Gaussian SVM for Forest Fire Dataset

Programming Project 3

Report

By Jigna Reshamwala and Jonathan Kocmoud

# 1. Problem Statement

In the paper “A Data Mining Approach to Predict Forest Fires using Meteorological Data”, the researchers utilize data modeling to find a mapping between meteorological data and the number and size of forest fires. In their results, the best model was created using a Support Vector Machine (SVM). The data was preprocessed and then fed into the SVM.

# 2. Implementation Details

1. Data Input:

Using the Pandas library, we read the data from the csv.

1. Preprocessing:

As explained in the research paper (page 8), we converted both the month and day to a *1-of-C* representation. Additionally, for the other inputs we normalize the inputs to a mean of 0 and a standard deviation of 1.

1. Feature Selection:

Based on the grouping developed in the research paper, a feature can be selected with the input parameters.

The program can be run with following command:

python FinalProject3.py DMC,DC,FFMC,month,day,X,Y,ISI

This will drop the features DMC, DC, FFMC, month, day, X, Y ,ISI.

1. Cross Validation: 10 cross validation

1. Evaluation

# 3. Results

The results:

1. Using best features:

Command:python FinalForetstFireSVM.py DMC,DC,FFMC,month,day,X,Y,ISI

Output:

Dropping features ['DMC', 'DC', 'FFMC', 'month', 'day', 'X', 'Y', 'ISI']

Features used: temp RH wind rain

For C = 0.0001 and gamma = 0.1 and epsilon = 4.07361277525 RMSE: 64.1965674093

For C = 1 and gamma = 0.1 and epsilon = 4.07361277525 RMSE: 63.9967634411

For C = 3 and gamma = 0.1 and epsilon = 4.07361277525 RMSE: 63.7582310243

For C = 100 and gamma = 0.1 and epsilon = 4.07361277525 RMSE: 62.3871220829

2. Using all Features:

Command: python FinalForetstFireSVM.py

Output:

Features used: X Y month day FFMC DMC DC ISI temp RH wind rain

For C = 0.0001 and gamma = 0.1 and epsilon = 130.144301433 RMSE: 133.427230031

For C = 1 and gamma = 0.1 and epsilon = 130.144301433 RMSE: 133.438868

For C = 3 and gamma = 0.1 and epsilon = 130.144301433 RMSE: 133.462238121

For C = 100 and gamma = 0.1 and epsilon = 130.144301433 RMSE: 134.241263971