

## Usando **python** & **sklearn**

### 1) Carregar o Iris dataset

- 3 classes (Versicolor, Setosa & Virginica)
- 4 atributos numéricos (Sepal Length, Sepal Width, Petal Length, Petal Width)



### 2) Usando o Iris dataset, escolha o melhor $k$ para o K-NN, `KNeighborsClassifier` do `sklearn.neighbors`

#### 2.1) usando 30 execuções em 60% / 40% (train / test)

- existe diferença estatística entre o resultado obtidos pelo classificador com o  $k$  de mais alta acurácia média com relação aos demais?

#### 2.2) usando validação cruzada de 5 *folds*

- existe diferença estatística entre o resultado obtidos pelo classificador com o  $k$  de mais alta acurácia média com relação aos demais?

Para cada um dos cenários ( (2.1) e (2.3) ), plote o gráfico acurácia vs  $K$

### 3) Carregar o Digits dataset em: <http://www.inf.ufpr.br/menotti/ci171-182/src/knn-digits.zip>

- 10 classes (0-9) e 2000 amostras/exemplos
- Atributos - investigue `digits.py`

#### 3.1) Gere diferentes representações (variando $X, Y$ ) e determine aquela que apresenta a melhor acurácia (classe totalmente balanceada) para os cenários 2.1 e 2.2

### 4) Usando o Digits dataset, compare a eficácia (acurácia) e eficiência (tempo) do K-NN original (NearestNeighbors) com algoritmos aproximados

#### 4.1) usando `KDTree` de `sklearn.neighbors`

#### 4.2) `Balltree` de `sklearn.neighbors`

5) Com relação aos algoritmos de Concept Learning e o conceito apresentado abaixo

5.1) Execute o algoritmo Find-S e apresente a hipótese gerada

5.2) Execute o algoritmo Eliminate Candidate e apresente o Espaço Versão gerado

Size	Color	Shape	Label
big	red	circle	No-
small	red	triangle	No-
small	red	circle	Yes+
big	blue	circle	No-
small	blue	circle	Yes+

6) Elabore

6.1) um relatório (relatorio.txt) com até 150 palavras discutindo os resultados obtidos em 2) e 4)

6.2) uma solução (solucao.txt) para o exercício em 5)

7) Entregue os códigos, relatório e solução em um único zip pelo moodle

- <http://moodle.c3sl.ufpr.br/>
- Cursos: CI171 - Aprendizado de Máquinas (CI171-182)
- **Não serão aceitas entregas em atraso**