

1. Considere o seguinte conjunto de dados:

Nome	Dívida	Salário	Casado	Risco
José	Alto	Alto	Sim	Alto
Ana	Baixo	Alto	Sim	Baixo
João	Alto	Alto	Não	Baixo
Maria	Alto	Baixo	Sim	Alto
Rui	Baixo	Baixo	Sim	Alto

Faça a previsão do Risco para a seguinte instância:

Dívida=Baixo; Salário=Alto; Casado=Não

- com base no Classificador K-Vizinhos-mais-Próximos, com $K=3$
- com base no Classificador Naive-Bayes, use a estimativa Laplace para cálculo das probabilidades condicionais.

2. Um problema de classificação envolve **4 Classes** (1, 2, 3, 4). Os dados de treino contêm 250 casos de cada classe, assim no total são 1000 casos. Suponha que um determinado teste baseado num atributo **Conta**, divide os exemplos em 2 grupos.

1º grupo (ex. Conta = sim) contém 600 casos	2º grupo (ex. Conta = não) contém 400 casos
250 exemplos da classe 1 150 exemplos da classe 2 150 exemplos da classe 3 50 exemplos da classe 4	0 exemplos da classe 1 100 exemplos da classe 2 100 exemplos da classe 3 200 exemplos da classe 4

- Apresente a matriz de confusão, assumindo a classificação apenas com base no atributo Conta.
- Calcule a taxa de erro do modelo. Que pode concluir.
- Qual o significado do valor no cruzamento das linhas identificadas com “Classe2” e “Classe3” e as colunas de previsão das classes?
- Quantos “positivos verdadeiros” vs. “positivos falsos” há em relação às classes 1 e 4? O que pode concluir?
- Calcule as medidas precision e recall para as classes 1 e 4? O que pode concluir?

1. Consider the following dataset:

Name	Debt	Salary	Married	Risk
José	High	High	Yes	High
Ana	Low	High	Yes	Low
João	High	High	No	Low
Maria	High	Low	Yes	High
Rui	Low	Low	Yes	High

Predict Risk for the following instance:

Divide = Low; Salary = High; Married = No

- based on the K-Nearest-Neighbors Classifier, with K=3
 - based on the Naive-Bayes Classifier, using the Laplace estimation to calculate the conditional probabilities, with laplace correction=1.
2. A classification problem involves **four classes** (1, 2, 3, 4). The training data contain 250 instances of each class, so the total are 1000 cases. Suppose that a particular test based on **Account** attribute divided the dataset into 2 groups of the examples.

1st group (eg Account = yes) contains 600 cases	2nd group (eg Account = no) contains 400 cases
250 Examples of Class 1 150 Examples of Class 2 150 Examples of Class 3 50 Examples of Class 4	0 Examples of Class 1 100 Examples of Class 2 100 Examples of Class 3 200 Examples of Class 4

- Develop and present the confusion matrix, assuming the classification is based solely on the Account attribute.
- Calculate the rate of error of the model. What can be concluded?
- What is the meaning of the value at the intersection of the lines marked with "Class 2" and "Class 3" and the columns for the prediction of classes?
- How many "true positive" and "false positive" are in relation to "Classe4"?
- Calculate the precision and recall measures for classes 1 and 4? What can you conclude?