

Questions principales:

1. Méthodes et critères d'évaluation de modèles en apprentissage supervisé

*Méthodes d'évaluation de modèle : motivation, resubstitution, validation croisée, bootstrap
Sélection de modèles par validation croisée : petite et grande base de données, biais de sélection
Principaux critères d'évaluation de modèle en classification (taux d'erreur, sensibilité/spécificité, courbes ROC...) et en régression (erreur quadratique, corrélation...), évaluation lors du test et lors de l'apprentissage*

2. Apprentissage supervisé par arbres de décision et de régression

*Algorithmes de développement, et d'élagage
Méthodes d'ensemble (bagging, boosting, random forests)
Noyau (kernel) associé à un arbre et à un ensemble d'arbres
Discuter la motivation des différents algorithmes et leur complexité.*

3. Apprentissage au sens des moindres carrés

*Régression linéaire (avec terme de régularisation), perceptrons multicouches pour la régression, méthodes de régression à base de noyaux.
Poser clairement le problème d'optimisation correspondant à l'algorithme étudié (fonction objectif à minimiser, paramètres ajustés, formulation des conditions d'optimalité, discussion des cas pathologiques le cas échéant)
Discuter la motivation et la complexité des différents algorithmes.*

4. Méthodes du plus proche voisin et méthodes à base de noyaux

*Algorithmes 1-NN et K-NN.
Interprétation kernel des arbres et de la classification linéaire.
Quel est l'intérêt de la méthode du plus proche voisin par rapport aux autres méthodes d'apprentissage supervisé ? Quels sont ses points faibles ?*

5. Machine à support vectoriel et méthodes à base de noyaux

*Notion de marge, formulation du problème d'optimisation, vecteurs de support, marge douce
Définition d'un noyau, truc à noyau, exemples de noyaux, méthodes à base de noyaux
Discuter les points forts et points faibles des machines à support vectoriel et l'intérêt des méthodes à base de noyaux en général.*

6. Compromis biais-variance et méthodes d'ensemble

*Décomposition biais-variance, sur et sous-apprentissage, paramètres influençant le biais et la variance, techniques de réduction de variance
Donner des exemples de techniques de réduction de variance vues au cours, discuter et comparer les différentes méthodes d'apprentissage du point de vue de leur biais et de leur variance.
Méthodes d'ensemble : motivation et algorithmes (bagging, random forests, boosting)*

7. Apprentissage non-supervisé

*Motivation
Méthodes de clustering : position du problème, méthode k-means et clustering hiérarchique, biais de sélection.
Méthodes de réduction de dimensionnalité : analyse en composantes principales, extensions.*

Questions secondaires:

1. Protocoles d'apprentissage automatique : Apprentissage supervisé, non-supervisé, apprentissage par renforcement, apprentissage en mode batch vs apprentissage on-line, apprentissage semi-supervisé et transductif. Donner un exemple de problème pratique et de méthode pour chaque type de protocole.
2. Critères d'évaluation de modèles en apprentissage, utilisés pour l'apprentissage supervisé et non-supervisé (Quantité d'information, erreur quadratique, vraisemblance, complexité)

3. Critères d'évaluation de modèles en apprentissage, utilisés lors de la phase de test en classification (taux d'erreur, sensibilité, spécificité, courbes ROC...) et en régression (erreur quadratique, corrélation...).
4. Sources d'erreur (erreur résiduelle, biais, variance) : discussion intuitive de l'influence du choix des attributs (nombre, types, combinaisons), de la complexité des modèles, de la taille de l'ensemble d'apprentissage).
5. Méthodes d'évaluation du taux d'erreur (resubstitution, ensemble de test, validation croisée, leave-n-out, bootstrap) :expliquer le principe et les avantages et inconvénients de ces méthodes.
6. Approches pour limiter ou diminuer le surapprentissage (relation entre sur/sous apprentissage et biais/variance; contrôle de la complexité; sélection de variables, méthodes d'ensemble).
7. Méthodes de clustering : position du problème, méthode k-means et clustering hiérarchique, biais de sélection.
8. Sélection de variables : motivation, filtrage, méthodes embarquées, wrapper, biais de sélection.