

ICT FOR SMART SOCIETIES

POLITECNICO DI TORINO

ICT FOR HEALTH LABORATORY

Authors:

PUTINA Andrian s226673

Professor:

VISENTIN Monica



November 18, 2016

Contents

1	Lab1: Linear Regression & Parkinson	2
1.1	Prepare the Data	2

1 Lab1: Linear Regression & Parkinson

The aim of the laboratory is understanding the basic idea of the regression and apply different techniques to obtain the desired results. In particular, it will be considered a database containing a real Parkinson's Disease dataset and different regression solutions will be applied:

- MSE (Minimum Square Error)
- Gradient Algorithm
- Steepest Descent Algorithm

Source: <https://archive.ics.uci.edu/ml/machine-learning-databases/parkinsons/telemonitoring/>

1.1 Prepare the Data

The first script takes as input variable the total number of patients `Npatients` and converts the fourth column (`test_time`) into its integer part. Furthermore creates a new matrix containing, for all the dates, only the mean of the measurements for each patient so at the end the new matrix has one row for each day (with the mean of the measured values of the features) for each patient.

```
close all
clear all
%Lab 1 - ICT HEALTH - Prepare the data

load('updrs.mat');

Npatients = 42;
parkinsonsupdrs(:,4) = abs(floor(parkinsonsupdrs(:,4)));

finalMatrix = [];
for patient = 1:Npatients

    rowPatient = parkinsonsupdrs(:, 1) == patient;
    matrixPatient = parkinsonsupdrs(rowPatient,:);

    count = 0;
    for day = 1:max(parkinsonsupdrs(:,4))+1
        indx = matrixPatient(:,4) == day-1;
        if (any(indx(:) == 1))
            count = count+1;
            meanMatrixPatient(count,:) = mean(matrixPatient(indx,:));
        end
    end

    finalMatrix = [finalMatrix; meanMatrixPatient];
    meanMatrixPatient=[];

end
save('finalMatrix','finalMatrix');
```

The next step is to divide the patients into two groups. The first sub-group will be used for the training phase of the algorithm while the second one will be used for the testing part. Anyway the data has to have zero mean and variance equal to 1 (EXPLAIN WHY!!!). The easiest way is to subtract the mean of each feature to all the patients and divide by the standard deviation. Obviously the mean and the standard deviation is computed on training data we decided before. The data is now normalized so has zero mean and variance = 1

```
close all
clear all
%Lab 1 - ICT HEALTH - Prepare data for train

load('finalMatrix.mat');

Npatients = 36;
rowPatients_train = finalMatrix(:,1) <= Npatients;
rowPatients_test = finalMatrix(:,1) > Npatients;

data_train = finalMatrix(rowPatients_train,:);
    m_data_train = mean(data_train,1);
    v_data_train = var(data_train,1);
    std_data_train = std(data_train,1);

    data_train_norm = data_train(:,5:22) - m_data_train(:,5:22);
    data_train_norm = data_train_norm ./ std_data_train(:,5:22);
    data_train_norm = [data_train(:,1:4), data_train_norm];

data_test = finalMatrix(rowPatients_test,:);
    data_test_norm = data_test(:,5:22) - m_data_train(:,5:22);
    data_test_norm = data_test_norm ./ std_data_train(:,5:22);
    data_test_norm = [data_test(:,1:4), data_test_norm];

save('data_train_norm','data_train_norm');
save('data_test_norm','data_test_norm');
```