

**CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
DEPARTAMENTO DE COMPUTAÇÃO
INTELIGÊNCIA COMPUTACIONAL**

REGRESSÃO LOGÍSTICA

CRISTIANO NEIVA ABRANTES E CHRISTIAN MEDEIROS

**BELO HORIZONTE
NOVEMBRO DE 2021**

1 Introdução

Nesta tarefa, utilizaremos o banco de dados Human Activity Recognition with Smartphones, que foi construído a partir das gravações nas quais os participantes do estudo realizavam atividades da vida cotidiana (AVD) enquanto carregavam um smartphone com sensores inerciais embutidos.

2 Objetivo

O objetivo é classificar as atividades em uma das seis atividades realizadas (caminhar, subir escadas, descer escadas, sentar, ficar em pé e deitar).

3 Desenvolvimento

3.1 Questão 1

Analisando o data base vemos que há um grande número de colunas, em um primeiro momento é feita a leitura de dados, e também um pré processamento.

Foi verificado que as variáveis já estavam normalizadas, ou seja possuíam valores entre -1 e 1, além disso foi realizada a verificação de bias, comparando a quantidade de dados para cada atividade, e o último passo foi a substituição das labels por inteiros.

3.2 Questão 2

O objetivo deste tópico é calcular as correlações entre as variáveis dependentes, criar um histograma dos valores de correlação e identificação daqueles que estão mais correlacionados.

Desenvolve-se a matriz de correlação obtendo-se os índices da diagonal principal e tornando todos os valores abaixo dela vazios, após a simplificação descrita acima, a matriz é convertida em um DataFrame, e o histograma de correlações é plotado, e a matriz é ordenada de forma a exibir as variáveis mais relacionadas.

3.3 Questão 3

O objetivo deste tópico é dividir os dados em conjuntos de dados de treinamento e teste, além de comparação entre a proporção de classes nas divisões de treino e de teste, de forma a manter a frequência de classes preditas. Então seguimos para a regressão a taxa foi de 70 e 30 para treino e teste.

3.4 Questão 4

Ajuste no modelo de regressão logística sem qualquer regularização usando todos os recursos e armazenamento do modelo, então usamos a regressão logística para prever valores da atividade, dados as features que são inputs do celular dos usuários.

Os parâmetros usados foram penalidade l2 $C = 10$, e devido ao tamanho da base de dados o número de iterações foi maior para obtermos a convergência (20000).

4 Acurácia

O algoritmo rodou por 209.7s e obtemos a acurácia de 98.54