

CS60050: Machine Learning
Autumn 2015, CSE, IIT Kharagpur
Assignment 2

Full marks: 100

Q1. Derive the dual problem of Hard-margin SVM. Solve the problem for the XOR dataset:

X1	X2	Y
0	0	0
0	1	1
1	0	1
1	1	0

[10 marks]

Q2. Define Bayes classifier. Show that the bayes classifier will achieve the best error rate, defined as:

$$E(f) = \int I(y \neq f(x)) p(x, y) dx dy$$

where $f(x)$ is the classifier, and $p(x, y)$ is the intrinsic data distribution. **[10 marks]**

Q3. Show that for a binary classification problem, least squares classification with labels $t = N/N_1$ for class 1 and $t = -N/N_2$ for class 2 gives the same solution as Fisher's linear discriminant analysis. **[10 marks]**

Q4. Download the breast cancer dataset from:

<https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary/breast-cancer>

Implement the naïve bayes classifier (using matlab / perl / python / C / Java). Report k-fold cross-validation error for $k = 3, 5, 10$. Also, report the algorithm implemented and attach the code. **[20 marks]**

Q5. Consider a 2 class classification problem with a dataset of inputs $\{x(1) = (-1, -1), x(2) = (-1, +1), x(3) = (+1, -1), x(4) = (+1, +1)\}$. Can this dataset be classified by SVM classifier with the following kernels:

- Linear kernel.
- Polynomial kernel of degree 2.
- Gaussian kernel.

Show the shatterings. **[10 marks]**

Q6. Implement a logistic regression solver and test it on the above mentioned breast cancer dataset using following two techniques:

- Iterative re-weighted least squares.
- Stochastic Gradient Descent.

Implement the algorithm in Matlab. Report the algorithms implemented and the matlab code. Show plots of objective function value as a function of iterations. **[40 marks]**