

Projeto Final de Inteligência Artificial - EC37C

Graduação em Engenharia da Computação

William Sdayle Marins Silva*

Departamento de Computação, Universidade Tecnológica Federal do Paraná - Cornélio Procópio - Pr

Alexandre Rossi Paschoal†

Departamento de Computação, Universidade Tecnológica Federal do Paraná - Cornélio Procópio - Pr

15 de abril de 2021

1 Trabalho de Machine Learning

Para o trabalho de machine learning pede-se que:

1. Baixe uma das bases do kaggle ou uci listadas no link [Dataset Projeto 1](#)
2. Apresentar matriz de dispersão (scatter plot) entre os parâmetros das amostras do dataset. Tendo na diagonal principal o histograma dos parâmetros e no restante da matriz a dispersão dos dados, conforme apresentado na Figura 1.
3. Apresentar box plot. (Matplotlib)
4. Utilizar os dados para treinamento de modelo de machine learning (KNN, Random Forest e Decision Tree), escolha um dos 3 citados, fica a critério do aluno utilizar mais de um modelo. (Scikit-learn)
5. Apresentar matriz de confusão dos resultados, do conjunto de teste. (Scikit-learn)
6. Apresentar Acurácia, Precisão, Revocação e F1-Score, do conjunto de teste. (Scikit-learn)

*E-mail: williamsilva.2019@alunos.utfpr.edu.br

†E-mail: paschoal@utfpr.edu.br

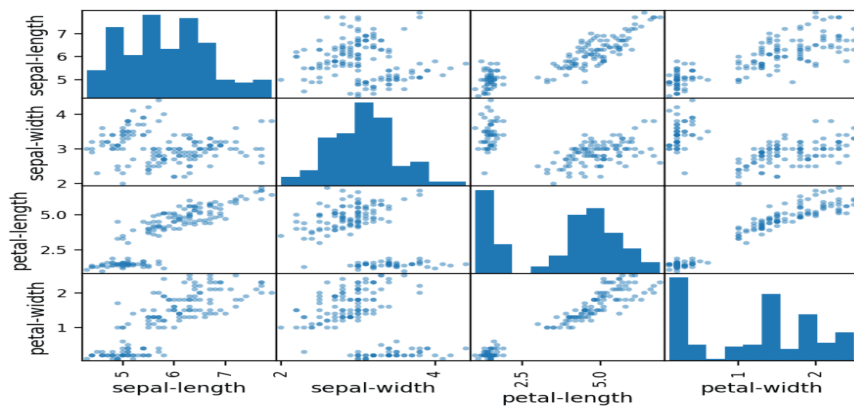


Figura 1: Matriz de dispersão

2 Trabalho de Deep Learning

Neste projeto, serão abordados conceitos de redes neurais convolucionais pré treinadas, ou seja, será utilizado o método de *transfer learning*.

Para o trabalho de Deep Learning pede-se que:

1. Baixe uma das bases listadas no link [Dataset Projeto 2](#)
2. Escolha um modelo pré treinado da literatura, presente no link [Modelos](#). Os pesos pré treinados podem ser da ImageNet.
3. Para extração de características, pode ser utilizado o método de *GlobalMaxPooling*, mas fica a critério do aluno qual método utilizar.
4. A ultima camada da rede, deve ter a mesma quantidade de classes que o dataset utilizado.
5. Como função de ativação da ultima camada, pode ser utilizada a *SoftMax* ou *Sigmoid*, fica a critério do aluno escolher.
6. Como otimizador pode ser utilizado o *SGD* ou *Adam*, que apresentam melhores resultados.
7. Como métrica de avaliação do treinamento, pode ser utilizada a acurácia.
8. Como métrica de erro, pode ser utilizada a *categorical crossentropy* para datasets não binários (com mais de uma classe) e o erro *binary crossentropy* para datasets binários (com apenas uma classe). Podem ser utilizadas outras métricas de erro que estão em [erros](#).

9. Crie um conjunto de treino e teste, sendo 80% das amostras do dataset para treino e 20% das amostras para teste. Não podem haver amostras iguais entre os conjuntos.
10. Para treinamento utilize:
 - Epochs: 500
 - Learning rate: 0.001
 - Batch size: 32
 - Dropout: 0.5
11. Faça uma gráfico de linhas mostrando a curva de aprendizado do modelo utilizado, utilizando a métrica acurácia. Fica a critério do aluno mostrar o gráfico com a otimização do erro.
12. Apresente a matriz de confusão do conjunto de teste.
13. Apresente a acurácia, precisão, revocação e f1-score do conjunto de teste.

Como auxílio para o trabalho 1, utilize o link [Machine Learning](#)

Obs: Não é necessário apresentar as curvas ROC.

Como auxílio para o trabalho 2, utilize o link [Redes Neurais Convolucionais](#)

3 Observações

- O trabalho deve ser feito em grupos.
- Caso o grupo queira, datasets fora da lista apresentada podem ser utilizados.
- Grupos distintos podem utilizar o mesmo dataset, porém, nesse caso, os modelos utilizados devem ser diferentes.
- Será avaliado o conhecimento do grupo como um todo, por isso, pede-se uma apresentação de no máximo 5 minutos sobre o desenvolvimento dos 2 projetos, bem como: dataset, modelo utilizado e como ele se comportou perante aos dados recebidos.
- Na apresentação deve conter os resultados dos modelos, bem como para os modelos de deep learning, deverá haver: gráfico com curva de aprendizado, matriz de confusão e acurácia do conjunto de teste.
- Até 3 grupos podem utilizar o mesmo dataset, caso mais que 3 grupos utilizem o mesmo dataset, será desconsiderado todos os grupos que usaram o mesmo dataset.