

Apprentissage Automatique Numérique

M1 AFD

TP Perceptron, Perceptron multi-couches

Objectif du TP

L'objectif de ce TP est d'implémenter l'algorithme du perceptron pour résoudre un problème de classification binaire. Le corpus utilisé est constitué de votes des membres du congrès américains sur des sujets sensibles ainsi que l'appartenance politique du membre (républicain ou démocrate).

Quelques exemples de données du corpus :

```
republican,n,y,n,y,y,y,n,n,n,y,?,y,y,y,n,y
republican,n,y,n,y,y,y,n,n,n,n,y,y,y,n,?
democrat,?,y,y,?,y,y,n,n,n,y,n,y,y,n,n
democrat,n,y,y,n,?,y,n,n,n,n,y,n,y,n,n,y
democrat,y,y,y,n,y,y,n,n,n,y,?,y,y,y,y
```

Les données et leur description sont disponibles suivantes :

<https://archive.ics.uci.edu/ml/datasets/Congressional+Voting+Records>

Au total, 435 exemples constituent ce corpus. 100 exemples seront utilisés en tant que corpus de validation et le reste sera utilisé pour estimer les paramètres du modèle (corpus d'apprentissage). Il est important de mélanger les données avant de créer les corpus. Pensez à utiliser une "graine" (seed) pour le générateur de nombres pseudo-aléatoires afin de pouvoir reproduire les résultats.

Travail à effectuer

1. Préparer des données : coder les données textuelles en valeurs numériques
2. Écrire une fonction **classify** qui permet de donner l'étiquette d'une observation à partir d'un vecteur de poids
3. Écrire une fonction **test** qui prend en paramètre un corpus et un vecteur de poids et qui renvoie le taux d'erreur
4. Implémenter l'algorithme du perceptron dans une fonction **train**
 - Limiter l'apprentissage à un nombre maximum d'itérations
 - Définir et utiliser un autre critère d'arrêt de l'apprentissage
5. Tracer la courbe d'évolution du taux d'erreur en fonction de l'avancement de l'apprentissage (numéro de l'itération).
6. Sommes nous sûrs d'atteindre une erreur nulle sur le corpus d'apprentissage ?
7. Intégrer un *learning rate* lors de la phase d'apprentissage :

$$W^{t+1} = W^t + \lambda \hat{x}$$

8. Déterminer un schéma permettant de faire évoluer le learning rate pendant l'apprentissage.