

# HW 4 – SVMs and other classifiers

Understanding of neural networks and implementing one from scratch

First Submission Due: Friday, November 1<sup>st</sup>, 11:59pm Pacific Time

Revision Due Dates will be updated after the grades are released

## Task 1 – the basic SVM classifier

[LP] In this homework, we will look at all the data points available in the digits dataset to perform classification for digit 1 and not digit 1. You are encouraged to reuse your preprocessing code from HW1. Different from HW1, you should not filter the dataset to remove any digits. Only convert the labels into labels for binary classification, i.e., '1' and '-1', where '-1' corresponds to digits other than 1. Extract the "intensity" and "symmetry" for the entire dataset similar to HW1. DO NOT look at the test dataset until you reach question 4.

For this homework, you may use a library implementation of Support Vector Machine classifiers, e.g., svm on sklearn provides the SVC class for binary classification.

Use the entire training set and train an SVM classifier using the linear kernel, the polynomial kernel (3<sup>rd</sup> order), and the RBF kernel. Use the default value for the regularization parameter ( $C=1$ ). Report the training error on all three kernels.

[HP] What is the best kernel for this dataset only based on the results above? Is this an appropriate way to select a model? Why or why not? Be brief in your answer.

## Task 2 – Cross validation

[LP 1] Use k-fold cross validation with  $k = 5$  to select the best kernel and regularization parameter among the following:

Kernel: linear, 3<sup>rd</sup> order polynomial, RBF

C: 0.01, 0.1, 1, 10, 100

Note that you will still be working only with the training dataset and splitting it into training and validation sets. You may use the sklearn's model selection class that provides you with the k-fold cross validator. You may also write your own implementation for cross validation.

Report the cross validation error on each combination of kernel and C value.

[LP 2] What kernel and C value you would choose as a result of this experiment?

[HP] Plot the cross validation error and in-sample error using an appropriate visualization and explain the result in a caption for the plot.

## Task 3 – The final hypothesis

[LP 1] Train your final SVM hypothesis based on cross validation on your entire training dataset. Report the in-sample error.

[LP 2] Extract the test dataset from the ZipDigits.test file. Note that you do not need to filter this dataset, similar to the training set in this homework. Report the final test error for the SVM using the parameters chosen from LP 1.

## Task 4 – Overall Learning approach

[HP 1] Collect all the results of your experiments on the digits dataset from previous homework by reporting the test errors in a table or graph (or whatever visualization you prefer). The models you should report for are PLA, pocket, pocket with 3<sup>rd</sup> order polynomial feature transform, and neural network trained using the three gradient descent approaches.

Briefly comment on the model with the best and worst test error. Why do you think the models show this performance? Be brief in your explanation. Please write in your own words, even if you discuss with others.

[HP 2] Is this approach to choose the best digit classifier appropriate? Comment on a good approach to choose the best digit classifier using the provided dataset. Think about the size of the dataset, the complexity of these models, and the parameters you can alter. Be brief in your answer. As always, please write in your own words, even if you discuss with others or refer to external sources.

## Submission instructions

As mentioned in class activity 1, we will use GitHub to share code and track progress in this class. If you have not already done so, create a GitHub private repository for the course and name it "CMPE257-Fall23-FirstName-LastName". Add the following users to the repository: mahima-as and rbpravin. You are also welcome to add the instructor and ISA to your Colab notebooks.

In your GitHub repo, create a branch called *homework-4*. Frequently commit your code and make sure it is shared with the instructor and ISA. Include a link to the GitHub repository in your submission pdf.

## Specifications

Tasks 1, 2, and 3 have components labeled [LP] and tasks 2 and 4 have components labeled [HP]. If you complete ALL the LP components satisfactorily, you will receive a grade of “low pass” on the homework. If you complete ALL the LP components and at least 2/4 HP components satisfactorily, you will receive a grade of “high pass”. If you do not meet the criteria for a “low pass”, the submission will be marked as “revision needed”.

Note the following statements from the syllabus:

*If a student receives a “low pass” or “revision needed” grade, the student may revise and resubmit their homework assignment by using one “token”.*

*For homework assignments, if the student fails to submit their assignment by the posted deadline, their submission will receive a grade of “revision needed”. If they fail to submit the assignment by the revision deadline, the submission will receive a grade of “fail”.*

At most two tokens may be used for the one-day deadline extensions (one token for each one-day extension), including the revision deadlines. Tokens will be automatically removed from your wallet if you submit late and/or resubmit.

**VERY IMPORTANT:** Include ALL the references you used for this assignment, including names of classmates you discuss with. Failure to cite your sources counts as an act of academic dishonesty and will be taken seriously without zero tolerance. You will automatically receive a “fail” grade in the homework and further serious penalties may be imposed.

**NOTE: You can look for help on the Internet but refrain from referencing too much. Please cite all your sources in your submission.**

When you submit your assignment, you automatically agree to the following statement. If you do not agree, it is your responsibility to provide the reason.

*“I affirm that I have neither given nor received unauthorized help in completing this homework. I am not aware of others receiving such help. I have cited all the sources in the solution file.”*