un algorithme de descente de gradient permettant d'entraîner ce réseau sur le dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i) : i = 1, \dots, N\}$. Pour simplifier, les gradients seront mis à jour après le processing de chaque data (pas de processing par batch).

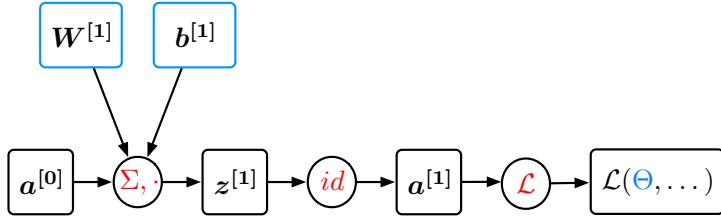


Figure 2: Graphe computationnel d'un réseau de neurones à une seule couche.

Exercice 3 (8 pt)

Soit un réseau convolutionnel (CNN) conçu pour classifier des images en niveaux de gris de taille $28 \times 28 \times 1$ (par exemple, le dataset MNIST). L'architecture du CNN est la suivante :

- **Première couche convolutionnelle (Conv 1)** : Applique 4 filtres de taille $3 \times 3 \times 1$ avec un *pas* (stride) de 1 et un *remplissage* (padding) de 1, suivie d'une activation ReLU.
- **Première couche de max-pooling (MaxPool 1)** : Applique une opération de max-pooling de taille 2×2 avec un *pas* (stride) de 2.
- **Deuxième couche convolutionnelle (Conv 2)**: Applique 8 filtres de taille $4 \times 4 \times k$ (valeur de k à déduire) avec un *pas* (stride) de 4 et un *remplissage* (padding) de 1, suivie d'une activation ReLU.
- **Deuxième couche de max-pooling (MaxPool 2)** : Applique une opération de max-pooling de taille 2×2 avec un *pas* (stride) de 2.
- **Couche totalement connectée (FC 1)**: Connecte la sortie de la couche précédente à une couche dense formée de 128 neurones et d'une activation ReLU.
- **Couche de sortie (FC 2)**: Connecte la sortie de la couche précédente à une couche dense formée de 10 neurones et d'une activation softmax.

1. Faites un schéma de ce CNN.
2. Calculez les dimensions de sortie après la première couche convolutionnelle (Conv 1) ainsi que le nombre de paramètres entraînables de cette couche.
3. Calculez les dimensions de sortie après la première couche de max-pooling (MaxPool 1) ainsi que le nombre de paramètres entraînables de cette couche.
4. Calculez les dimensions de sortie après la deuxième couche convolutionnelle (Conv 2) ainsi que le nombre de paramètres entraînables de cette couche.
5. Calculez les dimensions de sortie après la deuxième couche de max-pooling (MaxPool 2) ainsi que le nombre de paramètres entraînables de cette couche.
6. Calculez la dimension de sortie après chaque couche totalement connectée (FC 1 et FC 2) ainsi que le nombre de paramètres entraînables de ces couches.
7. Estimez le nombre de multiplications effectuées par la première couche convolutionnelle (Conv 1).

Exercice 4 (8 pt)

- Expliquez le fonctionnement d'un autoencodeur variationnel (VAE).
- Indiquez la différence principale avec un autoencodeur (classique) (AE). Quels sont les points d'améliorations apportés par le VAE par rapport au AE?

Exercice 5 (8 pt)

Le mécanisme d'auto-attention (self attention) peut être représenté comme dans la Figure 3.

- Expliquez ce mécanisme (brièvement).
- Donnez le formalisme mathématique qui correspond à ce mécanisme.
- Donnez les équations de la dynamique (forward pass) d'un réseau de neurones composé d'un mécanisme d'attention (comme en Figure 3) suivi d'une couche cachée.

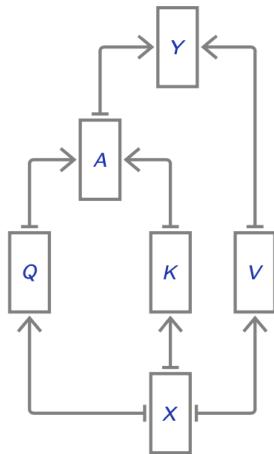


Figure 3: Mécanisme d'auto-attention.