

# Aprendizagem Automática 2022/23

## Ficha prática 4

### Exercício#4.1 - Algoritmo Naive Bayes

Neste exercício terá um contacto com o Algoritmo Naive Bayes aplicado num conjunto de dados de pequena dimensão.

Considere o seguinte conjunto de dados baseado no dataset IRIS, onde cada exemplo é caracterizado por 4 atributos discretizados com dois valores V/F e cada instância pertence a uma de 3 classes.

ID	Atr1>6	Atr2>3	Atr3>5	Atr4>1.5	Classe
1	F	V	F	F	Iris-setosa
2	F	V	F	F	Iris-setosa
3	F	V	F	F	Iris-setosa
4	V	F	F	F	Iris-versicolor
5	F	F	F	F	Iris-versicolor
6	F	F	F	F	Iris-versicolor
7	V	F	F	V	Iris-virginica
8	V	F	V	V	Iris-virginica
9	V	V	V	V	Iris-virginica

- 1) Calcule (à mão) a que classe pertence o exemplo {F, F, V, F} se o conjunto anterior for fornecido ao algoritmo Naive Bayes (Categorical Naive Bayes). Deverá usar o estimador simples ? ou um estimador “suavizado” como o estimador de laplace ?

## Exercício#4.2 - Algoritmo Naive Bayes

Neste exercício aplicará o algoritmo Naive Bayes e terá um primeiro contacto com a necessidade de pre-processamento dos dados para adaptar os dados à função CategoricalNB (Naive Bayes)

- 1) Carregue o conjunto de dados car.csv usando por exemplo a função `read_csv("car.csv")` do pandas. Este conjunto de dados também pode ser descarregado de

<https://archive.ics.uci.edu/ml/datasets/car+evaluation>

e posteriormente editado para um formato CSV.

- 2) Defina os atributos X e a classe y por exemplo usando,

```
X=data.iloc[:, :-1]
```

```
y=data.iloc[:, 1]
```

- 3) Faça um split treino teste
- 4) Faça um fit usando o algoritmo CategoricalNB e observe o resultado

Se obteve um erro, determina a origem do erro (veja o eventual erro e a documentação sobre CategoricalNB, em particular o tipo dos dados de entrada

- 5) Se necessitar de transformar os dados veja a definição de `preprocessing.LabelEncoder()`, e verifique se é adequada. Pode testar o efeito de:

```
le = preprocessing.LabelEncoder()
```

```
data['maint'] = le.fit_transform(data['maint'])
```

- 6) Adapte o seu código, faça um fit do modelo, proceda à avaliação de desempenho.
- 7) Investigue os parâmetros do algoritmo e faça testes sem smoothing, com smoothing (estimador de laplace), e com smoothing usando um  $\alpha=10$ , e compare os resultados.