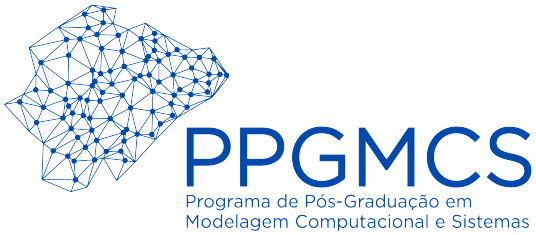

Machine Learning
Trabalho Prático 3 - Machine Learning
Prof. Dr. Honovan Rocha

O objetivo deste trabalho é implementar e avaliar Redes Neurais Mutilayer Perceptron (MLP) aplicadas a diferentes problemas de classificação.

1 Experimentos a serem realizados em todas as bases de dados

Deverão ser implementadas redes neurais do tipo Multilayer Perceptron, utilizando o framework **Keras/Tensorflow**, para obtenção de um modelo de classificação para cada uma das bases de dados descritas nas subseções a seguir. As redes deverão ser treinadas com o objetivo de gerar modelos de classificação com bom poder preditivo (bom desempenho para dados NÃO vistos no treinamento).

Para cada rede treinada, plote as curvas (iterações X Erro médio quadrático) e (iterações X acurácia) para os conjuntos de treinamento e validação (Obs: Curvas para treinamento e validação no mesmo gráfico). Para gerar estes gráficos, separe os dados de forma aleatória em 3 partes, Treinamento/Validação/Teste, utilizando a propoção 60/20/20. Lembre-se que os dados de validação são utilizados para encontrar os melhores hiperparâmetros dos modelos (caso queira, você pode utilizar um GridSearch para esta tarefa), enquanto os dados de teste medem o desempenho final do modelo.

No relatório do trabalho, apresente uma análise com detalhes acerca dos hiperparâmetros selecionados (taxa de aprendizado, quantidade de camadas, quantidade de neurônios por camada, otimizador, funções de ativação, constante momentum, etc). Analise também os gráficos gerados, verificando o comportamente das redes durante o treinamento e validação.

Após o treinamento, avalie o desempenho do modelo final obtido utilizando os dados de teste. Para fazer isso, utilize a função **classification_report** do **Scikit-Learn** e produza a matriz de confusão para os dados de teste.
Obs 1: Não se esqueça de normalizar os dados, para facilitar o trabalho do classificador.
Obs 2: A base de dados German Credit deve ser baixada do repositório, enquanto as outras 2 bases de dados são fornecidas junto ao tabalho.

1.1 Descrição da Base de Dados German Credit

A base de dados German Credit pode ser obtida no repositório da UCI [2]. Esta base foi fornecida pelo professor Hofmann [3] em Hamburgo e consiste de uma base de dados real com informações de 1000 clientes de uma instituição bancária. A base de dados original tem cada amostra definida por 24 variáveis categóricas e numéricas, entretanto, a base que deverá ser utilizada (incluída no mesmo arquivo do repositório) já foi pré-processada, contendo as 24 variáveis em formato numérico além de uma variável de rótulo, indicando se este cliente se tornou ou não inadimplente. Nesta parte do trabalho, você deverá implementar a rede MLP com o objetivo de realizar a previsão de se um novo cliente que solicitou empréstimo pagará ou não a dívida, caso o banco decida conceder o empréstimo. A Tabela 1 mostra as variáveis da base:

Tabela 1: Atributos da Base de Dados

nº	Atributos
1	Verificação saldo
2	Nº meses do emprestimo
3	Historico de credito
4	Crédito de imposto mín. alternativo - AMT
5	Saldo poupança
6	Trabalho Atual
7	Sexo
8	Tempo de residencia atual
9	Propriedade
10	Idade
11	Outros parcelamentos
12	Crédito em outro banco
13	Conta individual/conjunta
14	Telefone
15	Trabalho estrangeiro
16	Compra carro novo
17	Compra carro usado
18	devedor s/ avalista
19	Devedor c/ avalista
20	Aluguel casa
21	Possui casa
22	Desempregado
23	Trabalho informal
24	Trabalho formal

1.2 Base de Dados Desconhecida

Esta base de dados é uma base real e foi obtida junto a pesquisadores da área de Mineração. Ela consiste de diversas amostras de variáveis químicas e possui 8 possíveis classes (Y = [0...7]). A base contém 646 amostras e 25 variáveis, sendo que a primeira linha contém o nome da variável e a última coluna contem o rótulo da classe.

1.3 Base de Dados de Dígitos Manuscritos

Esta é uma base de dados clássica na literatura e consiste de diversas amostras de dígitos manuscritos. A base está dividida em 2 pastas (treinamento e teste), portanto, para esta base a divisão dos dados deve ser diferente das anteriores, de maneira que apenas os dados de treinamento devem ser divididos (70/30), gerando as bases de treinamento e validação.

O que deve ser entregue: Deve ser entregue um arquivo .zip contendo o código e um relatório contendo as análises das redes treinadas para cada base de dados. Um notebook contendo relatório e códigos também pode ser enviado.

Referências

D. Dua and C. Graff, "Uci - repositório de aprendizado de máquina," 2017. [Online]. Available: <https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data>

H. Hofmann, "German credit data," *UCI - Repositório de aprendizado de máquina*, 2000.