U.E. 5.2

Contrôle continu 3
le 02/10/2015

Nom, prénom : /10

Durée : 15 minutes.
***IMPORTANT !** Pour l'exercice 1, vous devez nous envoyer la réponse sous la forme d'un script python (coeff_regression_lineaire.py) par mail (lola.falletti@u-psud.fr et albenhenni@gmail.com) **et aussi** de mettre ce fichier sur dokeos.*

- **Exercice 1 : Résolution analytique pour les coefficients de la régression linéaire**

Soit le lot de données (x, y) défini de la manière suivante :

```
import numpy as np

params=(0,1)
x = 10*np.random.random(100)
y = params[0] + params[1]*x + np.random.normal(size=len(x))
```

Le résultat de la résolution analytique pour les coefficients de la régression linéaire vue en cours est la suivante :

$$\hat{\theta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{cov(x,y)}{var(x)}$$

$$\hat{\theta}_0 = \bar{y} - \hat{\theta}_1\bar{x}$$

$\hat{\theta}_1$ et $\hat{\theta}_2$ sont les coefficients que l'on souhaite estimer, tel que

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x$$

$\bar{x}$ et $\bar{y}$ sont les moyennes respectives des tableaux `x` et `y`.
La covariance et la variance sont définies par :

$$cov(x,y) = \frac{1}{N} * \sum(x_i - \bar{x})(y_i - \bar{y})$$

$$var(x) = \frac{1}{N} * \sum(x_i - \bar{x})^2$$

avec $N = len(x) - 1$

Nous allons vous demander de créer plusieurs fonctions pour calculer au final les paramètres $\hat{\theta}_1$ et $\hat{\theta}_2$ de façon analytique. Les questions de cet exercice 1 sont dépendantes.

Question 1.1 : Créer la fonction `moyenne(x)` retournant la valeur de la moyenne du tableau `x`. *(1 point)*

Question 1.2 : Créer la fonction `produit_scalaire(x,y)` permettant de retourner le résultat du produit scalaire du tableau `x` par le tableau `y`. *(1.5 point)*

Question 1.3 : Créer la fonction `variance(x,y)` permettant de retourner le résultat du calcul de la variance, définie plus haut. *(1.5 point)*

Indice : vous pouvez utiliser `np.square(...)` pour mettre un élément au carré.

Question 1.4 : Créer la fonction `covariance(x,y)` permettant de retourner le résultat du calcul de la covariance, définie plus haut. *(1.5 point)*

Question 1.5 : Créer la fonction `linear_parameters(x,y)` permettant de retourner les résultats de l'estimation des paramètres $\hat{\theta}_1$ et $\hat{\theta}_2$, d'après l'équation définie plus haut. *(1.5 point)*

- **Exercice 2 : QCM**

Question 2.1 : La méthode des moindres carrés consiste à fixer les paramètres pour la fonction $f(x_i)$ qui “fit” (ajuste) les y_i en minimisant l’expression (1 points)

[X] $\sum_i (y_i - f(x_i))^2$

[ ] $\sum_i (y_i^2 - f(x_i)^2)$

[ ] $\sum_i y_i^2 - \sum_i f(x_i)^2$

Question 2.2 : La méthode par descente de gradient met à jour la valeur des paramètres en suivant la formule $\theta_{i+1} \leftarrow \theta_i - \alpha \frac{\partial Erreur}{\partial \theta}$. Cocher les propositions suivantes qui vous semblent correctes *(2 points)*

[X] Le temps de convergence de l’algorithme dépend de α
[X] Le temps de convergence dépend de de la variance du terme d’erreur gaussien
[ ] La méthode permet de trouver systématiquement le minimum global
[ ] Les paramètres vers lesquels converge l’algorithme sont exactes