CSE/ECE 343/543: Machine Learning
Assignment-2

Max Marks: 140 **Due Date**: 24-Oct-2018, 11:59PM

Instructions

- Keep collaborations at high level discussions. Copying/Plagiarism will be dealt with strictly.
- Start early, solve the problems yourself. Some of these questions may be asked in Quiz/Exams.
- Submission Instructions: Submissions will be through backpack. Create a single *firstname-A2.zip* file containing a report **A2.pdf**, your source folder **A2-src** and theory questions solutions **A2-theory.pdf**. Report all your theory solutions and outputs of all programming questions e.g intrinsic and extrinsic parameters,figures, images etc in **A2.pdf**. List name of all the functions/scripts that you have implemented along with the two line summary in **A2.pdf**. Put all your programming functions/scripts in **A2-src**. You are allowed to use *numpy*, *scipy* and *matplotlib* only, unless specified otherwise. In case of any doubt, initiate a discussion on backpack or drop an email to Aradhya {`aradhyam@iiitd.ac.in`} **and** Lamha {`lamha15050@iiitd.ac.in`} with the subject line [**ML18-A2-Doubt**]. Emails with other subject lines may suffer delays in response.
- Report(A2.pdf) is **required**. 50% of the total points of the programming question will be deducted if the results are not reported in A2.pdf
- Late submission penalty: As per course policy.

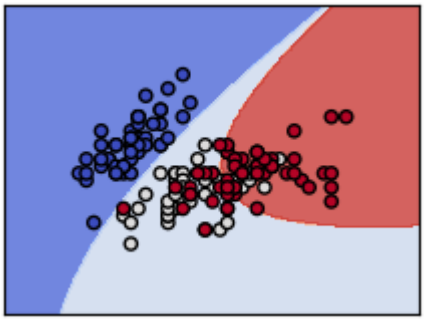
PROGRAMMING QUESTIONS

1. (85+10 points) **Support Vector Machines.**
   Note: Using any built-in function other than *sklearn*s .fit() is not allowed. You are to write your own functions, even for .predict() .

   i) (*3 points*) You are given four 2-dimensional datasets (data_1, data_2, data_3, data_4). Explore the datasets, plot them and write your observations and findings of the datasets. Particularly, comment on the number of samples, balance among classes, separability and noise in the dataset.

   ii) (*20 points*) For each dataset, write a kernel (if required) to make them linearly separable. Explain the choice of each kernel. Plot the datasets with decision boundaries corresponding to those kernels; something like this:

   iii) (*27 points*) Implement Soft margin SVM with linear kernel. Use the built in *sklearn*s binary classifier. Use the binary classifiers to implement a multi-class classifier for M classes. Split each dataset randomly into 80% training set and 20% test set.



Report the accuracy and F1 score on each dataset.
Note: For converting the binary classifier to multiclass, you need to use One-vs-Rest Classifier and One-vs-One Classifier You need to implement them yourself and write an analysis (based on running time and performance metric: accuracy in this case) on the two techniques. You can only use the .fit() function of the sklearns SVM module. You would have to write all the other functions including predict function (overall predict for the M-class classifier) yourself.
See [SS-Ch-17.1] for understanding One-vs-Rest and One-vs-One approaches (referred to as "One-versus-All" and "All-Pairs" respectively in the book).

iv) (*10 points*) Repeat the above part using an RBF kernel instead of the linear kernel.

v) (*25 points*) Use the RBF kernel with SVMs to classify the hindi handwritten character dataset. You are only given the training (attached as '`Train_val_Handwritten_Hindi_dataset.zip`' file) and the test dataset (attached as `Test_Handwritten_Hindi_dataset.zip` file). Split the data as you see fit and train your kernelized SVM. Report the training, validation and test error. Make sure that you do not use the test set for any training/cross-validation.

vi) (*25+10 points*) For both of the above parts, report the following analysis:

- What hyperparameters did you choose? Why? How did you choose the optimal values for the hyperparameters?
- Plot the support vectors and the margin separating hyperplane.
- Choose a value of hyperparameters such that the model overfits the training data. Discuss how overfitting affects the number of support vectors.
- Discuss your observations on performance of the linear kernel vs the RBF kernel.
- Plot confusion matrices for each dataset. You have to calculate it yourself, and cannot use any library for this.
- Plot ROC curves for data_1, data_2 and data_4 using the sklearn library.
- (**Bonus**) Plot multi-class ROC curve for data_3 and the hindi handwritten character data.

THEORY QUESTIONS

2. (10 points) (**Bonus**) The RBF kernel can map the given data into a higher dimensional space, with possibly infinite dimensions. One way to look at it is that every data points gets its own dimension, which would lead to overfitting. However, in reality, this generally does not happen. Why?

3. (10 points) Show that the sum of two convex functions is a convex function. Use this argument to show that the loss function for ℓ_1 regularized linear regression (with mean squared error) is a convex function defined over the parameters.