

Examen

Cours Apprentissage et Fouille de Données

7 novembre 2024 (sans document)

Nom :

Prénom :

Le contrôle comporte deux parties, l'une portant plus particulièrement sur le *cours d'apprentissage artificiel*, l'autre sur le *cours de fouille de données*. Afin de faciliter la correction, il vous est demandé de bien faire les deux parties du contrôle sur des feuilles séparées.

- **La partie suivante du contrôle porte sur l'apprentissage.** Il y a 13 questions (notées sur 22 points, avec une note finale de 20 au maximum).
- Pensez à **mettre votre nom sur votre copie.**

-
1. (3 points) Une question qui sera dévoilée en examen.
 2. (2 points) Une question qui sera dévoilée en examen.
 3. (2 points) Une question qui sera dévoilée en examen.
 4. (* 1 point) Soit une tâche d'apprentissage d'une fonction binaire : avec seulement deux classes. Supposez que l'algorithme d'apprentissage soit très simple (simpliste, mais pas absurde) : il prédit chaque nouvel exemple comme appartenant à la classe majoritaire dans l'ensemble d'apprentissage qui lui a servi à apprendre, et si il y a égalité entre les classes, il prédit au hasard.
Supposons maintenant que les données disponibles comportent 100 exemples, dont 50 de la classe '1' et 50 de la classe '2'.
Que pouvez-vous dire du **taux d'erreur** obtenu sur l'ensemble de test si l'on utilise un ensemble d'apprentissage de 70 exemples et un ensemble de test comportant les 30 exemples restants.
 5. (* 2 points) Expliquer ce que mesure la **courbe ROC**.
En particulier précisez ce que représentent les axes, ce que serait une courbe représentant un apprentissage parfait et une courbe qui représenterait une courbe correspondant au pire apprentissage.
 6. (* 2 points) Que mesurent le **rappel** et la **précision** dans le cas d'une classification binaire avec deux classes '+' et '-' ? Pourquoi y a-t-il un *compromis* entre précision et rappel ? Illustrez avec un algorithme de moteur de recherche qui retourne les liens jugés pertinents pour votre requête.

7. (* 2 points) Que prescrit le principe de **minimisation du risque empirique** (MRE)?
Pourquoi faut-il **régulariser** ce risque?
8. (1 point) Une question qui sera dévoilée en examen.
9. (* 1 point) Comment se fait l'exploration de l'espace des hypothèses dans l'apprentissage avec un réseau de neurones?
Qu'est-ce que l'on risque en utilisant ce type d'exploration?
10. (* 1 point) En quoi consiste le « **one-hot encoding** » (aussi appelé *codage disjonctif complet*)?
11. (1 point) Une question qui sera dévoilée en examen.
12. (1 point) Une question qui sera dévoilée en examen.
13. (3 points) Une question qui sera dévoilée en examen.