

Corrige Type

Questions de compréhension

1. Elle fait appel à plusieurs disciplines comme les mathématiques, l'informatique, la biologie, le traitement de signal, **1pt**

2. **1pt**

Reconnaissance faciale : Identifier une personne à partir d'une image ou d'une vidéo.

Analyse médicale : Détection de tumeurs ou de cellules anormales à partir d'images médicales.

3. **1pt**

Mettre tous les attributs numériques à la même échelle, évitant qu'un attribut à grande variance domine les autres.

Accélérer la convergence des algorithmes basés sur des calculs de distances.

4. **1pt**

Bayésien

Avantages : Facilité d'interprétation, efficace pour les données avec dépendances claires.

Inconvénients : Hypothèse de distribution préalable, moins performant avec des dépendances complexes.

CNN

Avantages : Puissant pour les données complexes (images, vidéos).

Inconvénients : Requiert beaucoup de données et de puissance de calcul, difficile à interpréter.

Exercice 1

1. $\text{input_dim} = 64 \times 64 \times 3 = 12\,288$ **1pt**

2. RNA de type PMC

3. Non. 3 neurones de sortie dans la couche de sortie.

4. Base de validation. Valider les hyperparamètres du classificateur. Taille de la base de validation = $50\,000 \times 20\% = 10\,000$ images.

5. PMC constitué d'une couche d'entrée (12 288 neurones) + 2 couches cachées constituées de 256 neurones chacune, une couche cachée de 3 neurones, fonction activation Relu et Softmax, Bias, Connections.

Exercice 2

1. $P(\text{Spam}) = 2/4 = 0.5$ et $P(\text{Non-Spam}) = 2/4 = 0.5$ **1 pt**

2. **1 pt**

Pour Spam, $P(\text{Has_Free} = \text{Non} \mid \text{Spam}) = 0/2 = 0$ et $P(\text{Has_Free} = \text{Oui} \mid \text{Spam}) = 2/2 = 1$.

Pour Non-Spam : $P(\text{Has_Free} = \text{Non} \mid \text{Non-Spam}) = 2/2 = 1$ et $P(\text{Has_Free} = \text{Oui} \mid \text{Non-Spam}) = 0/2 = 0$.

3. **1 pt**

$$\mu_{\text{Spam}} = \frac{120 + 150}{2} = 135$$

$$\sigma_{\text{Spam}}^2 = \frac{(120 - 135)^2 + (150 - 135)^2}{2} = \frac{225 + 225}{2} = 225$$

1 pt

$$\mu_{\text{Non-Spam}} = \frac{50 + 60}{2} = 55$$

$$\sigma_{\text{Non-Spam}}^2 = \frac{(50 - 55)^2 + (60 - 55)^2}{2} = \frac{25 + 25}{2} = 25$$

4. **1 pt**

Pour Spam ($\mu=135$, $\sigma^2=225$)

$$P(\text{Word_Count} = 140 \mid \text{Spam}) = 0.0265 \times 0.9459 \approx 0.0251$$

Pour Non-Spam ($\mu=55$, $\sigma^2=25$)

$$P(\text{Word_Count} = 140 \mid \text{Non-Spam}) \approx 0$$

5. Calcul les vraisemblances des classes 1 pt

Pour Spam

$$P(\text{Spam}) \times P(\text{Has_Free} = \text{Non} \mid \text{Spam}) \times P(\text{Word_Count} = 140 \mid \text{Spam}) \\ = 0.5 \times 0 \times 0.0251 = 0$$

Pour Non-Spam

$$P(\text{Non-Spam}) \times P(\text{Has_Free} = \text{Non} \mid \text{Non-Spam}) \times P(\text{Word_Count} = 140 \mid \text{Non-Spam}) \\ = 0.5 \times 1 \times 0 = 0$$

6. Le modèle échoue, car les deux probabilités sont nulles. **0.5 pt**

7. Calcul des probabilités conditionnelles pour l'attribut Has_Free en appliquant la correction Laplacienne $k=1$
1 pt

$$\text{Pour Spam, } P(\text{Has_Free} = \text{Non} \mid \text{Spam}) = (0 + 1/2) / (2 + 1) = 0,5 / 3 = 0.16$$

$$\text{Pour Non-Spam, } P(\text{Has_Free} = \text{Non} \mid \text{Non-Spam}) = (2 + 1/2) / (2 + 1) = 2.5 / 3 = 0.83$$

Calcul des vraisemblances des classes après application de la correction Laplacienne.

Pour Spam 1 pt

$$0.5 \times 0.16 \times 0.0251 = 0.002008$$

Pour Non-Spam

$$0.5 \times 0.83 \times 0 = 0$$

8. E5 est un spam. **0.5 pt**