Self test: Perceptrons

The following questions are to help you further understand the material about perceptrons.

Linear separability (1) We looked only at cases where the positive and negative examples were separable by a hyperplane passing through the origin, namely of form $\mathbf{w} \cdot \mathbf{x} = 0$, where $\mathbf{x}$ denotes a general point in the space.

The reason we did this is that hyperplanes through the origin are linear spaces while those that do not pass through the origin are not. More importantly, the case where the positive/negative examples are separated by a hyperplane not containing the origin is not really novel—in fact, it can be seen as a special case of when we can separate points with a hyperplane through the origin as you will show below.

A general hyperplane in d dimensions is of form $\mathbf{w} \cdot \mathbf{x} = b$ where b is not necesasrily 0 and $\mathbf{w}$ is a vector with d coordinates. As always, in this document $\mathbf{x}$ represents a general point in the space concerned.

Consider m examples $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \ldots, \mathbf{z}^{(m)}$ (each example being a vector with d coordinates) with labels $y_1, y_2, \ldots, y_m$. Then write for all $1 \leq i \leq m$, $\tilde{\mathbf{z}}^{(i)} = (\mathbf{z}^{(i)}, -1)$, namely the vector $\tilde{\mathbf{z}}^{(i)}$ has $d+1$ coordinates, and is obtained by concatenating the vector $\mathbf{z}^{(i)}$ with the number -1 in the last position.

Show that the points $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \ldots, \mathbf{z}^{(m)}$ with labels $y_1, \ldots, y_m$ are linearly separable by a general hyperplane in d dimensions (i.e. $\mathbf{w} \cdot \mathbf{x} = b$) iff the points $\tilde{\mathbf{z}}^{(1)}, \tilde{\mathbf{z}}^{(2)}, \ldots \tilde{\mathbf{z}}^{(m)}$ with labels $y_1, \ldots, y_m$ are linearly separable by a hyperplane (in $d+1$ dimensions) through the origin (i.e. of form $\tilde{\mathbf{w}} \cdot \tilde{\mathbf{x}} = 0$).

Linear Separability (2) We will do this in class when we look at Support Vector Machines, but it is a good idea to try your hand at this before.

Show that the points $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \ldots, \mathbf{z}^{(m)}$ are linearly separable by a general hyperplane in d dimensions (i.e. $\mathbf{w} \cdot \mathbf{x} = b$) iff there is a vector $\mathbf{w}'$ and a number b' such that for all $1 \leq i \leq m$,

$$y_i \Big(\mathbf{w}' \cdot \mathbf{z}^{(i)} - b' \Big) \geq 1.$$

Perceptron algorithm Suppose you run the Perceptron Algorithm through T training examples that are linearly separable, but have a margin that is quite small. Consequently, you find that at the end of going through all examples, the output hyperplane is yet unable to classify all examples properly.

In notation that we will use in the rest of the course, one training run through all the training examples is an *epoch*. Therefore after one epoch, the Perceptron Algorithm has not yet found the hyperplane that classifies all examples perfectly.

Is it a good idea to run through the same examples for another epoch? Given T training examples that are linearly separable, does the Perceptron Algorithm always find the separating hyperplane if it is run for enough epochs?