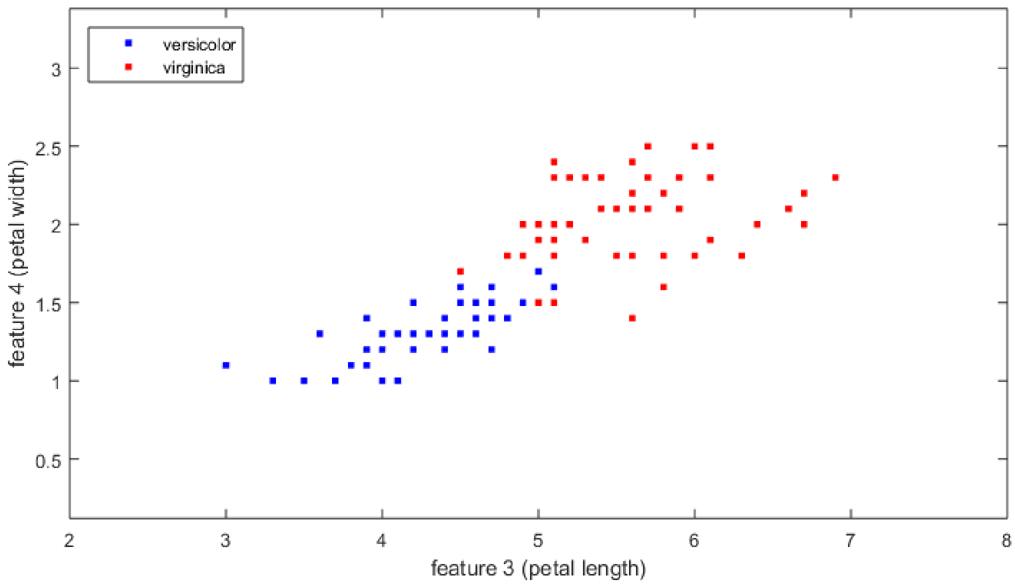
CS/ECE/ME 532
Homework 8: SVM and kernels

due: Monday December 5, 2016

1. Classification and the SVM. Revisit the `iris` data set from Homework 4. For this problem, we will use the 3rd and 4th features to classify whether an iris is *versicolor* or *virginica*. Here is a plot of the data set for this restricted set of features.



We will look for a linear classifier of the form: $x_{i3}w_1 + x_{i4}w_2 + w_3 \approx y_i$. Here, x_{ij} is the measurement of the j^{th} feature of the i^{th} iris, and w_1, w_2, w_3 are the weights we would like to find. The y_i are the labels; e.g. $+1$ for *versicolor* and -1 for *virginica*.

a) Reproduce the plot above, and also plot the decision boundary for the least squares classifier.

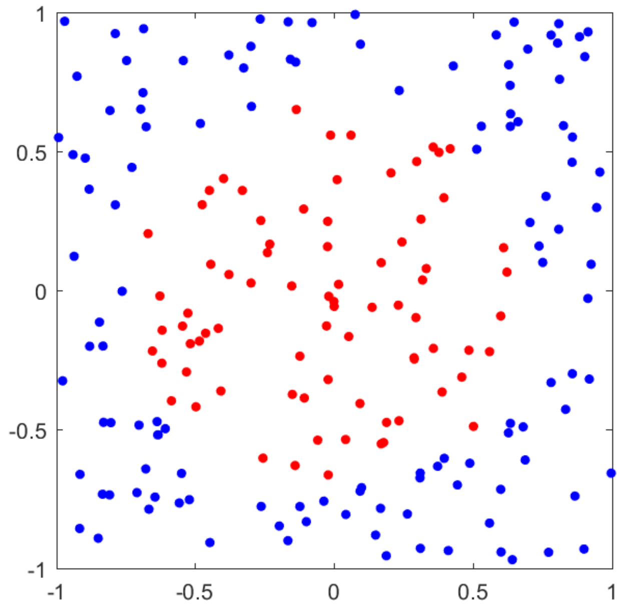
b) This time, we will use a regularized SVM classifier with the following loss function:

$$\underset{\boldsymbol{w}}{\text{minimize}} \quad \sum_{i=1}^{m} (1 - y_i \boldsymbol{x}_i^\mathsf{T} \boldsymbol{w})_+ + \lambda(w_1^2 + w_2^2)$$

Here, we are using the standard hinge loss, but with an ℓ_2 regularization that penalizes only w_1 and w_2 (we do not penalize the offset term w_3). Solve the problem by implementing gradient descent of the form $\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \gamma \nabla f(\boldsymbol{w}_t)$. For your numerical simulation, use parameters $\lambda = 0.1$, $\gamma = 0.003$, $\boldsymbol{w}_0 = \boldsymbol{0}$ and $T = 20{,}000$ iterations. Plot the decision boundary for this SVM classifier. How does it compare to the least squares classifier?

c) Let's take a closer look at the convergence properties of $\boldsymbol{w}_t$. Plot the three components of $\boldsymbol{w}_t$ on the same axes, as a function of the iteration number t. Do the three curves each appear to be converging? Now produce the same plots with a larger stepsize ($\gamma = 0.01$) and a smaller stepsize ($\gamma = 0.0001$). What do you observe?

2. Classification with kernels. Consider the two-dimensional classification problem based on the training data shown in the plot below. The data is provided in the file `circledata.csv`, where the first two columns are the x and y coordinates and the third column is the label ± 1.



a) Solve the standard least squares (linear) classification problem with ℓ_2 regularization parameter $\lambda = 10^{-5}$. Also solve the dual formulation and verify that both produce the same solution.

b) Design a least squares classifier using the Gaussian kernel

$$k(\boldsymbol{a}_i, \boldsymbol{a}_j) = \exp\left(-\tfrac{1}{2}\|\boldsymbol{a}_i - \boldsymbol{a}_j\|^2\right)$$

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

c) Design a least squares 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.

d) Now tackle the problem using hinge loss instead of squared error loss. You may do this using the Matlab function `svmtrain`, another package, or by writing your own code (e.g., GD or SGD). Design SVM classifiers using both Gaussian and polynomial kernels and compare your results to those obtained using least squares.