Homework 8

Mathematical and conceptual exercises

1. A decision boundary in R^2 is given by the equation

$$x_1 + 3x_1x_2 = 6x_2^2 + 8$$

This can be written in form $w \cdot \Phi(x) + b = 0$, where $x = (x_1, x_2)$, the basis expansion $\Phi(x) = \big(x_1, x_2, x_1^2, x_2^2, x_1x_2\big)$, and $b = -8$. What is w?

2. Data vectors $x = (x_1, \ldots, x_4)$ are augmented to give expanded features

$$\Phi(x) = \big(x_1, \ldots, x_4, x_1^2, \ldots, x_4^2, x_1x_2, \ldots, x_3x_4\big)$$

   (a) What is the dimension of $\Phi(x)$?

   (b) The Perceptron algorithm is run using the basis expansion $\Phi(x)$ and returns w, b. What is the dimension of w?

3. We have a data set, of d-dimensional points and their labels, that is linearly separable. However, we use a basis expansion

$$\Phi(x) = \big(x_1, \ldots, x_d, x_1^2, \ldots, x_d^2, x_1x_2, \ldots, x_{d-1}x_d\big)$$

and run the Perceptron algorithm with these expanded features.

   (a) Will the Perceptron algorithm necessarily converge?

   (b) Will the algorithm necessarily return a vector w in which the entries corresponding to quadratic terms in $\Phi(x)$ (such as x_1^2) are zero?

4. A data set consists of just four points in R^2:

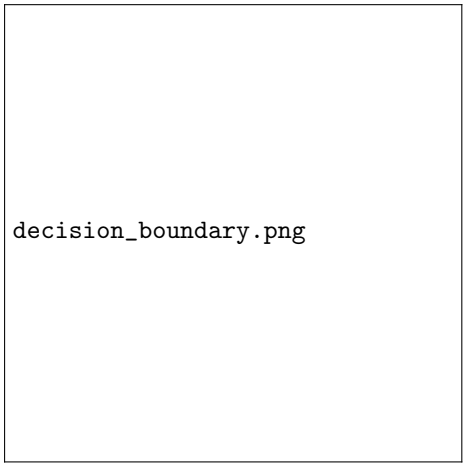
   - Label 1: points $x^{(1)} = (2, 3)$ and $x^{(2)} = (3, 6)$
   - Label -1: points $x^{(3)} = (1, 2)$ and $x^{(4)} = (0, 7)$

The kernel Perceptron algorithm (in dual form) is run on this data set, with a basis expansion $\Phi(\cdot)$ that produces a quadratic boundary. The algorithm converges after just five update steps: two updates on point $x^{(2)}$, two updates on point $x^{(3)}$ and one update on point $x^{(4)}$.

   (a) What vector α is returned?

(b) What value of b is returned?

5. Consider the following example from lecture, which shows the decision boundary resulting from applying kernel SVM (with quadratic kernel) to a small two-dimensional data set.



   (a) What is the dimension of α in this case?
   (b) How many entries in α are > 0?
   (c) How many entries in α are < 0?
   (d) For you to think about: why does the margin on the green side appear to be larger than the margin on the purple side?

Programming exercises

The data files for this week's lab are contained in `week8.zip`, which you can download from the course website.

1. Kernel Perceptron. Implement the kernel Perceptron algorithm, with the quadratic and RBF kernels. The data sets `data1.txt` and `data2.txt` contain 2-d data with two classes (coded as -1 and 1). Each row has three numbers: the two coordinates of the data points and the label.

   (a) Run the kernel Perceptron with quadratic kernel on these two data sets. In each case, show a plot that contains all the data points (with different colors and shapes for different labels) as well as the decision region.
   (b) Repeat for the RBF kernel. Show the results for two different settings of the scale parameter σ.

2. Multiclass kernel SVM. In this problem, we'll use support vector machines to classify the MNIST data set of handwritten digits.

(a) Load in the MNIST data: a training set of 60,000 points and a separate test set of 10,000 points.

(b) Learn a linear SVM classifier using `sklearn.svm.LinearSVC`. You will need to set `loss='hinge'`. Try different values of the tradeoff parameter: $C = 0.01, 0.1, 1.0, 10.0, 100.0$. In each case, report the training error and test error. Do you think this data is linearly separable?

(c) Now try kernel SVM with a quadratic kernel. You can do this with `sklearn.svm.SVC`, setting `kernel='poly'` and `degree=2`. Just try the single setting $C = 1.0$. Report the training error, the test error, and the number of support vectors.