CS436/536: Introduction to Machine Learning
Homework 5
Due 11/27 **before 11.59pm**

Instructions: To solve these problems, you are allowed to consult your classmates, as well as the class textbook (*Learning from Data* by Abu-Mostafa, Magdon-Ismail, and Lin, which we will call LFD) and the slides posted on Brightspace, but no other sources. If you have bought the book (in print or electronic form), you can access the online 'e-chapters' at http://www.amlbook.com/eChapters.html. You are encouraged to collaborate with other students, while respecting the collaboration policy (please see the module on Academic Honesty on Brightspace). Please write the names of all the other students you collaborated with on the homework. Everyone must write up their assignments separately.

Please write clearly and concisely, and use rigorous, formal arguments. Homework is due at the beginning of lecture, and homework turned in later will be considered late and will use up one of your late days. You must use Brightspace to submit the homework as a single neatly typed pdf file. Hand-drawn formulas or figures are okay and may be included as images within the pdf. If a programming assignment calls for plotting the results, axes must be clearly labeled, and its meaning must be obvious to anyone with only a rudimentary knowledge of machine learning and computer science. Emailed copies will not be accepted.

There is no supporting code for this homework. Please feel free to explore the documentation and examples for the `scikit-learn` python library. One of the objectives of this homework is to become familiar with reading through the documentation of existing python libraries and making the best use of the documentation, the small examples and recipes they provide to accomplish common tasks in machine learning. All the best!

(1) [400 points] The k-NN rule.
For this question, you will use the data you generated in HW3 from the MNIST Digits Dataset for classifying 1s vs. Not 1s, where you created $\mathcal{D}$ with 300 randomly selected data points and $\mathcal{D}_{\text{test}}$ consisting of the remaining data points.

You will have to implement the k Nearest Neighbors (k-NN) rule. You may use the helper code as a starting point.

(a) Use cross validation with $\mathcal{D}$ to select the optimal value of k for the k-NN rule. Show a plot of E_{CV} versus k and indicate the value of k you choose.

(b) For the chosen value of k, plot the decision boundary. Also compute and report its E_{in} and E_{CV}.

(c) Report E_{test} for the k-NN rule corresponding to the chosen value of k.

(2) [300 points] Support Vector Machines.
For this question, you will use the data the data you generated in HW2 from the MNIST Digits Dataset for the training and test datasets in `ZipDigits.train` and `ZipDigits.test` respectively with the two features you computed for the problem of classifying 1s vs. 5s.

(a) Use this method for training support vector machines and this method for model selection with cross validation from the `scikit-learn` python library to find the value C for the regularization parameter with the smallest cross validation error using 5-fold cross validation and the training dataset with two features you formed from the data in `ZipDigits.train`. Report E_{CV} for the best value for C.

(b) For the chosen value of C, learn a support vector machine using all of the training data. Compute and report its E_{in}.

(c) Use the test dataset from `ZipDigits.test` to compute E_{test} for the classifier you just learned and report it. Compare it with the results from HW2 using the linear model.

(3) [300 points] Support Vector Machines with the Polynomial Kernel.
For this question, you will use the data you generated in HW3 from the MNIST Digits Dataset for classifying 1s vs. Not 1s, where you created $\mathcal{D}$ with 300 randomly selected data points and $\mathcal{D}_{\text{test}}$ consisting of the remaining data points.

(a) Use this method (not the same as the one for the previous question) for training support vector machines using the kernel for the 10-th order polynomial feature transform and this method for model selection with cross validation from the `scikit-learn` python library to find the value C for the regularization parameter with the smallest cross validation error using 5-fold cross validation and the training dataset with two features you formed from the data in `ZipDigits.train`. Report E_{CV} for the best value for C.

(b) For the chosen value of C, learn a support vector machine using all of the training data. Compute and report its E_{in}.

(c) Use the test dataset from `ZipDigits.test` to compute E_{test} for the classifier you just learned and report it. Compare it with the results from HW3 using the linear model with the 10th order polynomial feature transform.

(4) [300 points] Compare Methods. Now that we have implemented several algorithms to learn from data, it is time to reflect upon them. Compare the final test error from your attempts to solve the 1s vs. Not 1s classification problem on the MNIST Digits Dataset, namely:

(i) Regularized Linear Model with 10-th order polynomial transform, where the regularization parameter λ is selected using cross validation.

(ii) The k Nearest Neighbors rule where k is selected using cross validation.

(iii) Neural Networks with early stopping.

Report and summarize all of your results and make some intelligent observations about the results you obtained.