- Nikhil Kamthe

- 861245635

- 10/20/2016

- CS 229

- PS 3

Solution: The loss function with regularization for a linear SVM can be written as

$$L = C \sum_i^m [1 - y_i f(x_i)]_+ + \frac{1}{2} w^T w \qquad \text{where } f(x_i) = w^T x_i + b$$

In order to find the general gradient descent learning rules for linear SVM, we need to find the iterative update in w and b. We can use the following equations to find the iterative updates in w and b.

$$w_{new} = w_{old} - \eta * \frac{dL}{dw} \qquad \text{where } \eta = \text{step size}$$

$$b_{new} = b_{old} - \eta * \frac{dL}{db}$$

IMPORTANT: Let me introduce following notation to make our life easier

$$[\phi(x_i)]_{svm} = \phi(x_i) \qquad \text{when } y_i f(x_i) < 1$$
$$[\phi(x_i)]_{svm} = 0 \qquad \text{when } y_i f(x_i) \geq 1$$

Lets compute $\frac{dL}{dw}$ and $\frac{dL}{db}$ in order to derive the update rules.

$$\frac{dL}{dw} = C \sum_i^m [-y_i x_i]_{svm} + w$$

$$\frac{dL}{db} = C \sum_i^m [-y_i]_{svm}$$

Therefore we can write the general gradient descent update rules for linear SVM as follows.

$$w_{new} = w_{old} - \eta * (C \sum_i^m [-y_i x_i]_{svm} + w_{old})$$

$$b_{new} = b_{old} - \eta * (C \sum_i^m [-y_i]_{svm})$$

Stochastic gradient descent update rules for linear SVM can be derived by evaluating general gradient descent update rules at a single data point k.

$$w_{new} = w_{old} - \eta * (C[-y_k x_k]_{svm} + w_{old})$$
$$b_{new} = b_{old} - \eta * (C[-y_k]_{svm})$$

From the update rule we can see that the Perceptron penalizes incorrectly classified examples whereas SVM penalizes incorrectly classified examples and correctly classified examples that lie within the margin. Also, it can be computationally expensive to reach the SVM solution by gradient descent due to quadratic computations and the regularization term. If the training set is not separable perceptron fails while svm finds the minimum cost separation.