

EST171 - APRENDIZADO DE MÁQUINA
Departamento de Estatística
Universidade Federal de Minas Gerais

Lista 2

Henrique Aparecido Laureano Matheus Henrique Sales

Outubro de 2016

Sumário

Exercício I

2

Exercício I

Baixe o conjunto de dados `titanic.txt`. Cada observação deste banco é relativa a um passageiro do Titanic. As covariáveis indicam características destes passageiros; a variável resposta indica se o passageiro sobreviveu ou não ao naufrágio.

Seu objetivo é criar classificadores para prever a variável resposta com base nas covariáveis disponíveis. Para tanto, você deverá implementar os seguintes classificadores, assim como estimar seus riscos via conjunto de teste:

- Regressão Logística. Mostre os coeficientes estimados.
- Regressão Linear. Mostre os coeficientes estimados.
- Naive Bayes.
- Análise Discriminante Linear.
- Análise Discriminante Quadrática.
- KNN. Para isso você precisará transformar as covariáveis categóricas em numéricas. Você pode usar variáveis dummies .

Responda ainda as seguintes perguntas:

- Qual o melhor classificador segundo o risco estimado? Discuta.
- Para os classificadores baseados em estimativas de probabilidade, faça também as curvas ROC com o conjunto de teste. Faça também a tabela de confusão quando o corte usado é 0.5 e também quando o corte é aquele que maximiza sensibilidade mais especificidade. Comente.

Regressão Logística

```
# <code r> ===== #  
path <- "C:/Users/henri/Dropbox/Scripts/aprendizado de maquina/list 2/"  
# </code r> ===== #
```

Regressão Linear

Naive Bayes

Análise Discriminante Linear

Análise Discriminante Quadrática

KNN

Qual o melhor classificador segundo o risco estimado?

Para os classificadores baseados em estimativas de probabilidade, faça também as curvas ROC com o conjunto de teste. Faça também a tabela de confusão quando o corte usado é 0.5 e também quando o corte é aquele que maximiza sensibilidade mais especificidade
