Ass3 Task 2.1

01 June 2023 17:33

1. Consider a training data set with N input vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ with corresponding target labels $y_1, \ldots, y_N$, where $y_i \in \{-1, +1\}$. The maximum margin solution for linear support vector classifiers without slack variables is found by solving

$$\min_{\boldsymbol{w}, b} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1} \max\left(0, 1 - y_i(\boldsymbol{w}^T\phi(\boldsymbol{x}_i) + b)\right) \tag{1}$$

where $\phi(\boldsymbol{x}) = \boldsymbol{x}$ is the identity function. This is equivalent to the formulation we have seen in the lecture

$$J(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{m} \alpha_i \left[y^{(i)}\left(\boldsymbol{w}^T\boldsymbol{x}^{(i)} + b\right) - 1\right],$$

where $\alpha_i \geq 0$ are Lagrange multipliers. Although $f(x) = \max(0, x)$ is a convex function, it is not differentiable. However, it is possible to express the gradient in a piece-wise fashion:

$$\frac{\partial f}{\partial x} = \begin{cases} \frac{\partial x}{\partial x}, & \text{if } x \geq 0 \\ \frac{\partial 0}{\partial x} & \text{else} \end{cases} = \begin{cases} 1, & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases}$$

Derive the gradients for $\boldsymbol{w}$ and b and fill in the missing code of the gradient descent routine. *Hint:* Don't forget to apply the chain rule of calculus!

Task: Find gradient for w and b of our SVM objective function

$$F = \underbrace{\frac{1}{2} w^T w}_{\text{regularization term}} + \underbrace{C \sum_{i=1} \max(0, 1 - y_i(w^T \phi(x_i) + b))}_{\text{hinge loss term}}$$

Derive gradient for w:

$$\frac{\partial F}{\partial w} = \frac{1}{2} \underbrace{\frac{\partial}{\partial w}(w^T w)} + \frac{\partial}{\partial w}\left(C \sum_{i=1} \max(0, 1 - y_i(w^T \phi(x_i) + b))\right)$$

$$w^T w = w_1 \cdot w_1 + w_2 \cdot w_2 + \ldots + w_n \cdot w_n$$
$$= w_1^2 + w_2^2 + \ldots + w_n^2$$
$$\frac{\partial}{\partial w} w^T w = \left[\frac{\partial}{\partial w}(w_1^2), \ldots, \frac{\partial}{\partial w}(w_n^2)\right] = [2w_1, 2w_2, \ldots, 2w_n] = 2w$$

$$= \frac{1}{2} \cdot 2w + C \sum_{i=1} \max(0, \underbrace{\frac{\partial}{\partial w}(1 - y_i(w^T \overbrace{\phi(x_i)}^{\phi(x_i) = x_i} + b))})$$

$$= \frac{\partial}{\partial w} 1 - \frac{\partial}{\partial w} y_i (w^T x_i) - \frac{\partial}{\partial w} y_i \cdot b$$

$$= -y_i \cdot x_i$$

$$= w + C \sum_{i=1} \max(0, -y_i \cdot x_i)$$

$$\frac{\partial F}{\partial w} = w + C \sum_{i=1} \begin{cases} -y_i \cdot x_i, & \text{if } (1 - y_i(w^T \phi(x_i) + b)) > 0 \\ 0, & \text{else} \end{cases}$$

Derive gradient for b

$$\frac{\partial F}{\partial b} = \frac{1}{2}\frac{\partial F}{\partial b}(w^T w) + \frac{\partial F}{\partial b}\left(C \sum_{i=1} \max(0, 1 - y_i(w^T \phi(x_i) + b)\right)$$

$$= C \sum_{i=1} \max\left(0, \frac{\partial F}{\partial b} 1 - \frac{\partial F}{\partial b} y_i(w^T \phi(x_i)) - \frac{\partial F}{\partial b} y_i \cdot b\right.$$

$$= C \sum_{i=1} \max(0, -y_i)$$

$$\frac{\partial F}{\partial b} = C \sum_{i=1} \begin{cases} -y_i, & \text{if } (1 - y_i(w^T \phi(x_i) + b)) > 0 \\ 0, & \text{else} \end{cases}$$