FTML Exercices 3

Pour le 4 avril 2024

TABLE DES MATIÈRES

1 RISQUES DE BAYES

Retrouver les valeurs théoriques des risques de Bayes des deux problèmes de l'exercice 3 du TP 1. Pour ces deux problèmes, nous avons vu en cours qu'on connaît le prédicteur de Bayes, et on peut ainsi calculer son risque car nous connaissons la distribution d'où les données sont issues.

Pour le premier problème (penalty), le calcul est similaire à ceux effectués au début du 3e cours magistral. Pour le deuxième problème (nombre de streams), une espérance non triviale est à calculer, il est possible de le faire numériquement.

2 ESTIMATEURS ET ESPÉRANCES

2.1 Moyenne empirique et espérance

1) On se donne une variable aléatoire réelle X ayant un moment d'ordre 1. On note $E[X]$ son espérance. On suppose que l'on dispose de n samples indépendants tirés de cette variable, $(x_1, \dots, x_n)$, qui suivent donc la loi de X. On considère leur moyenne empirique $S_n = \sum_{i=1}^{n} x_i$. Montrer que l'espérance de S_n est l'espérance de X, $E[X]$ (autrement dit, que S_n est un estimateur non biaisé de $E[X]$).

2) Simuler une variable de votre choix pour observer la convergence de S_n vers $E[X]$.

2.2 Espérance du risque empirique

1) On se donne un problème d'apprentissage supervisé usuel et un estimateur f, fixé et indépendant du dataset. Montrer que l'espérance du risque empirique de f est le risque réel de f. Autrement dit : le risque empirique de f est un estimateur non biaisé de son risque réel.

La notion d'espérance a bien un sens ici car il faut voir le dataset comme une variable aléatoire, ainsi le risque empirique de f est bien une variable aléatoire.

2) Simuler un problème de votre choix pour observer le résultat (c'est ce qui est fait dans simulations/lecture_3/empirical_risk_convergence)