

Python et Machine Learning

TRAVAUX DIRIGÉS

Quatrième partie

Salim Lardjane

Université de Bretagne Sud

1. Sauvegarder les données iris de R dans un fichier csv. Les lire sous Python dans un Data Frame pandas. Effectuer diverses représentations graphiques des données iris (nuages de points, boxplots, histogrammes, etc).
2. Le jeu suivant (Monty Hall) oppose un présentateur à un candidat. Le joueur est placé devant trois portes fermées. Derrière l'une d'elles se trouve une voiture et derrière chacun des deux autres se trouve une chèvre. Il doit tout d'abord désigner une porte. Puis le présentateur ouvre une porte qui n'est ni celle choisie par le candidat, ni celle cachant a voiture (le présentateur sait quelle est la bonne porte dès le début). Le candidat a alors le droit ou bien d'ouvrir la porte qu'il a choisie initialement, ou bien d'ouvrir la troisième porte. Que doit-il faire ? Quelles sont les chances de gagner la voiture en agissant au mieux ?
Simuler le jeu du Monty Hall, identifier la meilleure stratégie et calculer la probabilité de gagner la voiture avec cette stratégie. Essayer de retrouver le résultat avec le calcul des probabilités.
3. Créer un tableau numpy contenant les données 1, 2, 3, ... , 10. Calculer la moyenne et l'écart-type (échantillonal) correspondant.

4. Générer et représenter graphiquement la densité de probabilité (PDF) d'une distribution normale d'espérance 5 et d'écart-type 3.
 - a. Générer 1000 pseudo-réalisations de cette distribution.
 - b. Calculer une estimation de l'écart-type de la moyenne pour les données obtenues.
 - c. Représenter un histogramme des données.
 - d. Déterminer un intervalle de fluctuation contenant approximativement 95% des données.