

CS-470 Machine Learning

Assignment-1

Due Date: 17 November 2023

Task-1: Multivariate Gaussian Classifier

Using Fisheriris multivariate data (you already have it in .mat file), perform multivariate Gaussian classification. This a 4D data with three classes and labels are also given.

1. Randomly shuffle the data and corresponding labels.
2. Use first 100 samples (rows) in training dataset.
3. Use remaining 50 samples in test dataset.
4. Scatter plot first two dimensions of each class.
5. Using Gaussian discriminant function without dropping any term, perform classification and show the result in a form of a confusion matrix.
6. Can you drop any term in the discriminant function? Why?

Task-2: Multivariate Linear Regression

Using US CDC data of weekly flu estimates over a year, perform multivariate regression (in Matlab **load flu**). This data comprises weekly flu estimates of nine US regions (column-2 to 10). Column-1 is the date and last column (WtdILI) is the CDC's national estimate, take this as label **r**. Since there are nine regions against each date, this is $d = 9$ dimensional data. Visualize data as

```
Load flu
Y = double(flu(:,2:end-1));
[n,d] = size(Y);
x = flu.WtdILI;
figure;
regions = flu.Properties.VarNames(2:end-1);
plot(x,Y,'x')
legend(regions, 'Location', 'NorthWest')
```

1. Save .mat file for flu from Matlab to work in Python
2. Find the parameters $w_j, j = 0, 1, \dots, 9$ for the estimator
$$g(x|w) = w_0 + w_1x_1^t + w_2x_2^t, \dots, +w_dx_d^t$$
3. Observe the structure of $g(x|w)$ which should be $[N \times 1]$.
4. Plot both $g(x|w)$ and label vector **r** on the same figure to compare.

Deliverables:

1. Create a Kaggle (Jupyter) notebook. Each task will have a separate code section. You will have to upload the .mat data files to Kaggle repo of your current notebook.

2. A short pdf report on the assignment with Kaggle notebook link on top.