

# Examen blanc - Fondamentaux théoriques en Machine Learning

8 décembre 2018

## Exercice 1

Dans tout cet exercice, on considère le tableau de données suivant, avec les prédictions de deux estimateurs.

| X1 | X2 | Y | Y_EST_1 | Y_EST_1 |
|----|----|---|---------|---------|
| 3  | 2  | T | F       | T       |
| 2  | 2  | F | F       | F       |
| 3  | 1  | T | T       | F       |
| 1  | 2  | F | F       | F       |
| 1  | 2  | T | F       | F       |
| 3  | 2  | F | F       | F       |
| 2  | 3  | F | F       | F       |
| 2  | 2  | T | T       | F       |
| 2  | 1  | F | F       | F       |
| 2  | 3  | T | F       | T       |
| 1  | 2  | F | T       | F       |
| 1  | 1  | F | F       | F       |
| 3  | 1  | T | T       | F       |
| 1  | 3  | F | F       | F       |
| 2  | 2  | T | T       | T       |
| 3  | 1  | F | T       | F       |
| 1  | 2  | T | T       | T       |

- Produisez les matrices de confusion respectives des deux estimateurs.
- Pour chaque estimateur, trouvez une fonction de coût qui l'avantage par rapport à l'autre en terme de risque empirique.
- Le choix d'un SVM paraît-il pertinent ? Pourquoi ?
- Produisez l'estimateur bayésien naïf associé à une perte uniforme.
- Vérifiez qu'il prédit bien le résultat le plus fréquent sur l'échantillon dans les 5 cas où les fréquences ne sont pas égales. Expliquez empiriquement ses choix dans les cas (1,2) et (3,2).

f) Proposez un arbre de décision binaire qui sépare bien les 5 cas où les fréquences ne sont pas égales. Comparez ses décisions dans les cas (1,2) et (3,2) avec celles du naïve bayes.

### Exercice 2

a) Quelle est la taille maximale d'un polygone régulier qui puisse être pulvérisé par des triangles ? Déduisez-en une borne sur la dimension de Vapnik–Chervonenkis de l'ensemble des triangles sur le cercle unité.

b) Quelle fonction kernel pourriez-vous suggérer pour séparer les données suivantes avec un svm ?

