O conjunto de dados Iris foi usado em 1936 no artigo de R.A. Fisher: The Use of Multiple Measurements in Taxonomic Problems (O Uso de Múltiplas Medições em Problemas Taxonômicos). Esse conjunto de dados inclui 50 amostras de cada espécie.

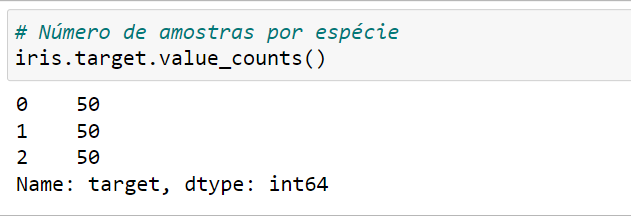


Figura 1 - Quantidade de amostras por espécie (0 - setosa, 1 - versicolor e 3 - virginica)

Duas espécies não são linermente separáveis uma da outra, mas uma espécie é linearmente separável das outras duas.

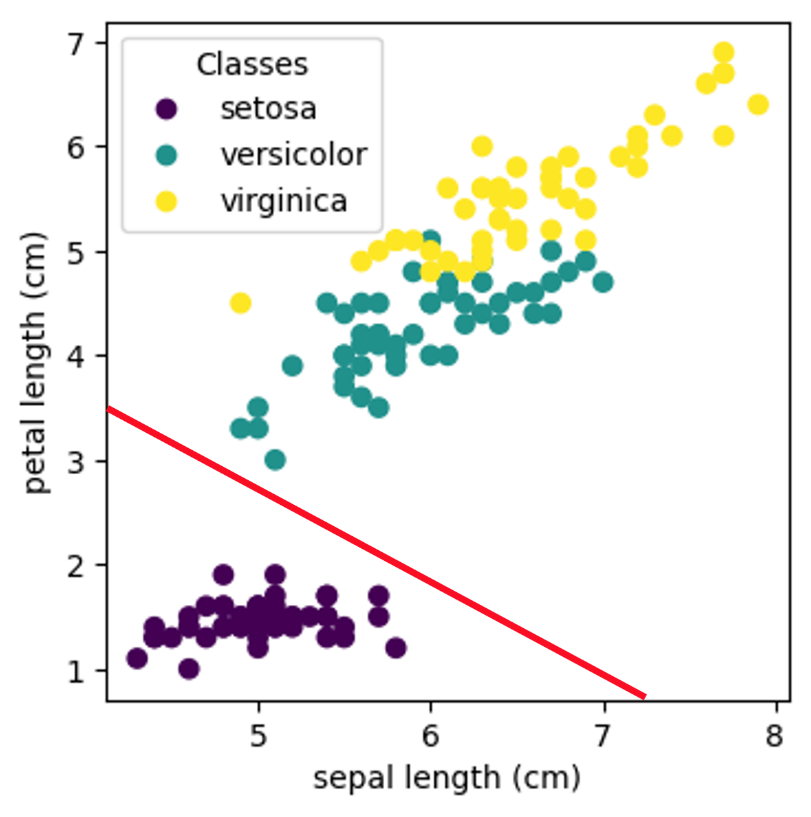


Figura 2 - Demonstrando graficamente que uma espécie é linearmente separável

As features (características usadas para fazer as previsões) são sepal length (comprimento da sépala), sepal width (largura da sépala), petal length (comprimento da pétala) e petal width (largura da pétala), todas as features em centímetro.

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Figura 3 - Features do conjunto de dados

É interessantes adicionar ao dataframe anterior (figura 3) colunas que identifica a espécie, conforme foi feito na figura 4.

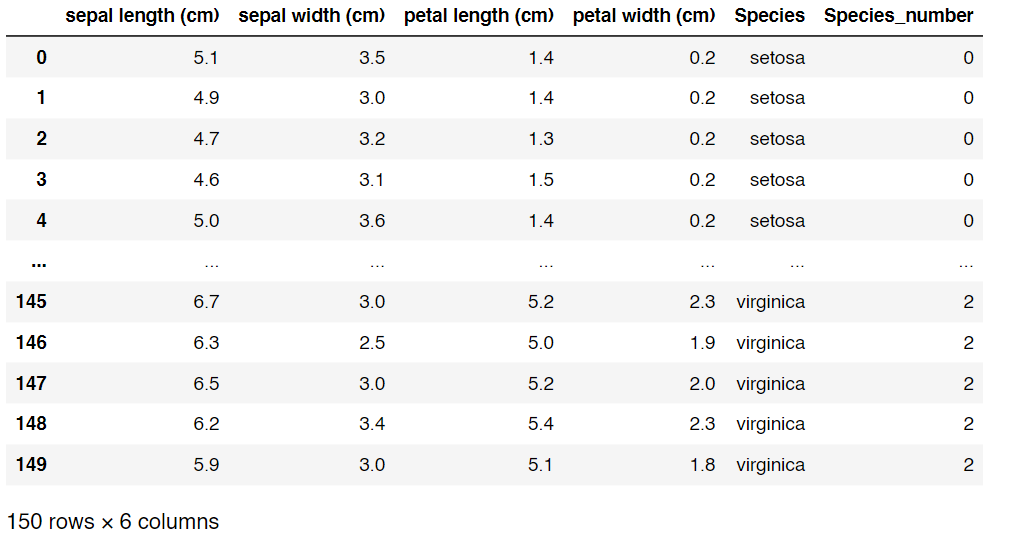


Figura 4 - Dataframe com a adição das colunas Species e Species\_number

Cada espécie pode assumir um determinado valor dentro de um intervalo de valores contidos em cada feature do dataframe, conforme se pode ver nas figuras 5 e 6.

Uma imagem com texto, captura de ecrã, Gráfico, diagrama

Descrição gerada automaticamente

Figura 5 - Intervalo de valores por espécie em cada feature (gráfico de dispersão)

Uma imagem com origami

Descrição gerada automaticamente

Figura 6 - Intervalo de valores por espécie em cada feature (gráfico de violino)

Três modelos de aprendizado de máquina foram treinados a fim de fazer predições: regressão logística, KNN e K-means. As acurácias obtidas no conjunto de teste são respectivamente: 95,56%, 100% e 100%. Abaixo, seguem os gráficos que demonstram onde o modelo errou.

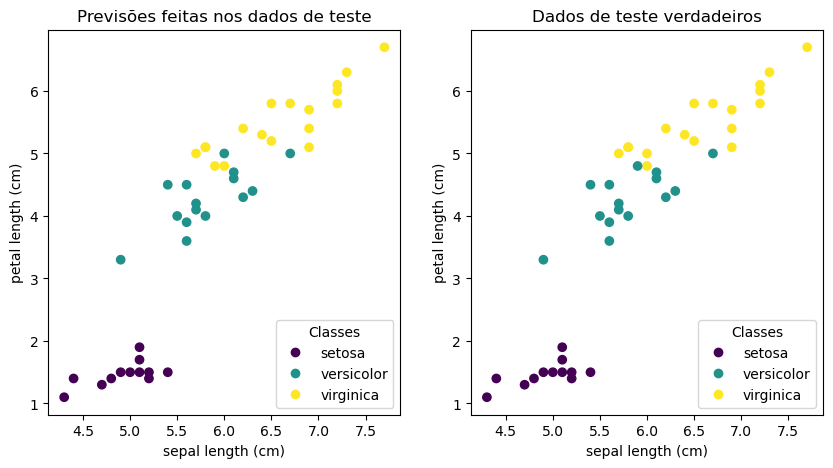


Figura 7 - Acertos e erros cometidos pela regressão logística

Uma imagem com texto, captura de ecrã, diagrama, Gráfico

Descrição gerada automaticamente

Figura 8 - O modelo KNN acertou todas, conforme se pode comprovar graficamente

Uma imagem com texto, captura de ecrã, diagrama, Gráfico

Descrição gerada automaticamente

Figura 9 - O modelo K-means acertou todas, conforme se pode comprovar graficamente