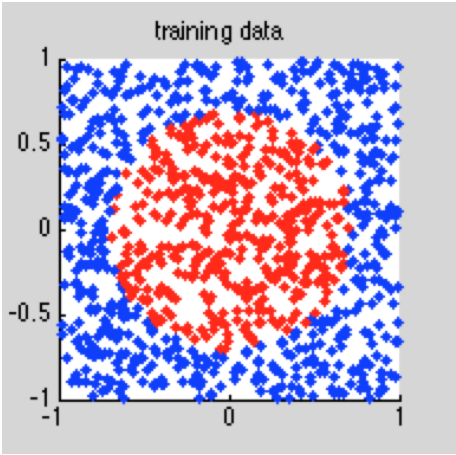
CS/ECE/ME 532
Homework 8: Hinge Loss and SVMs

Consider the classification problem discussed in class, where the goal is to design a classifier based on the training data shown in the plot below. The Matlab file `training_data.m` generates this dataset and also computes the standard least squares solution.



1. Express the standard least squares solution using the dual form using regularization parameter $\lambda = 10^{-5}$. Verify that it generates the same results as the primal solution.

2. Design a classifier using the Gaussian kernel

$$k(\boldsymbol{a}_i, \boldsymbol{a}_j) = \exp(-\frac{1}{2}\|\boldsymbol{a}_i - \boldsymbol{a}_j\|_2^2)$$

and regularization parameter $\lambda = 10^{-5}$. Compare its classification of the training data to the least squares classification.

3. Design a classifier using the polynomial kernel

$$k(\boldsymbol{a}_i, \boldsymbol{a}_j) = (\boldsymbol{a}_i^T \boldsymbol{a}_j + 1)^2$$

and regularization parameter $\lambda = 10^{-5}$. Compare its classification of the training data to the least squares classification and Gaussian kernel classifier.

4. Now experiment with these methods in the following way. Generate different sized sets of training data: $m = 10, 100, 1000$. Design the three types of classifiers using these data. Test the performance of the classifiers using independent sets of data of size 100. For each training data size, repeat the training and testing 100 times and average the results. This will provide a good indication of how the different approaches perform with different amounts of training data. Summarize your results by reporting the average test error for each of the three methods and each of the three training set sizes.

5. Now tackle the problem using hinge loss instead of squared error loss. You do this using the Matlab function `svmtrain`, another package, or by writing your own code (e.g., GD or SGD). Generate different sized sets of training data: $m = 10, 100, 1000$. Design SVM classifiers using both Gaussian and polynomial kernels. Test the performance of the classifiers using independent sets of data of size 100. For each training data size, repeat the training and testing 100 times and average the results. Compare your results to those obtained using least squares.

6. Consider the problem of trying to predict whether a person is a basketball player based on height. The training data consists of four people with heights of $5'10''$, $5'11''$, $6'1''$ and $6'10''$, and the latter two are basketball players. What classification rule do you obtain by minimizing hinge loss instead of squared error loss?

7. If you consider the dual solution in this case, which values of $\boldsymbol{\alpha} \in \mathbb{R}^4$ are nonzero (i.e., what are the "support vectors")?

8. The standard hinge loss is linearly decreasing up to 1 and then exactly 0 after. Let t_0 denote the point where the hinge loss first becomes 0. Does the solution change if you shift the t_0 within the range $t_0 \in [0, 1]$?