

Samet Aytac
1819044

HW2 Report

First of all, I will briefly explain what my algorithm does. Firstly, I read input and split it by using `strsplit`. It is built-in function in Matlab 2015 but not in Matlab 2011. I took it from <http://www.mathworks.com/matlabcentral/fileexchange/21710-string-toolkits/content/strings/strsplit.m>. I choose my features as (freesulfurdioxide, totalsulfurdioxide, alcohol).

For first part, I don't divide file as test and training. I use whole data set both test and training data. I find $tetas$ and $hvalues$ for every class and I use maximum of $hvalues$ to determine the class of data. In this way, I found accuracy=50 %.

Then, in second part, I divide data to 6 part. I use 5/6 of data as training data and 1/6 of data as test data. Again, I calculate $tetas$ and $hvalues$ and I found 6 accuracy. I took average of them which is accuracy₂=49,13 %.

At third part, I do feature mapping. Now, I have 9 features (6 of them are newly generated). Then, I calculate accuracy as same way as first part and find accuracy₃=52,25 %.

When I compare first part and second part, I observe that first accuracy is slightly greater than second accuracy. It means, at first part, there might be a little overfitting by comparing second part.

When I compare first part and third part, accuracy is greater at third part. It is totally understandable because I compare data with more features.