

## Exercício 2

---

Prof. A.P. Braga

18 de abril de 2022

### METRIC LEARNING + KNN

Este trabalho consiste na implementação de um classificador K-Nearest Neighbours (KNN) que utilize uma métrica de Mahalanobis, cujos parâmetros são obtidos através de um processo de otimização.

As classes a serem separadas devem ser geradas através do seguinte código:

```
if (!require(mvtnorm))
{
  install.packages("mvtnorm")
  library(mvtnorm)
}
n1 <- 140
n2 <- 60
xc1 <- rmvnorm(n1, mean = c(3,3), sigma = matrix(c(5,0,0,0.2), nrow = 2, byrow = T))
xc2 <- rmvnorm(n2, mean = c(3,4), sigma = matrix(c(5,0,0,0.2), nrow = 2, byrow = T))

X <- rbind(xc1, xc2)
Y <- c(rep(1, n1), rep(2, n2))
```

Como função objetivo para a otimização, busca-se maximizar a **Silhueta** média do conjunto de dados. ([https://en.wikipedia.org/wiki/Silhouette\\_\(clustering\)](https://en.wikipedia.org/wiki/Silhouette_(clustering)))

Devem ser observados o efeito da métrica sobre o erro de teste do classificador e sobre as distâncias relativas do conjunto de dados, comparando-se a métrica aprendida com a métrica euclidiana. Para esta verificação, devem ser criadas para cada métrica uma

Matriz de Distâncias  $D_{N \times N}$  onde o elemento  $d_{ij}$  corresponde à distância entre o elemento  $i$  e o elemento  $j$ . Tal matriz deve ser construída tanto para a distância euclidiana quanto para a distância aprendida, e ambas devem ser comparadas visualmente através da função *image()*, verificando-se o efeito da métrica.

A seguir, um exemplo de código para visualização da Matriz de Distâncias utilizando a distância euclidiana.

```
dist_euclidean <- as.matrix(dist(X))  
image(dist_euclidean)
```

Um relatório deve ser gerado descrevendo o que foi feito, como foi feito, mostrando os gráficos, as informações relevantes e explicando os resultados obtidos.