**I. Pen-and-paper**

1. Cálculo dos *priors*

Função de densidade de probabilidade de condicionada às classes – Distribuição normal

Função de massa de probabilidade de condicionada às classes

Função de densidade de probabilidade de e condicionadas às classes – Distribuição normal multivariada

=

=

Utilizando a Regra de Bayes concluímos que:

Logo, utilizando o MAP obtemos que será classificado como 1 se:

1. Calculando as probabilidades condicionadas às classes dados determinados input obtemos,

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  | Classe Prevista | Classe Real |
|  |  |  |  | N | N |
|  |  |  |  | P | N |
|  |  |  |  | N | N |
|  |  |  |  | P | N |
|  |  |  |  | N | P |
|  |  |  |  | P | P |
|  |  |  |  | P | P |
|  |  |  |  | P | P |
|  |  |  |  | N | P |
|  |  |  |  | P | P |

Exemplificação dos cálculos:

será classificado como negativo pois a probabilidade obtida é inferior a 50%, sendo que na realidade pertence à classe dos negativos. Trata-se assim de um verdadeiro negativo.

Construção da matriz de confusão:

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Valores Reais | |
| P | N |
| Valores Previstos | P | 4 | 2 |
| N | 2 | 2 |

1. O valor da medida F1 é dado por

Sendo que esta medida fica toma valores entre 0 e 1, sendo o valor 1 o caso ideal, podemos concluir que se trata de uma avaliação razoável do modelo treinado.

1. Organizando, por ordem crescente, os valores da probabilidade de as observações pertencerem à classe P, estabelecemos valores intermédios para utilizar como *treshold* e testar a eficácia do modelo em todos os valores possíveis. Os resultados obtidos encontram-se na seguinte tabela, onde as células sombreadas a verde significam uma previsão acertada, ao contrário das células sombreadas a verde.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | *Tresholds* | | | | | | | | | | | Classe Real |
|  | 18 | 23 | 48 | 54 | 63 | 70 | 84 | 89 | 93 | 94 |
|  |  | P | N | N | N | N | N | N | N | N | N | N | N |
|  |  | P | P | N | N | N | N | N | N | N | N | N | P |
|  |  | P | P | P | N | N | N | N | N | N | N | N | N |
|  |  | P | P | P | P | N | N | N | N | N | N | N | P |
|  |  | P | P | P | P | P | N | N | N | N | N | N | P |
|  |  | P | P | P | P | P | P | N | N | N | N | N | N |
|  |  | P | P | P | P | P | P | P | N | N | N | N | N |
|  |  | P | P | P | P | P | P | P | P | N | N | N | P |
|  |  | P | P | P | P | P | P | P | P | P | N | N | P |
|  |  | P | P | P | P | P | P | P | P | P | P | N | P |
| Eficácia(%) | | 60 | 70 | 60 | 70 | 60 | 50 | 60 | 70 | 60 | 50 | 40 |  |

Assim podemos concluir que a eficácia é maximizada em

onde toma o valor de 70%

Assim, o *treshold* normalmente utilizado de 50% não se encontra dentro dos intervalos de maximização pelo que, neste caso não deve ser utilizado para obter os melhores resultados possíveis.

**II. Programming and critical analysis**



1. Answer 6
2. Answer 7
3. Answer 8

**III. APPENDIX**

Paste your programming code here using Consolas 9pt or 10pt.

Use **highlighting** or colored text to facilitate the analysis by your faculty hosts.

**END**

**import** pandas **as** pd

**from** scipy.io **import** arff

**from** plotly.subplots **import** make\_subplots

**import** plotly.graph\_objects **as** go

data = arff.loadarff('breast.w.arff')

df = pd.DataFrame(data[0])

df.dropna(inplace=True)

df.replace(b'benign', 0, inplace=True)

df.replace(b'malignant', 1, inplace=True)

b = df[df["Class"] == 0]

m = df[df["Class"] == 1]

grid = make\_subplots(rows=3, cols=3)

i = 0

**for** col **in** ["Clump\_Thickness","Cell\_Size\_Uniformity","Cell\_Shape\_Uniformity","Marginal\_Adhesion","Single\_Epi\_Cell\_Size","Bare\_Nuclei","Bland\_Chromatin"

,"Normal\_Nucleoli","Mitoses"]:

fig = go.Histogram(x=b[col],nbinsx=10,opacity=0.5,name=col+"|benign")

grid.append\_trace(fig, i//3+1, (i%3)+1)

fig = go.Histogram(x=m[col],nbinsx=10,opacity=0.5,name=col+"|malignant")

grid.append\_trace(fig, i//3+1, (i%3)+1)

i += 1

grid.update\_layout(height=600, width=800, title\_text="Distribuições condicionadas às classes", barmode="overlay", showlegend=True)

grid.show()