

**Trabalho 3**  
**Classificadores Bayesianos, Critério MAP e LDA**  
**Disciplina de Reconhecimento de Padrões**  
**Hugo Silveira Sousa - 378998**

**2.1. Análise inicial dos dados**

**Arquivo:**

**Analise\_inicial\_dos\_dados.m**

**b) - Que informação, sobre os atributos, é obtida através da matriz de correlação?**

A matriz de correlação mostra os valores de correlação dos atributos da base, medindo o grau de relação entre cada par de atributos, é calculado a partir da covariância dos atributos normalizado pelos desvios padrões do par de atributos, tendo valores entre -1 e 1. Se os dois atributos crescem ou diminuem juntos, a correlação é positiva, mais próxima de 1, se quando um cresce o outro diminui, a correlação é negativa, mais próxima de -1, se o crescimento de um não influencia no outro, os atributos são descorrelacionados, com valores próximos de 0, atributos independentes implica em correlação nula, mas correlação nula não implica em atributos independentes.

**- Quais atributos são mais correlacionados?**

Correlação positiva: Considerei todos aqueles com correlação maior que 0.8

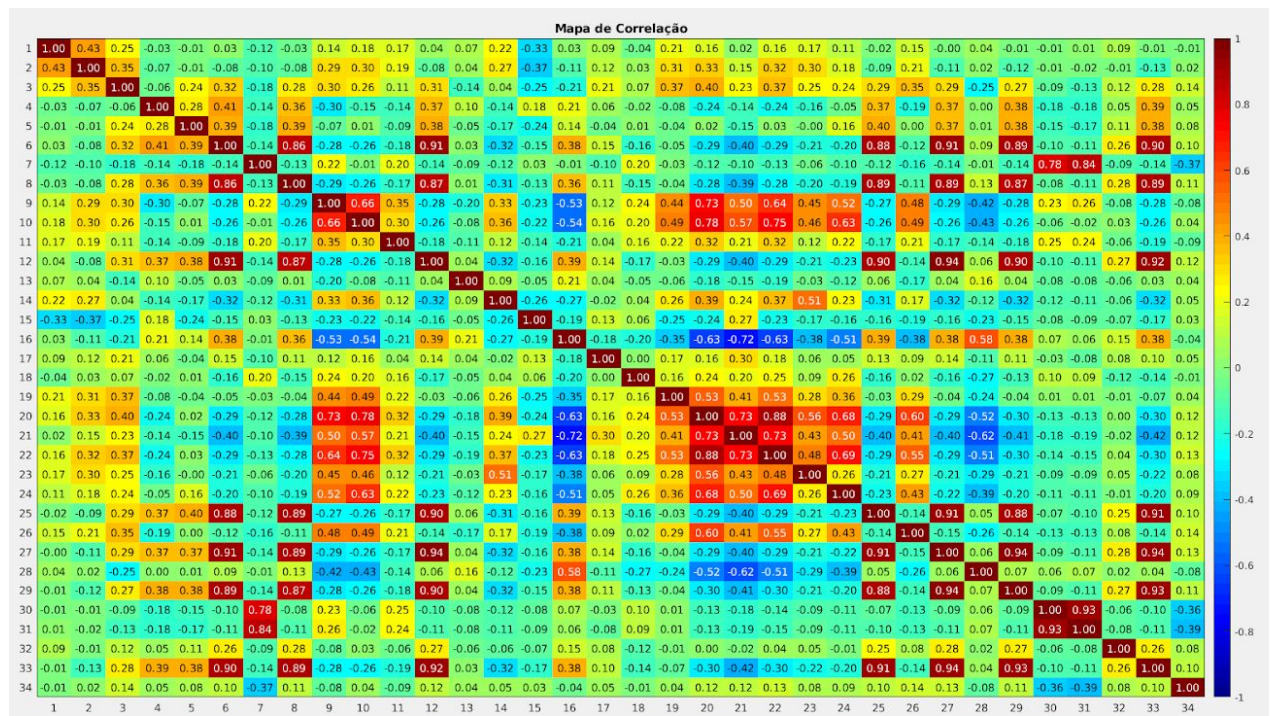
Atributo 1 - Atributo 2 (Correlação)

|                |                |                |
|----------------|----------------|----------------|
| 29 - 33 (0.93) | 27 - 29 (0.94) | 06 - 27 (0.91) |
| 27 - 33 (0.94) | 25 - 29 (0.88) | 12 - 25 (0.90) |
| 25 - 33 (0.91) | 12 - 29 (0.90) | 08 - 25 (0.89) |
| 06 - 33 (0.90) | 08 - 29 (0.87) | 06 - 25 (0.88) |
| 08 - 33 (0.89) | 06 - 29 (0.89) | 20 - 22 (0.88) |
| 12 - 33 (0.92) | 25 - 27 (0.91) | 12 - 08 (0.87) |
| 30 - 31 (0.93) | 12 - 27 (0.94) | 12 - 06 (0.91) |
| 07 - 31 (0.84) | 08 - 27 (0.89) | 08 - 06 (0.86) |

Correlação negativa: Considerei todos aqueles com correlação menor que -0.6

Atributo 1 - Atributo 2 (Correlação)

|                 |                 |                 |                 |
|-----------------|-----------------|-----------------|-----------------|
| 16 - 20 (-0.63) | 16 - 21 (-0.72) | 16 - 22 (-0.63) | 21 - 28 (-0.62) |
|-----------------|-----------------|-----------------|-----------------|



## 2.2) Classificadores Bayesianos

Arquivos:

qda.m

naive\_bayes.m

lda.m

- Qual classificador obteve a melhor taxa de acerto (precisão, acurácia) média?

- Em todos os algoritmos a base foi permutada aleatoriamente com a mesma semente(0) para a função rand.

- Média das Acurácias (K-fold=5):

1. QDA: 96.6588% (Desvio Padrão: 2.8782)
2. LDA: 96.6432% (Desvio Padrão: 3.5336)
3. Naive Bayes: 93.0321% (Desvio Padrão: 1.6078)

- Pode-se considerar esse problema linearmente separável? Justifique.

O LDA é o classificador com regiões de decisão do tipo linear, dos três usados, e como está apresentando taxas de acertos elevadas, nesse exemplo com 96.6432% de média de acurácia, na prática, podemos considerar esse problema linearmente separável.

## 2.3) LDA (CDA) como transformação linear.

Arquivos:

cda.m

- A cada iteração a base é permutada aleatoriamente, a semente do rand foi fixada em 0, para manter os mesmo resultados em todos os experimentos.
- Treinamento dos dados projetados foi realizado com QDA.
- Foi feito a normalização z-score antes da transformação linear, e depois da transformação linear, antes do classificador QDA.

- Taxa de Acurácia Média (50 iterações):

- Considerando os maiores autovalores para criação da matriz de projeção W

- 33 autovalores: 95.6667% (Desvio padrão: 1.8466)
- 32 autovalores: 95.3148% (Desvio padrão: 1.8931)
- 31 autovalores: 94.9074% (Desvio padrão: 1.7373)
- 30 autovalores: 94.6296% (Desvio padrão: 1.8893)
- 29 autovalores: 94.5185% (Desvio padrão: 1.9245)
- 28 autovalores: 94.5185% (Desvio padrão: 1.9783)
- 27 autovalores: 94.5000% (Desvio padrão: 1.5244)
- ...
- 20 autovalores: 94.3148% (Desvio padrão: 1.8801)
- ...
- 15 autovalores: 93.5556% (Desvio padrão: 2.5441)
- ...
- 10 autovalores: 91.0370% (Desvio padrão: 3.0015)
- ...
- 05 autovalores: 84.1852% (Desvio padrão: 4.3140) **(C-1)**
- 04 autovalores: 79.9444% (Desvio padrão: 5.0839)
- 03 autovalores: 74.1481% (Desvio padrão: 6.5120)
- 02 autovalores: 64.5741% (Desvio padrão: 7.4892)
- 01 autovalor: 55.5556% (Desvio padrão: 6.8195)