

Atividade 1 - Desenvolvimento de um Classificador Binário por Limiar

Prof. Dr. Juliano Henrique Foleis

Descrição da Atividade

Nesta atividade você vai implementar um classificador binário por limiar, conforme fizemos na Semana 2. A implementação deve ser feita em Python em um caderno no Jupyter.

Você vai trabalhar com um subconjunto da base de dados “Iris”. Esta base de dados é muito famosa e amplamente utilizada em um contexto educativo para estudar aprendizagem de máquina. Esta base de dados contém 4 medidas de flores de 3 espécies diferentes. O subconjunto separado para esta atividade (`iris_2class.csv`) contém instâncias de 2 espécies: *versicolor* e *virginica*. Cada espécie está representada por 50 instâncias.

Documente cada um dos passos indicados a seguir no Jupyter:

1. Calcule as estatísticas descritivas de cada um dos atributos (características) agrupando por espécie;
2. Faça um histograma para cada atributo. Faça barras separadas para cada espécie de flor.
3. Com base nos passos 1 e 2, faça *scatterplots* com os pares de atributos que parecerem ser os mais promissores para classificação das espécies.
4. A partir das observações que fizer nos passos 1–3, escolha o atributo que achar mais adequado para separar as espécies.
5. Faça uma busca exaustiva procurando maximizar a acurácia no conjunto de teste. Use 80% dos dados para o conjunto de treino e 20% para o conjunto de teste.
6. Execute a busca exaustiva implementada no passo 5 com 20 particionamentos diferentes. Reporte a acurácia máxima e mínima obtida. Calcule a acurácia média e o desvio padrão.

Instruções e Entrega

- A maioria dos passos acima estão prontos nos cadernos da Semana 2 disponibilizados no [Github](#).
- Capriche no seu *notebook*: coloque textos explicativos, customize os gráficos, etc. Aproveite para aprender como usar as ferramentas!
- A atividade deve ser feita em um Jupyter Notebook. Você pode usar o *Google Colab* se quiser, mas é necessário entregar o arquivo `.ipynb`.
- A entrega deverá ser realizada via Moodle, na *Atividade da Semana 2*.
- **Prazo para entrega:** 16/07/2021 às 23:55.
- O trabalho é individual.
- Não é permitido alterar o arquivo que contém a base de dados (`iris_2class.csv`)!

BONS ESTUDOS!