Gradient Descent

* general solution to optimization problem

Objective function | $\min_{w,b} L(w,b) = \min_{w,b} \sum_{n=1}^{N} \mathbb{I}(y_n(w^T x_n + b) < 0) + \lambda R(w,b)$

loss function   regularizer

Gradient Descent

loss function

$g^{(k)} \leftarrow \nabla_z F|_{z-1}$ the gradient of variable trying to optimize

$z^{(k)} \leftarrow z^{(k-1)} - \eta^{(k)} g^{(k)}$

Exponential Loss

$L(w,b) = \sum_n (\exp[-y_n(w \cdot x_n + b)]) + \frac{\lambda}{2}\|w\|^2 \rightarrow L_p$ norm, $p=2$ $\|w\|^2 = w_1^2 + w_2^2 + \ldots + w_D^2$

$$\frac{\partial L}{\partial b} = \sum_n \frac{\partial}{\partial b}(-y_n(w \cdot x_n + b)) \cdot \exp[-y_n(w \cdot x_n + b)] + \frac{\partial}{\partial b}\left[\frac{\lambda}{2}\|w\|^2\right]$$
$$= \sum_n -y_n \cdot \exp[-y_n(w \cdot x_n + b)] + 0$$

$$\frac{\partial L}{\partial b} = -\sum_n y_n \exp[-y_n(w \cdot x_n + b)]$$

$$\nabla_w \|w\|^2 = \begin{bmatrix} \frac{\partial}{\partial w_1}[\|w\|^2] \\ \frac{\partial}{\partial w_2}[\|w\|^2] \\ \frac{\partial}{\partial w_3}[\|w\|^2] \\ \vdots \\ \frac{\partial}{\partial w_D}[\|w\|^2] \end{bmatrix} = \begin{bmatrix} 2w_1 \\ 2w_2 \\ 2w_3 \\ \vdots \\ 2w_D \end{bmatrix} = 2w$$

$$\nabla_w L = \nabla_w \sum_n (\exp[-y_n(w \cdot x_n + b)]) + \nabla_w \frac{\lambda}{2}\|w\|^2$$
$$= \sum_n (\nabla_w \exp[-y_n(w \cdot x_n + b)]) + \lambda w$$
$$= \sum_n ((-y_n x_n)\exp[-y_n(w \cdot x_n + b)]) + \lambda w$$

$$\nabla_w L = -\sum_n y_n x_n \exp[-y_n(w \cdot x_n + b)] + \lambda w$$

---

negative

$b_{new} = b_{prev} - \eta \cdot \frac{\partial L}{\partial b}$

$W_{new} = W_{prev} - \eta \nabla_w L$ $W \in \mathbb{R}^{D \times 1}$, $\nabla_w L \in \mathbb{R}^{D \times 1}$

$$\nabla_w w = \begin{bmatrix} \frac{\partial}{\partial w_1}[w_1] \\ \frac{\partial}{\partial w_2}[w_2] \\ \vdots \\ \frac{\partial}{\partial w_D}[w_D] \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

---

matrix: $\|w\|^2 = w^T w$