

Exercice 1 : (2 points :0.5+0.5+0.5+0.5)

1-Quelle est la différence entre la classification et la régression ?

La variable de sortie dans la classification est discrète et finie tandis que dans la régression est continue.

2-En classification, quelles sont les erreurs les plus coûteuses les FP ou bien les FN ? Justifier.

FN est plus couteuse, car dans FN, le classifieur déclare un patient par exemple en bonne santé alors son cas est positive

3-Quelle est la différence entre la régression et l'interpolation ?

La régression sert pour la prédiction et l'interpolation sert pour interpoler entre les points

4- Quelle est la différence entre l'étape de l'apprentissage et le test ?

L'apprentissage permet de créer un modèle, alors le test sert pour valider ou tester le modèle.

Exercice 2 : (6points :3+1+1+1)

Soit le tableau suivant :

| | | | | | |
|---|----|---|---|---|---|
| X | -5 | 0 | 1 | 2 | 3 |
| Y | -1 | 0 | 4 | 6 | 8 |

1- Trouver la droite de régression (par l'approche matricielle).

$$X = \begin{pmatrix} 1 & -5 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \text{ et } X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -5 & 0 & 1 & 2 & 3 \end{pmatrix}$$

$$X\vec{\beta} = Y \leftrightarrow X^T X \vec{\beta} = X^T Y \text{ avec } \vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 5 & 1 \\ 1 & 39 \end{pmatrix}, X^T Y = \begin{pmatrix} 17 \\ 45 \end{pmatrix}$$

$$\begin{pmatrix} 5 & 1 \\ 1 & 39 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 17 \\ 45 \end{pmatrix} \rightarrow \beta_0 = 3.18 \text{ et } \beta_1 = 1.07$$

Donc l'équation de la droite est $\hat{y} = 1.07x + 3.18$

2- Calculer les erreurs quadratiques moyennes suivantes : $\left\{ \begin{matrix} err = \sum_{i=1}^N (y - \hat{y})^2 / 10 \end{matrix} \right.$

| y | \hat{y} | $y - \hat{y}$ | $(y - \hat{y})^2$ |
|----|-----------|---------------|-------------------|
| -1 | -2,17 | 1,17 | 1,38 |
| 0 | 3,18 | -3,18 | 10,14 |
| 4 | 4,25 | -0,25 | 0,06 |
| 6 | 5,32 | 0,67 | 0,44 |
| 8 | 6,40 | 1,59 | 2,55 |

Err=1.45

3- Calculer les indices de corrélation r_{xy} et $r_{x\hat{y}}$.

$$\sigma_x = 2.78, \sigma_y = 3.44, \sigma_{\hat{y}} = 2.98, \sigma_{x\hat{y}} = 8.32, r_{xy} = 0.82 \text{ et } r_{x\hat{y}} \approx 1.$$

4- Calculer la droite de régression par l'approche de Mayer ($\beta_1 = \sigma_{xy}/\sigma_x^2$ et $\beta_0 = \bar{y} - \beta_1 \bar{x}$)

$$\beta_1 = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{8.32}{2.72^2} = 1.07$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 3.4 - 1.07 \times 0.2 = 3.18$$

Exercice 3 (6points :2+2+2+2)

| | | Cas prédits | |
|-----------|---------------|-------------|--------|
| | | Test + | Test - |
| Cas réels | Cancer | 950 | 50 |
| | Pas de Cancer | 60 | 940 |

- 1- Le nombre total de cas réels : 950 +50(+) et 60+940 (-) → 2000
- 2- Le nombre total de cas prédits : 950+60 (+) et 50+940 (-) → 2000
- 3- Donner les valeurs de TP, FP, FN, et TN respectivement ?
TP=950 FN=50 FP=60 TN=940
- 4- Calculer les métriques de performance suivantes : accuracy, precision, sensitivity, et specificity.
Accuracy = 0.9450
precision = 0.9406
Sensitivity = 0.9500
specificity = 0.9400

Exercice 4 : (6points:4+2)

1-

- Tableau de vraisemblance Ciel :

| jouer | Non | Oui | |
|-------------------|-----|-----|-----|
| Ensoleillé | 3/4 | | 3/8 |
| Pluie | 1/4 | 2/4 | 3/8 |
| Couvert | 0 | 2/4 | 2/8 |
| Total | 4 | 4 | |
| | 4/8 | 4/8 | |

- Tableau de vraisemblance Temp :

| jouer | Non | Oui | |
|----------------|-----|-----|-----|
| Chaude | 2/4 | 1/4 | 3/8 |
| Tiède | 1/4 | 1/4 | 2/8 |
| Fraiche | 1/4 | 2/4 | 3/8 |
| Total | 4 | 4 | |
| | 4/8 | 4/8 | |

- Tableau de vraisemblance Humidité :

| jouer | Non | Oui | |
|---------------|-----|-----|-----|
| Elevé | 3/4 | 2/4 | 5/8 |
| Normal | 1/4 | 2/4 | 3/8 |
| Total | 4 | 4 | |
| | 4/8 | 4/8 | |

- Tableau de vraisemblance Vent :

| | | | |
|---------------|------------|------------|-----|
| jouer | Non | Oui | |
| Faible | 2/4 | 4/4 | 6/8 |
| Fort | 2/4 | 0/4 | 2/8 |
| Total | 4 | 4 | |
| | 4/8 | 4/8 | |

2-

a. Pour décider Oui/Non, on devrait calculer et comparer :

$$\begin{cases} P(\text{couvert}/\text{non}) \times P(\text{Tiede}/\text{non}) \times P(\text{normal}/\text{non}) \times P(\text{fort}/\text{non}) \\ P(\text{couvert}/\text{Oui}) \times P(\text{Tiede}/\text{Oui}) \times P(\text{normal}/\text{Oui}) \times P(\text{fort}/\text{Oui}) \end{cases}$$

$$= \begin{cases} 2 \\ \frac{2}{4} \end{cases} * \begin{cases} 1 \\ \frac{1}{4} \end{cases} * \begin{cases} 2 \\ \frac{2}{4} \end{cases} * 0 = 0$$

Comme les deux termes sont équivalents, les deux possibilités sont vraies.

b. Pour décider Oui/Non, on devrait calculer et comparer :

$$\begin{cases} P(\text{ensol}/\text{non}) \times P(\text{Tiede}/\text{non}) \times P(\text{elevé}/\text{non}) \times P(\text{faible}/\text{non}) \\ P(\text{ensol}/\text{Oui}) \times P(\text{Tiede}/\text{Oui}) \times P(\text{elevé}/\text{Oui}) \times P(\text{faible}/\text{Oui}) \end{cases}$$

$$= \begin{cases} 3 \\ \frac{3}{4} \end{cases} * \begin{cases} 1 \\ \frac{1}{4} \end{cases} * \begin{cases} 3 \\ \frac{3}{4} \end{cases} * \frac{2}{4}$$

Donc la décision est Non