**Exercícios: aprendizagem *bayesiana***

1. Suponhamos que recebeu uma mensagem de *e-mail* com o assunto "Olá". De acordo às estatísticas realizadas com as mensagens que tem recebido, somente 10% do total das mensagens recebidas são *spam*, 50% do *spam* tem como assunto "Olá" e 2% das mensagens que não são *spam* têm o assunto "Olá". Qual é a probabilidade de a mensagem recebida ser *spam*.
2. Uma rara doença genética foi identificada recentemente. A boa notícia é que apenas uma pessoa por cada milhão de pessoas tem a referida doença e o teste genético é extremamente bom; se um indivíduo tem a doença o teste é sempre positivo e se não tem a doença o teste retorna um falso positivo em apenas 0,1% dos casos. A má notícia é que você foi dado como positivo num teste. Qual é a probabilidade de que você tenha a doença?
3. Considere o seguinte conjunto de treino, o qual tem dois atributos com três valores possíveis cada (0, 1 e 2) e duas possíveis classes (x e y).



Que classe seria atribuída pelo algoritmo Naïve Bayes para a seguinte instância de teste?



1. Considere o seguinte conjunto de dados acerca das características e origens de uma série de indivíduos. Se observamos um novo indivíduo *alto* e com *cabelo loiro*, qual seria o país mais provável atribuído ao mesmo pelo algoritmo Naïve Bayes?



1. Dado o conjunto de exemplos de treino caracterizados por 4 atributos, com dois valores possíveis cada (0 e 1), qual seria a classe correspondente ao exemplo de teste <0, 0, 1, 1>? Qual seria a classe atribuída ao mesmo exemplo utilizando estimação - m?



1. Utilize o método Naïve Bayes para determinar a classe associada a uma instância com *Nível = 3* e *Estado = Mau*, a partir do conjunto de exemplos de treino abaixo indicados. Considere que *Nível* é um atributo com valores reais e *Estado* um atributo com valores no conjunto {Bom, Mau}.



1. Uma empresa decide criar um classificador de documentos. Para começar, se trata de classificar os documentos em duas categorias, *desportos* ou *política*. Cada documento é representado como um vector de atributos, descrevendo cada atributo a presença ou ausência das palavras seguintes no documento, **x** = (golo, futebol, golfe, defesa, transgressão, portinhola, gabinete, estratégia). Os dados de treino são apresentados a continuação.



Utilizando um classificador Naïve Bayes, qual é a probabilidade de que o documento   
**x** = (1, 0, 0, 1, 1, 1, 1, 0) seja sobre política?

1. Suponhamos que se conhece que a probabilidade *a priori* de reincidência para um paciente que sofre por primeira vez um ataque de coração é de 0,30.

Suponhamos que no modelo construído para predizer a reincidência existem duas variáveis predictoras que se supõe que são independentes dado o valor da classe. As duas variáveis predictoras são *a realização de exercício moderado de forma assídua* e a *idade do paciente*. Enquanto a primeira é discreta com dois valores possíveis (*sim*, *não*), a segunda é *contínua e distribuída normalmente para cada valor da classe*.

Entre os pacientes que não têm reincidência, 80% realizam exercícios moderados de forma assídua, enquanto entre os que têm reincidência esta probabilidade diminui para 10%.

A idade dos pacientes não reincidentes segue um modelo normal com média de 60 anos e desvio padrão de 10 anos, enquanto no caso dos pacientes com reincidência a média é de 75 anos e 15 de desvio padrão.

Efectuar o diagnóstico para um paciente com 58 anos que não pratica exercícios de maneira assídua.