
CLASSIFICADOR BAYESIANO MULTIVARIADO - HEART

Gabriel Saraiva Espeschit - 2015065541

19 de agosto de 2020

Para classificação da base de dados *Heart*, tomou-se os seguintes passos:

1. Leitura dos dados *Heart* da seguinte forma:
df_heart = pandas.read_csv('heart.dat', header=None, delim_whitespace = True)
2. Divisão dos dados em dados em 90% para treino e 10% teste utilizando a função própria *train_test*^[1].
3. Cálculo das médias, desvios padrões, correlações e probabilidades marginais para cada classe do grupo de treino utilizando a função *media_dp_cor_pm*^[2].
4. Alimentação dos dados calculados acima na função de classificação para obter a classe sob os casos de teste usando a função *classe*^[3].
5. Cálculo da porcentagem de erros e acertos da classificação obtida por meio da função *classe* em relação aos dados originais.
6. Repetição dos passos 3, 4 e 5 porém com 70% de dados de teste e 30% de treino.
7. Repetição dos passos 3, 4 e 5 com 20% de dados de teste e 80% de treino.

No primeiro caso (90%, 10%) podemos verificar que o desempenho nos dados de teste foram melhor que o esperado, podendo ter ocorrido um *overfitting*:

Dados de teste:

Número de acertos: 25

Números de erros: 2

Porcentagem de acertos: 92.59%

Dados de treino:

Numero de acertos: 210

Números de erros: 33

Porcentagem de acertos: 86.42%

No segundo caso (70%, 30%) podemos verificar que o desempenho nos dados de teste foram conforme o esperado:

Dados de teste:

Número de acertos: 72

Números de erros: 9

Porcentagem de acertos: 88.89%

Dados de treino:

Numero de acertos: 169

Números de erros: 20

Porcentagem de acertos: 89.42%

No terceiro caso (20%, 80%) podemos verificar que o desempenho nos dados de teste foram abaixo do esperado, e ocorreu um *underfitting*:

Dados de teste:

Número de acertos: 158

Números de erros: 48

Porcentagem de acertos: 77.78%

Dados de treino:

Numero de acertos: 52

Números de erros: 2

Porcentagem de acertos: 96.3%