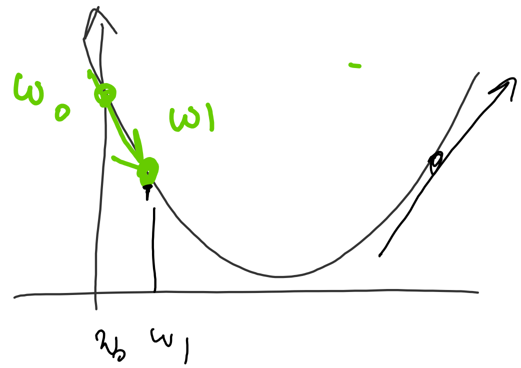
Lecture 10/8: Stochastic Gradient Descent

Tuesday, October 8, 2019 2:47 PM

Convexity



$$w^1 \leftarrow w_0 - \eta \nabla L(w)$$

$$\nabla L(w) = -1$$

$$\overset{=0}{w_0} - (-1) = 1$$

$$\nabla \mathcal{L}_S(\vec{w}) = \nabla\left(R(\vec{w}) + \frac{C}{n}\sum_i L(\vec{w}; \vec{x}_i, y_i)\right)$$

$$= \nabla R(\vec{w}) + \frac{C}{n}\sum_i \nabla L(\vec{w}, \vec{x}_i, y_i)$$

$$R(\vec{w}) = \frac{1}{2}\vec{w}\cdot\vec{w} \qquad \nabla R(\vec{w}) = \frac{2}{2}\vec{w} = \vec{w}$$

gradient

$$\max\left(1 - y_i(\vec{w}\cdot\vec{x}_i), 0\right) = \begin{cases} 0 & y_i(\vec{w}\cdot\vec{x}_i) \geq 1 \Rightarrow (0, 0, \ldots, 0) \\ \ldots \end{cases}$$

$$\left(1 - y_i(w \cdot x_i)\right) \quad y_i(w \cdot x_i) < 1 \implies -y_i x_i$$

$$\vec{w}^{(t+1)} \leftarrow \underbrace{\vec{w}^{(t)} - \eta_t \vec{w}^{(t)}}_{} - \eta_t \frac{C}{n} \sum_{i=1}^{n} -y_i \vec{x}_i \mathbb{1}\left(y_i (\vec{w}^{(t)} \cdot \vec{x}_i) \leq 1\right)$$

$$\underline{(1-\eta_t)} w^{(t)} + \underline{\eta_t \frac{C}{n}} \sum_{i=1}^{n} y_i \vec{x}_i \; \mathbb{1}\left(y_i (\vec{w}^{(t)} \cdot \vec{x}_i) \leq 1\right)$$

Weighting Scheme

Stochastic Gradient Descent SVM

$$\vec{w}^{(t+1)} \leftarrow (1-\eta_t) w^{(t)} + \eta_t C \; y_i \vec{x}_i \; \underline{\mathbb{1}\left(y_i (\vec{w}^{(t)} \cdot \vec{x}_i) \leq 1\right)} \qquad (\vec{x}_i, y_i) \sim S$$

$$\text{Indicator} = \begin{cases} 1 & \text{if condition is met.} \\ 0 & \text{o/w.} \end{cases}$$

At each step $(\vec{x}, y) \sim S$

If $(\vec{w}^{(t)} \cdot \vec{x}) y \leq 1$ then Update.

$$\vec{w}^{(t+1)} \leftarrow (1-\eta_t) \vec{w}^{(t)} + \eta_t C \; y_i \vec{x}_i$$

Else $\vec{w}^{(t+1)} \leftarrow (1-\eta_t) \vec{w}^{(t)}$.

$$\min \; \overset{\circ}{\longmapsto} \; + \frac{1}{\cdot} \sum^{n} \max\left(0, -y_i \vec{w} \cdot \vec{x}_i\right) \qquad (\vec{w} \cdot \vec{x}) y \geq 0$$

$\vec{w}$ Regularization n $i=1$

$\nabla L = 0$

gradient

$$\max(0, -y_i \vec{w} \cdot \vec{x}_i) = \begin{cases} 0 & y_i \vec{w} \cdot \vec{x}_i > 0 \\ -y_i (\vec{w} \cdot \vec{x}_i) & y_i \vec{w} \cdot \vec{x}_i \leq 0 \end{cases}$$

0

$-y_i \vec{x}_i$

$$w^{(t+1)} \leftarrow w^{(t)} - \eta_t (-y_i \vec{x}_i) \; \mathbb{1}(y_i \vec{w} \cdot \vec{x}_i \leq 0)$$

Alg: $(\vec{x}, y) \sim S$

If $(\vec{w}^{(t)} \cdot \vec{x}) y \leq 0$ Update

$\vec{w}^{(t+1)} \leftarrow \vec{w}^{(t)} + \eta_t \, y_i \vec{x}_i$

Perceptron