Chwan-Hao TUNG
SID: 861052182
PS 3

The total loss for a linear SVM:

$$C\sum_i [1 - y_i f(x_i)]_+ + \frac{1}{2}w^T w$$

To get the gradient decent for a linear SVM, we first take a partial derivative with respect to the weight vector w and b on the loss function,
Since the linear classifier is of the form $f(x_i) = w^T x_i + b$.

$$C\sum_i [1 - y_i(w^T x_i + b)]_+ + \frac{1}{2}w^T w = C\sum_i [1 - y_i w^T x_i + y_i b]_+ + \frac{1}{2}w^2$$

$$\frac{\partial}{\partial w} = C\sum_i [-y_i x_i]_+ + w$$

$$\frac{\partial}{\partial b} = C\sum_i [-y_i]_+$$

The update rule will thus be,

$$w \leftarrow w - \eta \cdot (C\sum_i [-y_i x_i]_+ + w)$$

$$b \leftarrow b - \eta \cdot (C\sum_i [-y_i]_+)$$

where η is the step size.

Generally speaking, the perceptron learning rule tend to end up with classification that can be very close to a set of data points, in other words, a very narrow margin because perceptron converges as soon as all data are separated. In the case of linear SVM, we can choose to have a larger margin by adjusting the impact of the constraints so we can have better generalization over the data, usually this entails some misclassified data points or training error.