

Corrigé type interrogation en apprentissage automatique

Exercice 01 (10 pts).

- 1) Définir c'est quoi un apprentissage automatique ?
- 2) Pouvez-vous donner quatre problèmes qui peuvent être traités efficacement par des techniques d'apprentissage automatique.
- 3) C'est quoi le rôle de l'ensemble d'entraînement ?
- 4) Quelles sont les deux tâches supervisées les plus courantes ?
- 5) Quel type d'algorithme d'apprentissage utilisez-vous pour segmenter vos clients en plusieurs groupes ?
- 6) Le problème de détection du spam peut-il être classé comme un problème d'apprentissage supervisé ou non supervisé ?
- 7) Que se passe-t-il si votre modèle fonctionne bien sur les données d'apprentissage, mais se généralise mal aux nouvelles instances ? Pouvez-vous donner une solution possible ?
- 8) Quel est le rôle de la fonction d'erreur pour un problème de classification ?
- 9) Quel est l'avantage de la régression logistique par rapport au classificateur de Bayes ?
- 10) A quoi consiste l'hypothèse de Bayes naïve ?

Solution.

- 1) **L'apprentissage automatique (machine Learning) est une discipline émergente pour développer une intelligence artificielle capable :**
 - **d'analyser automatiquement des données ;**
 - **de détecter des motifs et associations de données ;**
 - **d'utiliser des motifs pour la prédiction de données ;**
 - **de prendre des décisions automatiques ;**
- 2) Les problèmes qui peuvent être traités par des techniques d'apprentissage automatique : **Par exemple le forage de données, la reconnaissance des formes, la vision artificielle, la classification de documents textuels, ... etc.**
- 3) Le rôle de l'ensemble d'entraînement : **L'ensemble d'entraînement est utilisé pour faire le système apprendre des exemples.**
- 4) Les deux tâches supervisées les plus courantes : **Régression et Classification**
- 5) Quel type d'algorithme d'apprentissage utilisez-vous pour segmenter vos clients en plusieurs groupes : **Clustering.**
- 6) Le problème de détection du spam peut être classé comme **un problème d'apprentissage supervisé.**
- 7) Que se passe-t-il si votre modèle fonctionne bien sur les données d'apprentissage, mais se généralise mal aux nouvelles instances ? Pouvez-vous donner une solution possible ? **Problème overfitting. Solution : utiliser plus de données pour l'apprentissage, utiliser d'autres modèles, utiliser la technique de régularisation ... etc.**
- 8) Quel est le rôle de la fonction d'erreur pour un problème de classification ? **Dans le contexte de l'apprentissage automatique, la fonction d'erreur permet d'évaluer la qualité d'une solution, par conséquent comparer des solutions.**
- 9) L'avantage de la régression logistique par rapport au classificateur de Bayes est qu'elle n'est pas limitée par l'hypothèse de Bayes naïve donc elle peut classer efficacement les données aberrantes (Outliers) qui perturbent la frontière de décision.
- 10) L'hypothèse de Bayes naïve (HBN) consiste à dire que x est une conjonction de valeurs d'attributs et que ces attributs sont des variables aléatoires indépendantes.

Exercice 02 (10 pts).

Soient les informations des symptômes et du diagnostic des patients suivants :

Douleurs	Fatigue	Mal de tête	Fièvre	Grippe ?
Oui	Non	Doux	Oui	Non
Oui	Oui	Néant	Non	Oui
Oui	Non	Fort	Oui	Oui
Non	Oui	Doux	Oui	Oui
Non	Non	Néant	Non	Non
Non	Oui	Fort	Oui	Oui
Non	Oui	Fort	Non	Non
Oui	Oui	Doux	Oui	Oui

En utilisant la classification de Bayes naïve, prédire l'état d'un patient ayant les symptômes suivants :

Douleurs	Fatigue	Mal de tête	Fièvre	Grippe ?
Non	Non	Doux	Oui	?

Solution.

Soit la donnée $x = (\text{Douleurs} = \text{Non}, \text{Fatigue} = \text{Non}, \text{Mal de tête} = \text{Doux}, \text{Fièvre} = \text{Oui})$, on peut prédire la classe de x en prenant **le maximum de probabilité a posteriori** :

$$HBN = \operatorname{argmax}_{h \in \{\text{Oui}, \text{Non}\}} p(\text{Grippe} = h | x)$$

Si on utilise la règle de Bayes, on aura :

$$p(\text{Grippe} = h | x) = \frac{p(x | \text{Grippe} = h) * p(\text{Grippe} = h)}{p(x)}$$

$$\frac{p(x | \text{Grippe} = h) * p(\text{Grippe} = h)}{\sum_{a \in \{\text{Oui}, \text{Non}\}} p(x | \text{Grippe} = a) p(\text{Grippe} = a)}$$

Application numérique :

$$P(\text{Grippe} = \text{Oui}) = \frac{5}{8}$$

$$P(\text{Grippe} = \text{Non}) = \frac{3}{8}$$

$$P(\text{Douleurs} = \text{Non} | \text{Oui}) = \frac{2}{5}$$

$$P(\text{Fatigue} = \text{Non} | \text{Oui}) = \frac{1}{5}$$

$$P(\text{Douleurs} = \text{Non} | \text{Non}) = \frac{2}{3}$$

$$P(\text{Fatigue} = \text{Non} | \text{Non}) = \frac{2}{3}$$

$$P(\text{Mal de tête} = \text{Doux} | \text{Oui}) = \frac{2}{5}$$

$$P(\text{Fièvre} = \text{Oui} | \text{Oui}) = \frac{4}{5}$$

$$P(\text{Mal de tête} = \text{Doux} | \text{Non}) = \frac{1}{3}$$

$$P(\text{Fièvre} = \text{Oui} | \text{Non}) = \frac{1}{3}$$

$$\begin{aligned}
p(x|Grippe = Oui) &= P(Douleurs = Non|Oui) \times P(Fatigue = Non|Oui) \\
&\times P(Mal de tête = Doux|Oui) \times P(Fièvre = Oui |Oui) \\
&= \frac{2}{5} \times \frac{1}{5} \times \frac{2}{5} \times \frac{4}{5} = \frac{16}{625}
\end{aligned}$$

$$\begin{aligned}
p(x|Grippe = Non) &= P(Douleurs = Non|Non) \times P(Fatigue = Non|Non) \\
&\times P(Mal de tête = Doux|Non) \times P(Fièvre = Oui |Non) \\
&= \frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{4}{81}
\end{aligned}$$

Donc, on aura :

$$\begin{aligned}
p(Grippe = Oui|x) &= \frac{\frac{16}{625} \times \frac{5}{8}}{\frac{16}{625} \times \frac{5}{8} + \frac{4}{81} \times \frac{3}{8}} \approx 0.4635 \\
p(Grippe = Non|x) &= \frac{\frac{4}{81} \times \frac{3}{8}}{\frac{16}{625} \times \frac{5}{8} + \frac{4}{81} \times \frac{3}{8}} \approx 0.5365
\end{aligned}$$

Conclusion :

Le résultat prédit pour la donnée $x = (\text{Douleurs} = \text{Non}, \text{Fatigue} = \text{Non}, \text{Mal de tête} = \text{Doux}, \text{Fièvre} = \text{Oui})$ sera ***Grippe = Non***.