



APPRENTISSAGE BAYÉSIEN ET EXPLORATION DE TEXTES

BAYESIAN MACHINE LEARNING AND TEXT MINING

Responsable(s): Alexandre SAIDI, Mohsen ARDABILIAN, Stéphane DERRODE

| Cours : 12 | TD : 0.0 | TP : 0.0 | Autonomie : 0.0 | BE : 8 | Projet : 0.0 | Langue du cours : FR

Objectifs de la formation

On se focalisera sur la famille des méthodes bayésiennes, qui se distingue par son optimalité au sens de certains critères, par son coût réduit d'un point de vue algorithmique et par l'interprétabilité de ses résultats. Nous étudierons aussi les solutions qui s'offrent au Data Scientist lorsque l'échantillon d'apprentissage est peu important au regard du nombre de paramètres à apprendre, ou lorsque l'apprentissage doit se faire de manière non supervisée. En termes d'application, nous mettrons l'accent sur l'exploration d'un corpus textuel pour découvrir par exemple de nouveaux clients éligibles à la vente d'un service/produit, de prévoir les sentiments (avis) des clients ou comprendre les comportements qui prédisent une fraude.

Mots-clés : Théorie de la décision bayésienne, Apprentissage non supervisé, Modèles de Markov cachés, Fouille de textes, Analyse de sentiments, Chatbot, Traitement de Langues Naturelle.

Programme

- Décision bayésienne (2h)
- Modèle de mélange gaussien (2h)
- Chaîne de Markov cachée (2h)
- BE sur l'apprentissage bayésien (2h)
- Technologies et méthodes de traitement de langues naturelles et fouille de texte (8h)
- Restitution d'une lecture scientifique par groupe (4h)

Compétences

- Sélectionner la ou les méthodes de ML adaptées à son problème de classification, en tenant compte de différents critères.
- développer des programmes mettant en œuvre ces méthodes pour analyser ses propres données.
- De mettre en œuvre une chaîne de traitements pour interpréter des textes (p. ex. tweet).
- De se familiariser avec les techniques et outils modernes de fouille de texte et de lire des papiers de recherche récents sur les sujets évoqués

Travail en autonomie

Objectifs : Les étudiants, répartis en groupes, devront lire et comprendre un article scientifique sur l'un des sujets abordés en cours. Les articles seront proposés par les encadrants ou par les étudiants (après accord des encadrants).

Méthodes : Le travail se fera en autonomie et la restitution orale se fera lors d'une séance collective de restitution.

Bibliographie

M. R. Gupta and Y. Chen, • *THEORY AND USE OF THE EM ALGORITHM*, Foundations and Trends in Signal Processing, Vol. 4(3), pp. 223–296, 2011
M. Watanabe and K. Yamaguchi *THE EM ALGORITHM AND RELATED STATISTICAL MODELS*, Statistics: Dekker series of textbooks and monographs, 2004
Michael W. Berry, Jacob Kogan *TEXT MINING: APPLICATIONS AND THEORY*, Willey, 2010

Contrôle des connaissances

Note = 50% savoir + 50% savoir-faire
Note de savoir = 100% examen terminal
Note de savoir-faire = BE text Mining (50%) + restitution d'un article scientifique (50%)