Data Science et Machine Learning

TP1 sur le modèle général de la régression linéaire

A réaliser dans l'environnement Python avec la bibliothèque Numpy

Nous souhaitons expliquer le coût de fonctionnement mensuel d'une entreprise avec les quantités produites de ses deux principaux produits. Nous disposons pour cela des données de 12 mois fournies dans le tableau Excel joint.

1. Ecrivez un programme Python qui utilise le calcul matriciel de la bibliothèque NumPy pour trouver les paramètres du modèle de la régression linéaire multiple correspondant à ce problème.

   Indication :

   ```
   colonne_un = np.ones((n,1))
   X = np.concatenate((colonne_un,X), axis=1)
   Xt = np.transpose(X)
   XtX = np.dot(Xt,X)
   Inv = np.linalg.inv(XtX)
   ```

2. Calculez le coefficient de corrélation linéaire, la somme des carrés totaux, la somme des carrés résiduels, la somme des carrés expliqués et le coefficient de détermination de deux manières.

   Indication :

   ```
   R = np.corrcoef(Y,Y_ch)[0,1]
   SCT = n*np.var(Y)
   ```

3. Calculer les variances empiriques expliqués et résiduelles et déduisez l'écart-type des résidus.
4. Réalisez le test de Fisher pour évaluer la fiabilité du modèle avec une zone de rejet de 5%.
5. Réalisez une prédiction du coût de fonctionnement le mois prochain sachant que l'entreprise planifie de produire 65 unités du produit 1 et 30 unités du produit 2.

   Indication :

   ```
   prediction = np.dot(xnew,a)
   ```