

DCA0200 - INTELIGÊNCIA ARTIFICIAL- 2022.2

Primeira Lista de Exercícios e Trabalhos

Data de apresentação da lista e dos trabalhos: 20/09/2022

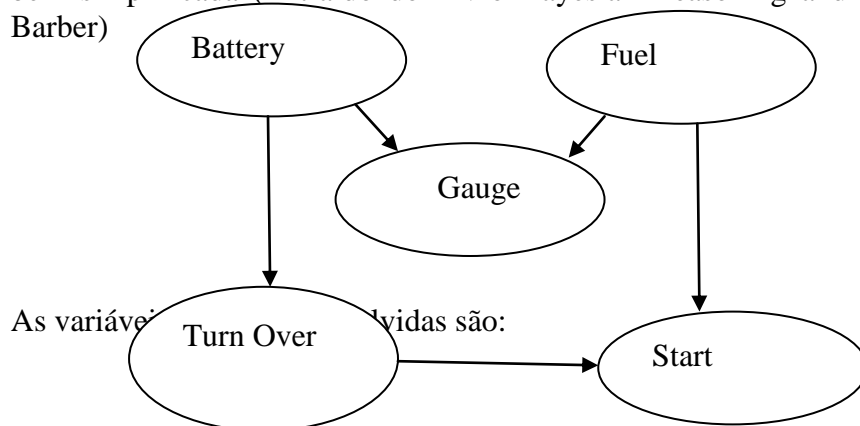
Obs. A lista, os trabalhos e o projeto podem ser feitos em grupo de até dois alunos. No dia da entrega da lista os alunos devem apresentar os problemas da lista e os trabalhos.

Lista de Problemas

1-) A tabela de dados abaixo ilustra a aplicação do método Naïve-Bayes. Um determinado banco deve decidir se um cliente deve ou não receber um empréstimo bancário em função da sua condição de bom ou mau pagador. Considerando os dados de treinamento abaixo, aplique o classificador Naive-Bayes, para atribuir a classe (rótulo) para os registros 12 e 13:

Registro	Tem casa própria	Estado Civil	Possui Carro	Rendimentos	Bom Pagador
1	Sim	Solteiro	Sim	Alto	Não
2	Não	Casado	Sim	Médio	Não
3	Não	Solteiro	Não	Baixo	Não
4	Sim	Casado	Sim	Alto	Não
5	Não	Divorciado	Não	Médio	Sim
6	Não	Casado	Não	Baixo	Não
7	Sim	Divorciado	Sim	Alto	Sim
8	Não	Solteiro	Sim	Médio	Sim
9	Não	Casado	Sim	Baixo	Não
10	Não	Solteiro	Não	Médio	Sim
11	Sim	Divorciado	Não	Médio	Não
12	Não	Divorciado	Sim	Alto	?
13	Sim	Solteiro	Não	Médio	?

2-) A rede bayesiana abaixo concerne ao problema de partida de um carro, de uma forma bem simplificada (Extraído do Livro Bayesian Reasoning and Machine Learning - D. Barber)



b=battery, g=gauge, f=fuel, t =turn over, s=start
fa=false, tr= true

As probabilidades referentes a rede bayesiana são dadas por:

$p(b=bad)=0.05$	$p(f=empty)=0.1$
$p(g=empty b=good,f=not\ empty)=0.05$	$p(g=empty b=good,f=empty)=0.98$
$p(g=empty b=bad,f=not\ empty)=0.07$	$p(g=empty b=bad,f=empty)=0.97$
$p(t=fa b=good)=0.1$	$p(t=fa b=bad)=0.96$
$p(s=fa t=tr, f=not\ empty)=0.01$	$p(s=fa t=tr, f=empty)=0.92$
$p(s=fa t=fa, f=not\ empty)=1.0$	$p(s=fa t=fa, f=empty)=0.99$

Um agente inteligente com base nas inferências, isto é, o cálculo da $P(f=empty|s=no)$ (a probabilidade do tanque estar vazio dado que o carro não deu partida) e da $P(b=bad|s=no)$ (a probabilidade da bateria estar descarregada e o carro não deu partida), deve decidir qual o problema mais provável pela não partida do carro. Apresente a solução e implemente os cálculos de forma computacional

3-) Uma rede de crença (ou rede bayesiana), modela a relação entre as variáveis: oil (price of oil), inf (inflation) , eh (economy health), bp (British Petroleum Stock price), rt (retailer stock price). Cada variável tem dois estados (l:low) e (h:high), exceto a variável bp que tem adicionalmente o estado (n: normal). A rede de crença modela as variáveis de acordo com a tabela abaixo. (Extraído do Livro Bayesian Reasoning and Machine Learning - D. Barber)

$P(eh=l)=0.7$	
$P(bp=l oil=l)=0.9$	$P(bp=n oil=l)=0.1$
$P(bp=l oil=h)=0.1$	$P(bp=n oil=h)=0.4$
$P(oil=l eh=l)=0.9$	$P(oil=l eh=h)=0.05$
$P(rt=l inf=l,eh=l)=0.9$	$P(rt=l inf=l,eh=h)=0.1$
$P(rt=l inf=h,eh=l)=0.1$	$P(rt=l inf=h,eh=h)=0.01$
$P(inf=l oil=l, eh=l)=0.9$	$P(inf=l oil=l, eh=h)=0.1$
$P(inf=l oil=h, eh=l)=0.1$	$P(inf=l oil=h, eh=h)=0.01$

- a-) Determine o gráfico da rede de crença (rede bayesiana) para este problema
b-) Dado que a $bp=n$ e $rt=h$, qual é a probabilidade de que a inflação seja alta?

Apresente a solução e implemente os cálculos de forma computacional

4-) Considere o problema de tomada de decisão caracterizado por uma sequência de eventos que podem ser apresentado por um grafo conhecido como rede de decisão. O problema em questão consiste das escolhas e das decisões por parte de uma empresa de petróleo. Uma determinada empresa petrolífera obteve a concessão para explorar uma certa região. Os estudos anteriores (testes preliminares) estimam a probabilidade de existir petróleo nessa região em 30 %. A companhia pode optar por um novo teste, que custa US\$ 100.000,00, sendo que, se realmente existe petróleo, esse teste dirá com uma probabilidade de 0.85 que existe, e se realmente não existe, dirá com probabilidade 0.75 que não existe. Considerando que o custo de perfuração será de US\$ 1.000.000,00 e que, se for encontrado petróleo, a companhia receberá US\$ 20.000.000,00 pela produção. Considere, portanto os seguintes eventos e os seus complementos: (i) Evento T (a companhia faz o teste); (ii) Evento F (o teste é favorável à existência de petróleo; (iii) Evento P (a companhia perfura o poço); (iv) Evento E (existe petróleo).

- a-) Construa a rede indicando os nós de decisões e os nós ao acaso (variáveis aleatórias). Considere as funções de utilidade, representadas por losangos, como sendo o lucro =receita - despesas, calculado em cada percurso da árvore.
- b-) Determine em cada nó dos percursos da árvore a utilidade esperada.
- c-) Usando o critério da utilidade máxima esperada, determine a melhor decisão.
- d-) Qual o valor esperado do lucro da companhia se for tomada a melhor decisão?
- e-) Apresente também a solução deste problema através de um programa computacional e simule diferentes situações alterando o valor das probabilidades.

Observações:

- (i) O evento inicial da árvore é se companhia faz ou não faz o teste.
- (ii) Para cada evento tem o seu complementar: Exemplo: T: Faz o teste, $\neg T$: Não faz o teste
- (iii) Para o cálculo da utilidade esperada determine antes as probabilidades condicionais a posteriori com base no teorema de Bayes.

4-) Considere os dados apresentados na tabela abaixo. Assumindo que as distribuições associadas a cada classe são gaussianas com probabilidades a priori dadas por $P(\omega_1) = P(\omega_2) = 1/2$ busque solucionar as questões abaixo.

Classe	ω_1		ω_2	
Variáveis	x_1	x_2	x_1	x_2
1	-5.01	-8.12	-0.91	-0.18
2	-5.43	-3.48	1.30	-2.06
3	1.08	-5.52	-7.75	-4.54
4	0.86	-3.78	-5.47	0.50
5	-2.67	0.63	6.14	5.72
6	4.94	3.29	3.60	1.26
7	-2.51	2.09	5.37	-4.63
8	-2.25	-2.13	7.18	1.46

9	5.56 2.86	-7.39 1.17
10	1.03 -3.33	-7.50 -6.32

a-) Estime o vetor média e a matriz de covariância de cada distribuição gaussiana multivariada. Para estimativa dos vetores média e das matrizes utilize os estimadores de máxima verossimilhança dados abaixo.

Estimador de média: $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$ e

Estimador dos elementos das matrizes de covariância: $s_{ij} = \frac{\sum_{l=1}^N (x_{il} - \hat{\mu}_i)(x_{jl} - \hat{\mu}_j)}{N - 1}$

b-) Determine a superfície de decisão fazendo $g_1(\mathbf{x})=g_2(\mathbf{x})$, $g_1(\mathbf{x})$ e $g_2(\mathbf{x})$ as funções discriminantes e $\mathbf{x}=[x_1, x_2]^t$.

As funções discriminantes $g_i(\mathbf{x})$, $i=1,2$ podem ser calculadas para estas condições pela equação abaixo.

$$g_i(\mathbf{x}) = -1/2((\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) - 1/2 \ln |\boldsymbol{\Sigma}_i|) \quad i=1,2$$

$\boldsymbol{\mu}$: Vetor Média

$\boldsymbol{\Sigma}$:

$\boldsymbol{\Sigma}^{-1}$: Inversa da Matriz de Covariância

$|\boldsymbol{\Sigma}_i|$: Determinante da matriz de covariância

c-) Classifique os pontos (padrões) $(1, 2)^t$, $(5, 3)^t$, $(0, 0)^t$, $(1, 0)^t$, decidindo pela classe com maior valor das funções discriminantes, isto é, $\max\{g_i(\mathbf{x})\}$ ou faça pela região definida pela superfície de separação (função discriminante)

5-) Apresente um estudo sobre as seguintes medidas de desempenho de classificadores de padrões: Acurácia, Precisão e F-Score

Trabalhos:

Trabalho1: Pesquise e apresente um trabalho sobre o algoritmo Naïve-Bayes para a detecção de Spam em mensagens de email ou para classificar páginas de texto como interessante ou não interessante com base em um tema de interesse (esporte, política, ...etc.) presente nas palavras que aparecem nas páginas.

Trabalho 2: Pesquise e apresente um trabalho sobre Random Forest com uma aplicação de livre escolha.