

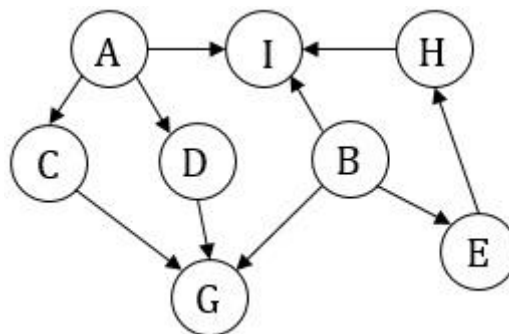
Exercícios de Aprendizado de Máquina

Data de Entrega: 18/11/2019

Não utilize funções prontas de algoritmos aprendidos em sala de aula. Programe as suas e apresente-as na lista. Faça um relatório explicando como foram resolvidos os exercícios e envie junto com o código fonte. Envie a lista para patrick.ciarelli@ufes.br

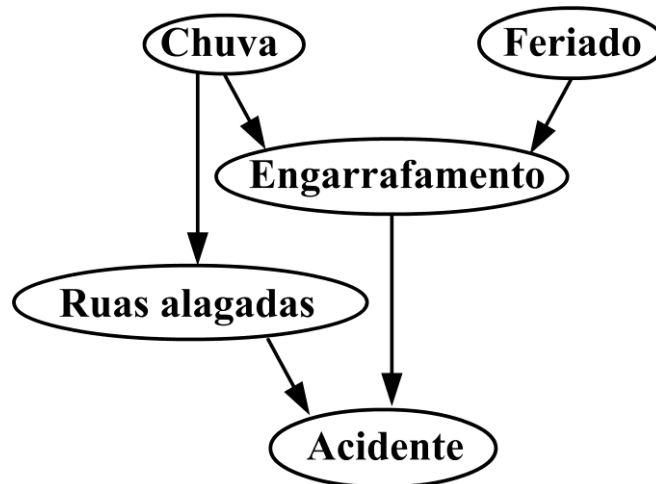
Parte IV – Métodos de Classificação Baseados em Probabilidade

- 1) Para a base Car Evaluation (disponível em <http://archive.ics.uci.edu/ml/>), considerando que o primeiro atributo é x_1 , o segundo é x_2 e assim por diante, estime as probabilidades:
 - a) $P(x_1 = \text{med})$ e $P(x_2 = \text{low})$
 - b) $P(x_6 = \text{high} | x_3 = 2)$ e $P(x_2 = \text{low} | x_4 = 4)$
 - c) $P(x_1 = \text{low} | x_2 = \text{low}, x_5 = \text{small})$ e $P(x_4 = 4 | x_1 = \text{med}, x_3 = 2)$
 - d) $P(x_2 = \text{vhigh}, x_3 = 2 | x_4 = 2)$ e $P(x_3 = 4, x_5 = \text{med} | x_1 = \text{med})$
- 2) Aplique o *Naive Bayes* sobre a base de dados Monk's Problems (disponível em <http://archive.ics.uci.edu/ml/>). Obtenha a acurácia, treinando com `monks-2.train` e testando em `monks-2.test`. Realize os experimentos:
 - a) Considerando uma distribuição Gaussiana dos atributos;
 - b) Discretizando os valores em intervalos de tamanho 1;
 - c) Discretize os valores da mesma forma que em b) usando a suavização de Laplace.
- 3) Para a rede bayesiana da figura abaixo, verifique as seguintes afirmações, indicando se é falso ou verdadeiro e fornecendo a devida explicação.



- a) A é independente de B
- b) A é independente de B tal que foi observado G
- c) C é independente de B tal que foi observado H
- d) G é independente de E tal que foi observado B e I
- e) G é independente de I tal que foi observado A e B

- 4) Dada a rede bayesiana abaixo e a base de dados trânsito.txt, obtenha (na base de dados os valores 0 indicam que o evento não aconteceu, enquanto 1 aconteceu, sendo C chuva, F feriado, E engarrafamento, R ruas alagadas e A acidente):



- As probabilidades condicionais e a priori de cada nó necessários para o uso da rede.
- A probabilidade de acontecer acidente tal que foi observado ruas alagadas ($P(\text{Acidente} = 1 \mid \text{Ruas alagadas} = 1)$)
- A probabilidade de ser feriado tal que foi observado chuva e engarrafamento ($P(\text{Feriado} = 1 \mid \text{Chuva} = 1, \text{Engarrafamento} = 1)$)

Parte V – Métodos de Classificação Baseados em Procura

- Para a base Nursery (disponível em <http://archive.ics.uci.edu/ml/>):
 - Construa uma árvore de decisão com dois níveis de nó de decisão (isto é, o primeiro nó de decisão (primeiro nível), os nós de decisão abaixo dele (segundo nível) e em seguida os nós folha) usando a medida de Ganho de Informação. Selecione aleatoriamente 10000 amostras dos dados para treinamento que serão usados para construir a árvore. Retorne a estrutura da árvore construída.
 - Use os restantes dos dados para avaliação. Retorne a acurácia obtida.
 - Tente obter as regras de decisão a partir da árvore construída.
- Para a base Servo (disponível em <http://archive.ics.uci.edu/ml/>):
 - Construa uma árvore de regressão com dois níveis de nó de decisão (isto é, o primeiro nó de decisão (primeiro nível), os nós de decisão abaixo dele (segundo nível) e em seguida os nós folha) usando a medida de redução de desvio

- padrão. Selecione aleatoriamente 75% dos dados para treinamento que serão usados para construir a árvore. Retorne a estrutura da árvore construída.
- b) Use os restantes 25% dos dados para avaliação. Retorne as medidas MAPE e RMSE.
- c) Tente obter as regras de decisão a partir da árvore construída.

Questões Teóricas

- 1) Prove porque uma variável constante não tem efeito em um classificador naïve Bayes.
- 2) Em uma residência há um alarme que detecta incêndio e soa um alarme. Considere as variáveis booleanas A (o alarme soa), AD (alarme defeituoso), F (fumaça de cigarro), I (incêndio), T (temperatura alta), S (sensor de temperatura) e SD (sensor defeituoso). Sabendo que o sensor mede o excesso de temperatura (causado pelo incêndio) e nesta condição aciona o alarme, que também pode ser acionado pela fumaça, trace uma rede bayesiana para esse domínio, e justifique o modelo selecionado.
- 3) Como pode ser usada uma árvore (de regressão ou de decisão) para avaliar uma amostra quando ela possui uma ou mais variáveis faltantes?
- 4) Explique o algoritmo *Random Forest*.