

ING2-INFO – Ingénierie Mathématique 3

Matière pour le partiel

Augustin Cosse

Fall 2024

1 Théorie

1. Vous devez être capables de dériver la fonction de coût correspondant à la somme des carrés des résidus pour le modèle de régression linéaire. Vous devez être capables de dériver les itérations de gradient sur cette fonction de coût ainsi que la forme close du vecteur des coefficients de régression (via les équations normales).
2. Vous devez être capables de lister et caractériser les trois grandes approches de régularisation (Ridge, Lasso, Sélection du meilleur sous-ensemble)
3. Vous devez être capables d'expliquer comment la validation croisée peut être utilisée dans le cadre de la sélection de modèles (par ex. dans le cadre de la sélection du meilleur sous-ensemble).
4. Vous devez être capables de décrire les hypothèses statistiques sous-jacentes aux modèles de régression linéaire, régression Ridge et régression Lasso. Vous devez être capables de décrire les distributions de probabilité sous-jacentes.
5. Vous devez être capables de comparer les approches de régularisation en fonction de leurs efficacités respectives (sélection de caractéristiques) et de leur complexité algorithmique.
6. Vous devez être capables d'expliquer la notion d'équilibre biais-variance et comment cet équilibre peut être utilisé pour expliquer le lien entre complexité du modèle et overfitting.
7. Vous devez être capables d'expliquer comment le modèle linéaire et la fonction de coût donnée par la somme des carrés des résidus peuvent être utilisés afin d'entraîner un modèle de classification binaire
8. Vous devez être capables d'expliquer comment le modèle de classification binaire peut être généralisé à K classes (i.e. via les approches "un contre un" et "un contre tous")
9. Vous devez être capables de dériver et expliquer le modèle de régression logistique (y compris les hypothèses en terme de distributions de probabilité).
10. Vous devez être capables de motiver l'utilisation de la fonction d'entropie binaire croisée à partir de la minimisation de la fonction de vraisemblance pour le modèle de régression logistique.
11. Vous devez être capables de fournir l'expression générale d'un réseau de neurones et représenter ce réseau sous forme d'un diagramme
12. Vous devez être capables d'expliquer et d'implémenter l'algorithme de backpropagation (lister les étapes principales et fournir les équations associées).

2 Programmation

Vous devez être capables de fournir le pseudo-code pour les fonctions suivantes:

1. Descente de gradient appliquée à la somme des carrés des résidus pour le modèle linéaire, l'extension aux caractéristiques polynomiales ainsi que la régularisation Ridge.
2. La résolution des équations normales et leur extension à la régularisation Ridge.
3. La validation croisée (e.g. validation à K compartiments)
4. L'apprentissage d'un modèle de classification binaire via la descente de gradient et son extension à plusieurs classes via les approches de type "un contre tous" et "un contre un"
5. La propagation forward et la backpropagation au sein d'un réseau de neurones simple (quelques couches cachées et quelques neurones par couche)

Vous devez être capables d'entraîner et d'appliquer les différents modèles SCIKIT-LEARN utilisés lors des séances de TP.