

Lab. 3 - Aprendizado de Máquina Probabilístico

CMC-13 Introdução a Ciência de Dados

(Trabalho Individual ou em Grupo de dois ou três alunos)

Prof. Paulo André Castro

1. Objetivo

Exercitar e fixar conhecimentos adquiridos sobre Ciência de Dados e preparação de dados utilizando uma base de dados fornecida.

2. Descrição do Trabalho

2.1. Base de dados (dataset)

O dataset incluem dados de votações no congresso dos EUA para 14 questões relevantes e a informação se o congressista era republicano ou democrata. Os votos são simples sim (y) ou não (n). Valores faltantes, o que incluem faltas ou abstenções na votação estão representados como '?'. O dataset está disponível no arquivo 'republican_democrat.csv', no Google Classroom. Mais informações sobre o dataset estão disponíveis no arquivo republican_democrat.txt. Há um arquivos de dados que deve ser us

2.2. Tarefas a Realizar

1. Modelo probabilístico baseado em Nãive Bayes ou outras formas de Redes Bayesianas

Crie um modelo probabilístico baseado em Bayes Ingênuo ou outro padrão de rede bayseiana para estimar a probabilidade de um indivíduo ser republicano dados seus votos nas votações indicadas como atributos. Observe que talvez não haja informação para todas as votações.

2. Criação de Ranking

Utilizando o modelo probabilístico do item anterior, crie um ranking usando os dados do arquivo 'republican_democrat_test.csv'. Observe que este arquivo tem a mesma estrutura do dataset, exceto pelo atributo Target (democrat/republican) ter sido substituído por um atributo identificador do indivíduo (ID). O ranking deve ordenar os identificadores ID em ordem decrescente de probabilidade do indivíduo ser republicano.

3. Material a ser Entregue e Prazo

Devem ser entregues um relatório e um notebook com o código

Entregar através do Google Classroom!

OBS: Não compacte os arquivos em um zip (ou qq outro formato), faça os uploads dos dois arquivos!

A. Relatório em formato pdf (ver detalhes abaixo)

B. Código em formato Notebook (ve detalhes abaixo)

Prazo de Entrega: 17/junho/2022;

Item A: Relatório

Estrutura do Relatório do Projeto (arquivo em formato pdf)

Título: Lab. 3 - Aprendizado de Máquina Probabilístico

Equipe: Nomes do membros da Equipe

1. Modelo Probabilístico

Descrever modelo, procedimentos realizados para concluir esta tarefa

2. Criação de RankingExploratória

Descrever procedimentos e o ranking elaborado nesta tarefa

4. Conclusões: Comentários e sugestões sobre o trabalho (complexidade/facilidade, sugestões, etc.).

Item B: Código do Projeto

Código do Projeto (Formato jupyter notebook, Linguagem: Python , R ou Julia)

Siga a estrutura do relatório, para organizar o código no notebook

1. Modelo Probabilístico

Códigos relacionados

2. Criação de RankingExploratória

Códigos relacionados

Obs:

Use células markdown antes de cada célula de código, para descrever qual o propósito do código em seguida

Bom Trabalho!
Prof. Paulo André Castro