

Lab. 3 - Aprendizado de Máquina Probabilístico

CMC-13 Introdução a Ciência de Dados

(Trabalho em Grupo de dois, três ou quatro 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

Deve ser entregue um jupyter notebook (formato .ipynb) com descrição, comentários e código-fonte

OBS: Entregar através do Google Classroom! Não compacte o arquivos em um zip (ou qq outro formato),!

A. Notebook com descrição, comentários e código-fonte (ver detalhes abaixo)

Prazo de Entrega: 26/junho/2023;

Estrutura do Notebook

OBS:(arquivo em jupyter Notebook) Intercalar células(tags) de texto e código fonte observando estrutura indicada abaixo. Comentar o código nas célula de código

Título: Lab. 3 - Aprendizado de Máquina Probabilístico

Equipe: Nomes do membros da Equipe

1. Modelo Probabilístico

Descrever modelo, procedimentos realizados para concluir esta tarefa

Códigos relacionados

2. Criação de RankingExploratória

Descrever procedimentos e o ranking elaborado nesta tarefa

Códigos relacionados

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

Bom Trabalho!

Prof. Paulo André Castro