

Lista #1

Eric Azevedo de oliveira

22/02/2022

Questão 01

Descreva o processo de KDD, explicando cada etapa e o que normalmente é realizado em cada um delas.

R:

- 1) Domain Understanding : É o ato da aprendizagem e estudo relacionados a base de dados que será abordada.
- 2) Selection : Ato de fazer uma seleção dos atributos relevantes da base de dados.
- 3) Preprocessing Data Cleaning : Fazer o processamento dos dados em níveis diferentes .
- 4) Transformation : Quando os dados são transformados em apenas um dado para que ele possa ser utilizado para retirar as informações.
- 5) Data Mining: Conhecer os algoritmos e executa-los para minerar os dados.
- 6) Evaluation and interpretation :Ato de analisar e interpretar os dados obtidos, sendo possível se necessário refazer as etapas anteriores.
- 7) Discoveres Knowledge : Descoberta de um ou mais conhecimentos que possam ser utilizados no processo.

Questão 02

Quais são os principais problemas de aprendizado de máquina existentes? Explique e forneça exemplos.

R:

- 1)Regressão: Ele tem sem princípio ativo interligado ao da classificação, mas os atributos que serão classificados serão numéricos, irão prever um número de sorteio
- 2)Classificação :É um dos tipos mais utilizados, sendo assim responsável pela classificação dos tipos são problemas nominais no caso se um animal for macho ou femea.

3) Agrupamento: Nesse caso não terá um tributo de classificação para conseguir se prever, sendo responsável no ato de agrupar os atributos com uma maior compatibilidade. Ex agrupar perfil de uma sala.

4) Regras de associação: É utilizado para buscar atributos que tem uma semelhança, se um Cliente utiliza uma marca específica ele tem a probabilidade de gostar de vários produtos da mesma marca

Questão 03

Quais são os principais métodos de aprendizado existentes? Explique e forneça exemplos.

R:

1) Deep Learning: Ele consegue reconhecer situações que tem uma ordem de complexidade maior, sendo um exemplo as redes neurais.

2) Supervisionado :É um método que tem a opção de ser supervisionado tendo em si o atributo da classificação como a árvore de decisão.

3)Não Supervisionado: É o contrário do supervisionado, não sendo possível a supervisão e não tem o atributo de classificação, como os mapas auto-organizáveis.

4)Semi Supervisionado: ela tem apenas um dos seus dados rotulados, sendo possível encontrar em um algoritmo SVM.

5) Reforço: É responsável pela maximização das recompensas do resultado , como nas Redes neurais

Questão 04

Considerando-se a base de dados sobre “Esperar ou não pelo restaurante” (verificar base de dados disponibilizada no CANVAS), pede-se:

1) Calcular o ganho de informação de cada atributo. Que atributo é a raiz da árvore?

R:

Ganho

Bar:0

Sex/Sab: 0.02

Fome: 0.19

Cliente: 0.54

Preço: 0.52

Chuva: 0.02

Res: 0.02

Tipo: 0

Tempo: 0.2

Com base nos dados da base de dados o atributo que será definido com raiz será Cliente, pois ele causa a maior influência da árvore.

2) Que atributo estará no segundo nível da árvore. Faça os cálculos e apresente a árvore gerada.

R:

Ganho

Bar:0

Sex/Sab: 0.01

Fome: 0.25

Cliente: 0.0

Preço: 0.25

Chuva: 0.04

Res: 0.25

Tipo: 0.25

Tempo: 0.25

Como temos 5 atributos com o mesmo peso (Preço, Chuva, Res, Tipo, Tempo) teremos que olhar o mais simples que nesse contexto seria Fome, pois ele só apresenta 2 tipos de variância Sim e Não.

3) Quais as regras obtidas a partir desta árvore? Qual a cobertura de cada regra?

R: Após obter o Cliente como raiz partimos para três conjuntos de regras em si, sendo elas, se algum cliente a sua resposta será sim, e se tiver nenhum a resposta será não, se tiver cheio irá partir para outro conjunto de regras que serão resultados de Fome, se for sim irá dar sim e se for não irá dar não conforme o gráfico apresentado abaixo.

Cobertura

Cliente	Cg	CpC
Sim	0.3	0.6
Não	0.1	0.3

Fome

Sim	0.1	0.3
Não	0.1	0.3

Questão 05

a) Utilizando-se a base de dados “Esperar ou não pelo restaurante”, altere os parâmetros indicados na figura abaixo. Mande gerar a árvore novamente e veja o que acontece. Explique.

PS: Para alterar algum parâmetro da função, clique sobre o nome da função, ao lado da função “Choose”.

R: Irá acontecer uma mudança na estrutura base da árvore, além disso alguns valores irão ser alterados.

b) Investigue o significado dos parâmetros ConfidenceFactor e NimNumObj. A opção More explica cada parâmetro.

R : CF: É responsável pelas podas feitas na árvore (tirar os atributos menos importantes).

MNO: É responsável pelo número mínimo de instâncias por uma folha na árvore.