

Universidade do Estado do Amazonas

Escola Superior de Tecnologia

Data: 29 de Outubro de 2018

Professora: Elloá B. Guedes

Disciplina: Redes Neurais Artificiais

PROJETO PRÁTICO # 4

Três variedades de trigo (Kama, Rosa e Canadian) possuem sementes muito parecidas, entretanto diferentes. Um grupo de pesquisadores poloneses coletou 70 amostras de cada tipo e, usando uma técnica particular de raio-X, coletou medidas geométricas destas sementes, a citar: área, perímetro, compactude, comprimento, largura, coeficiente de assimetria e comprimento do sulco da semente. O dataset detalhado contendo as medições pode ser encontrado em: <https://archive.ics.uci.edu/ml/datasets/seeds#>

Equipes de três estudantes devem conduzir um processo de Aprendizado de Máquina com tarefa de classificação multiclasse que utilize redes neurais artificiais do tipo *multilayer perceptron* para endereçar esta tarefa. Cada equipe deve produzir um Jupyter Notebook contendo boas práticas de programação que ilustrem os passos detalhados desta tarefa, com justificativas e comentários sempre que necessários. O notebook produzido deve incluir claramente o nome dos integrantes da equipe e deve estar compactado juntamente com o dataset utilizado.

Como discutido nas aulas, uma das principais dificuldades consiste em escolher os parâmetros e hiperparâmetros adequados das redes neurais. Assim, neste projeto prático você deve construir diversas redes utilizando uma *busca em grade* (**GridSearchCV**) a qual deve variar a função de ativação e o número de neurônios nas camadas ocultas. No tocante à este último parâmetro, em particular, utilize a regra da pirâmide geométrica:

$$N_h = \alpha \cdot \sqrt{N_i \cdot N_o}, \quad (1)$$

em que N_h é o número de neurônios ocultos (a serem distribuídos em uma ou duas camadas ocultas), N_i é o número de neurônios na camada de entrada e N_o é o número de neurônios na camada de saída. Teste os valores de α como sendo iguais a 0.5, 2 e 3. Espera-se que cada equipe teste, no mínimo, 30 redes neurais diferentes para este problema.

Além do que foi exposto, a busca em grade deve considerar a realização de uma validação cruzada com $k = 3$ folds e a métrica de desempenho a ser contabilizada será o F -Score médio dentre os folds. A equipe deve apresentar claramente qual o melhor modelo que endereça a tarefa e métricas obtidas nesta tarefa. Por fim, a equipe também deve obrigatoriamente incluir uma justificativa acerca do método de otimização a ser utilizado (*solver*) para treinar todas as redes.

Apresentação da tarefa: 30/10/2018

Deadline para entrega: 05/10/2018 as 18h.