

TP 2 - Méthodes de régression pour l'inférence de réseaux.

Olivier Goudet

February 26, 2024

Chaque TP de cette UE est noté. Il est à faire individuellement et à déposer avant la fin du semestre sur le dépôt Moodle de l'UE.

Questions

1. Utiliser la méthode Lasso pour retrouver la couverture de Markov (Markov Blanket) de la variable V_0 pour le réseau 1. (Voir Cours 2)
2. Quels variables sont dans cette couverture de Markov ?
3. Appliquer cette méthode pour retrouver des couvertures de Markov à toutes les variables du graphe 1 de façon à obtenir un fichier de prédiction, sous la forme d'une liste de liens "Cause", "Effet", "Score" dans un fichier, avec un tri suivant le score en valeur absolue du plus fort au plus faible (cf. TP 1 exo 2).
4. Utiliser cette méthodes pour les autres graphes de 2 à 5 et comparer ces résultats avec la méthode de corrélation paire à paire.
5. Calculer les score de précision moyenne obtenus pour les différents datasets avec cette méthodes. Les comparer avec ceux obtenus au TP1.
6. Utiliser la méthode GraphicalLasso de scikit-learn pour l'ensemble des graphes (slide 63 cours 2). Comparer avec les résultats obtenus par la méthode Lasso.
7. Réimplémentez la méthode GENIE3.
(<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0012776>). On pourra utiliser la méthode RandomForestRegressor de scikit-learn. Voir cours 4.
8. Proposer une autre méthode sélection de variables basée sur la régression (de préférence non-linéaire) de façon à obtenir les graphes d'indépendance.
9. En quoi ces méthodes de sélection de variables peuvent donner de l'information sur le sens de la causalité ?
10. Est-ce que tous les liens obtenus par ces méthodes de sélection de variable sont théoriquement dans le vrai graphe causal qui a généré les données ?
11. Comparez les résultats obtenus par les différentes méthodes de régression que vous avez implémenté pour l'inférence de réseaux. Pour faire ces comparaisons, pour un même dataset, on pourra afficher les cours de précision/recall correspondant au différentes méthodes implémentée sur un même graphique (cf. TP1).