
Corrigé type de l'examen

Questions de cours (4 pts)

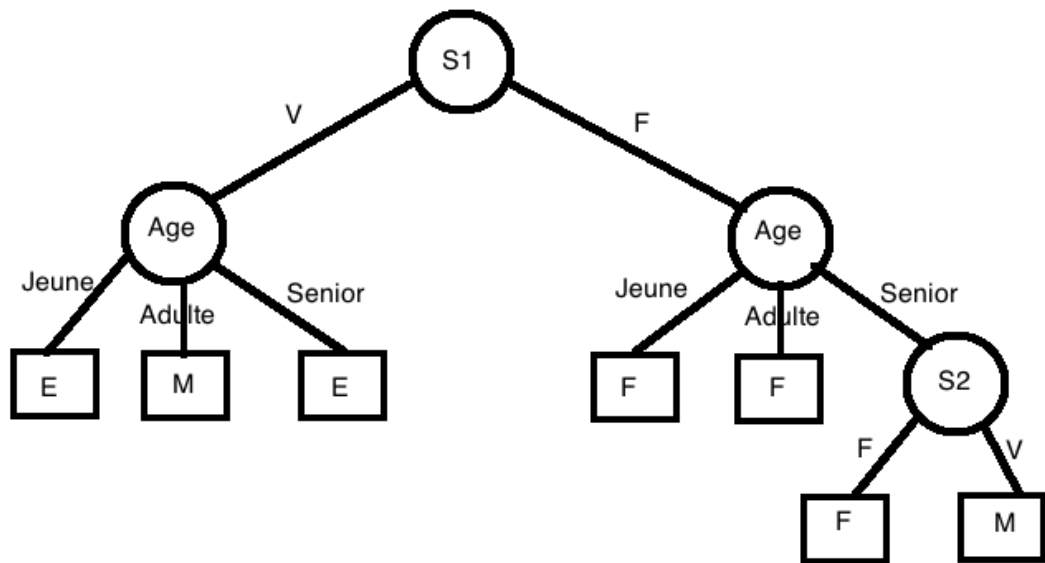
1. En apprentissage supervisé, la différence fondamentale entre classification et régression est que la classification est utilisée pour prédire les classes discrètes (catégorielles) tandis que la régression est utilisée pour prédire les classes continues (numériques).
2. Le phénomène de sur-apprentissage apparaît lorsqu'on obtient des modèles qui reconnaissent parfaitement les exemples utilisés lors de l'apprentissage et échouent face aux nouveaux exemples.
3. La stratégie utilisée pour éviter le sur-apprentissage lors de l'apprentissage d'arbres de décision est l'élagage.
4. Les méthodes d'évaluation des techniques d'apprentissage sont diverses telles que la validation croisée, le bootstrap, leave one out, ...etc. (voir le cours pour les principes)

Exercice 1 Motifs fréquents (10.5 pts : 4 + 2.5 + 2.5 + 1.5)

1. Les motifs fréquents :
F1 = { Lait, Couches, Pain, Beurre, Jus, Sucre }
F2 = { Lait Couches, Lait Pain, Lait Beurre, Couches Pain, Couches Beurre, Couches Jus, Couches Sucre, Pain Beurre, Jus Sucre }
F3 = { Lait Couches Pain, Lait Couches Beurre, Lait Pain Beurre, Couches Pain Beurre }
F4 = { Lait Couches Pain Beurre }
Motifs fréquents = { F1 \cup F2 \cup F3 \cup F4 }
2. Motifs fréquents fermés = { Lait, Couches, Jus, Sucre, Lait Couches, Couches Jus, Couches Sucre, Pain Beurre, Jus Sucre, Lait Pain Beurre, Couches Pain Beurre, Lait Couches Pain Beurre }
3. Motifs fréquents maximaux = { Couches Jus, Couches Sucre, Jus Sucre, Lait Couches Pain Beurre }
4. Les règles solides d'association de type $A, B \Rightarrow C$ sont :
Lait Beurre \Rightarrow Pain
Lait Pain \Rightarrow Beurre
Couches Beurre \Rightarrow Pain
Couches Pain \Rightarrow Beurre

Exercice 2 Classification (5.5 pts : 4 + 0.5 + 1)

1. L'arbre souhaité en utilisant l'algorithme ID3 :



2. Le risque du patient ayant les attributs (Jeune,V,F) selon l'arbre construit est **E**.

3. Le risque du même patient en utilisant la classification bayésienne naïve :

$$X = (\text{Jeune}, V, F)$$

$$\begin{aligned} P(X/\text{Risque}=F) &= P(\text{Age}=\text{Jeune}/\text{Risque}=F) \times P(S1=V/\text{Risque}=F) \times P(S2=F/\text{Risque}=F) \\ &\quad \times P(P(\text{Risque}=F)) \\ &= 0.5 \times 0 \times \dots = 0 \end{aligned}$$

$$\begin{aligned} P(X/\text{Risque}=E) &= P(\text{Age}=\text{Jeune}/\text{Risque}=E) \times P(S1=V/\text{Risque}=E) \times P(S2=F/\text{Risque}=E) \\ &\quad \times P(P(\text{Risque}=E)) \\ &= 1/3 \times 3/3 \times 1/3 \times 3/10 = 9/90 = 0.1 \end{aligned}$$

$$\begin{aligned} P(X/\text{Risque}=M) &= P(\text{Age}=\text{Jeune}/\text{Risque}=M) \times P(S1=V/\text{Risque}=M) \times P(S2=F/\text{Risque}=M) \\ &\quad \times P(P(\text{Risque}=M)) \\ &= 0 \times \dots = 0 \end{aligned}$$

Le risque est celui de la classe de probabilité max c-à-d **E**.

Dr A.Djeffal