

TP5 : Les réseaux de neurones

Objectifs

Dans ce TP, nous allons étudier les réseaux de neurones en classification puis en régression.

Classification

Dans cette partie nous allons travailler sur la base de données Human resources.

Comme pour les TPs précédents, vous devez analyser les données, créer le modèle à l'aide d'un réseau de neurones, puis mesurer les performances obtenues sur une base de test.

Analyse des données

Vous constaterez que la base ne contient pas que des données numériques mais aussi des données "catégoriques". Il va donc falloir dans un premier temps transformer chaque catégorie en un entier correspondant. Heureusement, la bibliothèque sklearn fournit des outils nous aidant à cette tâche.

Etudiez le code suivant :

```
def label_encode(data):  
    # Transforme un type catégorie en entier  
    le = LabelEncoder()  
    # On récupère tous les noms de catégories possibles  
    unique_values = list(data['col'].unique())  
    le_fitted = le.fit(unique_values)  
    # On liste l'ensemble des valeurs  
    values = list(data['col'].values)  
    # On transforme les catégories en entier  
    values_transformed = le.transform(values)  
    # On fait le remplacement de la colonne dans le dataframe d'origine  
    data['col'] = values_transformed
```

Apprentissage

Vous allez devoir utiliser `from sklearn.neural_network import MLPClassifier`.

Exemple de résultats attendus :

```
Train score: 0.9593294599485666, Test score 0.9524444444444444  
[[3358  93]  
 [ 121 928]]
```

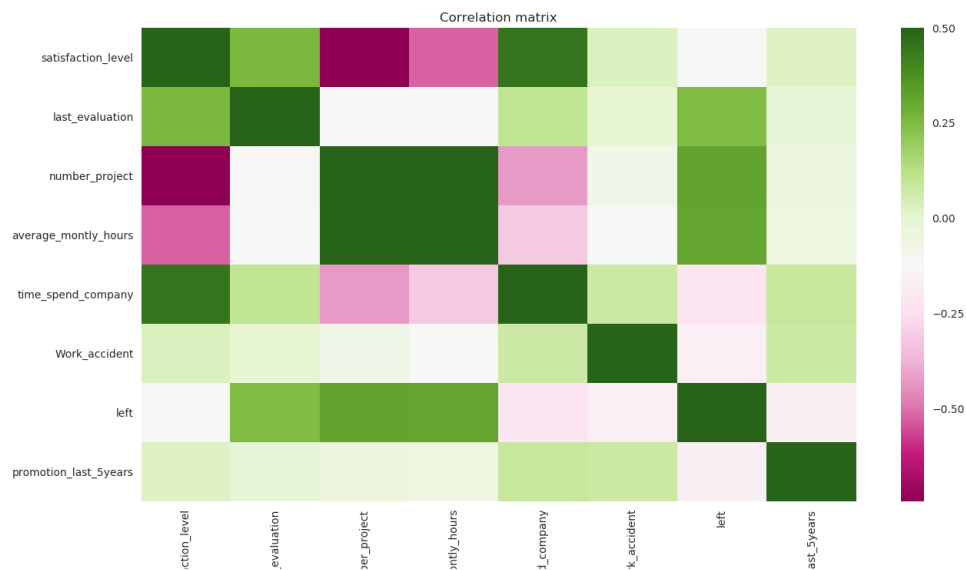
Analyse avancée des données

L'entreprise qui vous a fourni les données, souhaite retenir les bons employés. Elle vous demande donc d'analyser les données afin de comprendre pourquoi les bons employés partent de l'entreprise.

Un bon employé est défini par un employé :

- qui travaille sur plus de projets que la moyenne,
- qui effectue plus d'heures par mois que la moyenne,
- qui a obtenu une meilleure évaluation que la moyenne.
- qui travaille dans la société depuis plus longtemps que la moyenne.

Exemple de résultat :



Que pouvez-vous conclure à l'entreprise ?

Regression

Objectif

Les réseaux de neurones peuvent aussi être utilisés pour des problèmes de régression. Dans cette seconde partie, nous allons étudier la base de données bike sharing by hour. Il s'agit de prédire le nombre de locations de vélos en fonction de différents critères (météo, week-end etc). La même base est disponible mais avec une répartition journalière.

A vous de jouer

Comme pour la première partie, récupérez la base, analysez les données, et effectuez un apprentissage avec les réseaux de neurones.

Exemple de résultat :

```
Coefficient of determination: 0.99999932807  
MAE: 0.110978270303  
MSE: 0.021783586532
```

Résumé

- Les réseaux de neurones sont des algorithmes puissants pour la régression et pour la classification.
- Ils souffrent parfois d'une convergence lente.