

Istanbul Technical University- Fall 2017

BLG527E Machine Learning

Homework 2

Purpose: Multivariate Classification, Dimensionality reduction.

Total worth: 6% of your grade.

Handed out: Wednesday, March 8, 2017.

Due: Wednesday, March 22, 2017 22:00. (through ninova!)

Instructor: Zehra Cataltepe (cataltepe@itu.edu.tr),

Assistant: Mahiye Uluyağmur- Öztürk (muluyagmur@itu.edu.tr),

Policy: Collaboration in the form of discussions is acceptable, but you should write your own answer/code by yourself. Cheating is highly discouraged for it could mean a zero or negative grade from the homework.

If a question is not clear, please let us know (via email, during office hour or in class).

Submission Instructions: Please submit through the class ninova site.

Please zip and upload all your files using filename studentID_HW2.zip. You must provide all functions you wrote with your zipped file. Functions you do not submit may cause you lose a portion of your grade. You must also include a .doc or pdf file with answers to the questions and how to call your python or matlab functions for each question so that we can run and check the results.

QUESTIONS:

Dataset:

Optdigits data by Alpaydin and Kaynak, from UCI Machine Learning Repository:

<http://ftp.ics.uci.edu/pub/ml-repos/machine-learning-databases/optdigits/>

You need the files:

optdigits.names	explanation of data
optdigits.trn	training data
optdigits.tes	test data

Q1) [6 points] [Multivariate Analysis]

Assume that each class i 's ($i=0..9$) inputs are distributed according to a normal with mean μ_i and covariance matrix Σ_i . Compute and report the training error and test error of each classifier, $g(x)$, for each of the assumptions Q1a and b:

Q1a) $\Sigma_i = \Sigma$ (common covariance matrix) and Σ is diagonal.

Implement $g(x)$ discriminant function clearly (add comments) in your code and also write its formula into the report.

Q1b) $\Sigma_i = \Sigma$ (common covariance matrix) and $\Sigma = s^2 I$ for some $s > 0$.

Implement $g(x)$ discriminant function clearly (add comments) in your code and write its formula into the report.

Q1c) Which assumption results in the best test error? Report also **test error per class** and **confusion matrix** on the test set of classifiers for each of the assumptions. Which classes are confused most with each other?

Hint1: First of all, remove the features with 0 variance.

Hint2 : You can use built-in functions for mean and variance calculations.

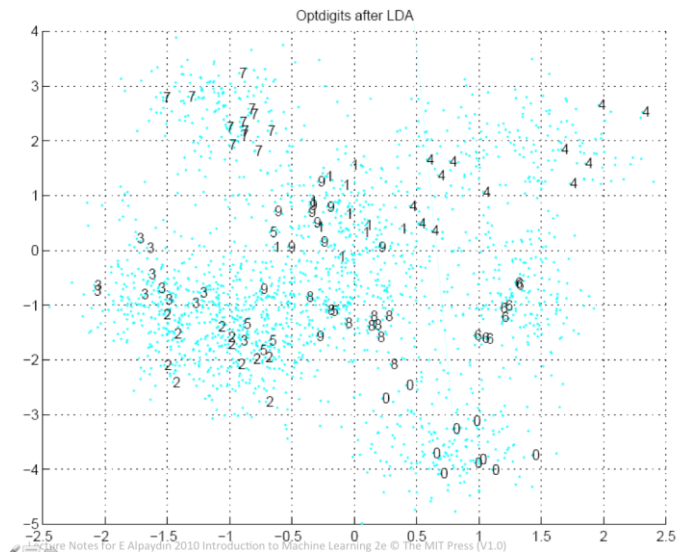
Q2)[6 points] [Dimensionality Reduction]

Project the instances in optdigits.tra dataset using the LDA algorithm into a two dimensional space as shown in the figure below.

Using the parameters of the LDA transformation you computed for the optdigits.tra dataset, show the two dimensional projection of optdigits.tes dataset in another plot.

Hint1: Do not re-compute the LDA parameters for the optdigits.tes set, just reuse the ones you computed for the optdigits.tra set.

Hint2 : Do not use a built-in LDA function, implement it yourself.



Output format for Q2: The result for Q2 will be a figure as it is shown in the figure above.

Output format for Q1:

Results of your program must be written into **output.txt** text file in the following format:

Question1 results:

Training error of Q1a:

Test error of Q1a:

Training error of Q1b:

Test error of Q1b:

Test error for each class for Q1a:

Test error for each class for Q1b:

Confusion matrix of the Q1a:

Confusion matrix of the Q1b: