Computer Vision
HW#3
Due March 24, 2020
Submission:

- *Please use your programming language of choice, MATLAB, C, python, …*
- *Please prepare a .pdf report file containing your answers.*
- *Compress your report file and your codes into a single zip file.*
- *Name the file as your last name followed by hw3 for example: shirani_hw3.zip*
- *Submit by uploading to Avenue (under Assessment/Assignment). Please do NOT email me your file.*

1. In this question you will implement linear regression with Gaussian bases. Start by downloading the dataset from Avenue. The dataset is the AutoMPG dataset from the UCI repository. The task is to predict fuel efficiency (miles per gallon) from 7 features describing a car. Load the data and normalizing the features and targets to have 0 mean and unit variance if necessary.
   Implement linear basis function regression with Gaussian basis functions. For the centers μ_j use randomly chosen training data points. Set covariance matrix $\sum = 2I$. Perform the following experiments:
   a) Using the first 100 points as training data and the remainder as testing data. Fit a Gaussian basis function regression using 5, 15, 25, . . ., 95 basis functions. Do not use any regularization. Plot training error and test error (in RMS error) versus number of basis-functions. Put this plot, along with a brief comment on what you see, in your report.
   b) Implement L2-regularized regression. Again, use the first 100 points. Fit a regression model with 90 basis functions using $\lambda = \{0, 0.01, 0.1, 1, 10, 100, 1000\}$. Use 10-fold cross-validation to decide on the best value for λ. Produce a plot of validation set error versus regularizer value. Use a semilogx plot, putting regularizer value on a log scale. Put this plot in your report and note which regularizer value you would choose from the cross-validation.

2. Download the training data set “musicdata.txt” and the test data set “musictestdata.txt” from Avenue. The data consist of 91 variables concerning commercial songs released between 1922 and 2011. The first variable (the song's release year) is the response variable that we are trying to predict based on the remaining 90 input variables which consist of various measures of timbre. There are 1000 songs in the training data and 300 in the test data. As there are a considerable number of input variables over fitting is a serious issue. In order to avoid this, implement the regularized ML estimation with lasso regularizer. Your function should accept a scalar value of λ, a vector-valued response variable (y) and a

matrix of input variables (X) and it should output a vector of coefficient values w.

Once you have implemented a lasso solver function run the solver using 100 different values of λ (try different range of values for λ). In your analysis, include:

1. A plot of log(λ) against the squared er**r**or in the training data.
2. A plot of log(λ) against the squared erro**r** in the test data.
3. A plot of λ against the number of nonzero coefficients.

3. This question reviews the principles of linear classifiers. Use the Australian Crab dataset (Text file) available on Avenue. The columns of the data are 'sp' 'sex' 'index' 'FL' 'RW' 'CL' 'CW' 'BD'. Species is either 1 or 2, Sex is 1 or 2, index is a number of a particular data point for a log book. FL is the frontal lobe size, RW is the rear width of the shell, CL is the carapace (the shell covering the body) length, CW is the carapace length, BD is the body depth. Each row is a data point.
    a) The two classes are the two species of crab. Use the measurement data (FL,RW,CL, CW,BD) to classify the crabs into one of the species. Normalize your data to the range [-1, 1] by either mapping it linearly into a range or by subtract the mean for each feature and divide by the standard deviation. Use 75% of the data as training data and 25% as test data. Use the logistic regression to learn the hyperplane between the data sets. Give your results in a confusion matrix.
    b) Design a linear SVM to classify the data. Give the results in a confusion matrix.