

Trabalho Final

Dia 1 - Apresentação das propostas

Abaixo seguem as propostas apresentadas, por ordem de envio.

Grupo 1

Banco de Dados: 'Default'

Problema: Newton Raphson estimação de proporção de maus pagadores

Integrantes: Gabriel, Alisson e Veronica

Grupo 2

Metodologia:

a. Vamos criar um modelo de regressão (logística) pra definir qual a variável do banco mais influencia na proporção de pessoas com dívidas

b. Depois de descobrir qual a variável mais influente, vamos criar uma função não linear(a partir da logística) onde: $Y(\text{proporção de dívida}) = \text{variável influente}_1 + \text{variável influente}_2 + \dots$

c. Vamos usar um dos algoritmos indicados para maximizar essa função(ou seja, fazer a verossimilhança)

d. vamos comparar os resultados com o estimador de máxima verossimilhança(tradicional), que já é sabido para a variável resposta (bernoulli)

Integrantes: Rafaela Galetto, Gustavo Utpott e Rafael Chaves

Grupo 3

Banco de Dados: Credit Card Fraud Detection - Mostra variáveis referentes à compras em cartões de crédito e a classificação se foram feitos de cartões fraudulentos ou não.

Banco de dados obtidos no kaggle: <https://www.kaggle.com/mlg-ulb/creditcardfraud>

Variável Resposta: Classificação do cartão em Fraudulento ou não. (Distribuição Bernoulli)

Método de Estimação/Otimização:

Regressão logística (Estimação dos parâmetros por máxima verossimilhança utilizando newton-raphson)

Integrantes: Vitor Borges e Leonardo Almeida

Grupo 4

Dados: Iremos utilizar o banco de dados de passageiros do Titanic: <https://www.kaggle.com/c/titanic/data>.

Problema: Temos dados como idade, sexo, cabine que viajou etc e queremos estimar baseado nestes dados se um passageiro sobreviveu ou não ao acidente do títanic.

Variável resposta: Se o passageiro sobreviveu ou não

Método: Primeiramente iremos fazer descritivas para observar a importância de cada variável na previsão, escolhendo uma delas para fazermos uma previsão inicial (Supomos primeiramente utilizar a idade, em vista que crianças eram resgatadas primeiro). Após, iremos utilizar regressão logística trabalhando com a função de máxima verossimilhança e otimizando pelo método de Newton-Raphson. Buscaremos explicitar os cálculos e implementar o Newton-Raphson de forma aberta, comparando com as funções já existentes nas linguagens de programação.

Integrantes: Roger Moreira da Silva e Leonardo Zarth

Grupo 5

Problema:

Dois problemas de seleção de modelos e estimação:

* Dados sobre intervalo entre passagem de carros em uma rodovia

* Experimento do pote de bolas do UFRGS Portas Abertas

+ Obter o modelos mais adequados para estes dados.

Método de Estimação/Otimização: AIC, Máxima verossimilhança e Newton-Raphson

Aplicar o AIC para testar alguns modelos como $Gamma(a, b)$, $Normal(u, s^2)$, distribuição histograma com diferentes números de “passos”, para um conjunto de dados referentes a intervalo entre a passagem de dois carros em uma rodovia.

Aplicar o AIC também para os dados do experimento do pote (distribuição empírica e alguma multimodal).

Integrantes: Lucas da Rocha Schwengber

Grupo

Integrantes: ?

Grupo

Integrantes: ?

