

## MAE 5905: Introdução à Ciência de Dados

Lista 3. Primeiro Semestre de 2025. Entregar 22/05/2025.

1. Para o conjunto de dados Iris, use somente o comprimento de pétalas ( $X_1$ ) e comprimento de sépalas ( $X_2$ ) como preditores e a variável resposta  $Y$  = espécie (Setosa, Versicolor, Virgínica). Construa uma árvore para classificação. Escreva com detalhes as regiões no plano e faça o gráfico da árvore e das regiões, usando o pacote de sua preferência. Obtenha a taxa de erro de classificação.
2. Considere o conjunto de dados **rehabcardio**, sendo preditores  $X_1$  =HDL,  $X_2$ =LDL,  $X_3$  =Trigl,  $X_4$ =Glicose e  $X_5$ =Peso e resposta  $Y$ =Diabete (presente=1, ausente=0). Utilize um subconjunto em que as amostras têm todas as medidas completas. Construa árvores usando bagging e floresta aleatória. Usando a taxa de erro de classificação, escolha o melhor classificador.
3. Dadas as observações abaixo ( $n = 7$ ,  $p = 2$ ), faça um gráfico no plano ( $X_1, X_2$ ).
  - (a) Use o classificador de margem máxima e dê a regra de classificação. Obtenha a equação do hiperplano separador.
  - (b) No gráfico, indique a margem e os vetores suporte para o CMM.
  - (c) Calcule o valor da margem. O que acontece com o classificador se modificarmos um pouco a sétima observação?

Obs.	$X_1$	$X_2$	$Y$
1	3	4	Azul
2	2	2	Azul
3	4	4	Azul
4	1	4	Azul
5	2	1	Vermelho
6	4	3	Vermelho
7	4	1	Vermelho