

perceptron algorithm

$$f(x) = \text{sign}(wx + b)$$

$$L(w, b) = \max(0, -y(wx + b)) \rightarrow \text{hinge loss}$$

update rule: $w_j := w_j - \alpha \frac{\partial L}{\partial w_j}$

$$b := b - \alpha \frac{\partial L}{\partial b}$$

$$\frac{\partial L}{\partial w_j} = \begin{cases} -y x_j & \text{if } y(wx + b) \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial L}{\partial b} = \begin{cases} -y & \text{if } y(wx + b) \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\Rightarrow w := w + \alpha y x \quad \text{and} \quad b := b + \alpha y$$

example

Sample	x_1	x_2	y
1	2	3	+1
2	1	1	-1
3	2	1	-1
4	3	3	+1
5	5	5	+1

initialize: $\omega = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $b = 0$ $\alpha = 1$

↑ ignore

iteration #1 $x = [2 \ 3]$ $y = +1$

$$f(x) = \text{sign}\left(\underbrace{[2 \ 3]}_x \begin{bmatrix} 0 \\ 0 \end{bmatrix}_\omega + \begin{bmatrix} b \\ 0 \end{bmatrix}\right) = 0 \Rightarrow \text{wrong}$$

↑
prediction

$$\omega = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + (1) \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

↓
y
↓
x

$$b = 0 + (1) = 1$$

iteration #2 $x = [1 \ 1]$ $y = -1$

$$f(x) = \text{sign}\left([1 \ 1] \begin{bmatrix} 2 \\ 3 \end{bmatrix} + 1\right) = +1 \Rightarrow \text{wrong}$$

$$\omega = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + (-1) \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$b = 1 + (-1) = 0$$

iteration #3 $x = [2 \ 1]$ $y = -1$

$$f(x) = \text{sign}\left([2 \ 1] \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 0\right) = +1 \Rightarrow \text{wrong}$$

$$\omega = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + (-1) \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$b = 0 + (-1) = -1$$

iteration #4 $x = [3 \ 3]$ $y = +1$

$$f(x) = \text{sign}([3 \ 3] \begin{bmatrix} -1 \\ 1 \end{bmatrix} - 1) = -1 \Rightarrow \text{wrong}$$

$$w = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + (1) \begin{bmatrix} 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

$$b = -1 + (1) = 0$$

iteration #5 $x = [5 \ 5]$ $y = +1$

$$f(x) = \text{sign}([5 \ 5] \begin{bmatrix} 2 \\ 4 \end{bmatrix} + 0) = +1 \Rightarrow \text{correct}$$

no update!

and so on ...

linear sum

- if x is on the decision boundary:

$$wx + b = 0$$

- if x is on the margin:

\downarrow
support vector

$$wx + b = \pm 1$$

+ve: upper margin
-ve: lower margin

- distance from x to the boundary:

$$d = \frac{|w x + b|}{\|w\|}$$

- we want the margin to be 1 unit away from decision boundary $\Rightarrow w x + b = 1$

$$\Rightarrow d = \frac{1}{\|w\|} \text{ for one "side"}$$

$$\Rightarrow \frac{2}{\|w\|} = \text{margin distance}$$

- goal: $\max_w \frac{2}{\|w\|}$ or $\min_w \frac{\|w\|^2}{2}$
- makes it convex
(global minima)
- simplifies derivation

- constraint: $y_i (w x_i + b) \geq 0 \quad \forall i$

- Lagrange multiplier stuff bc constraints: (whatever that means)

$$\underbrace{L(w, b, \alpha)}_{\substack{\text{minimize} \\ \text{maximize}}} = \frac{\|w\|^2}{2} - \sum_{i=1}^n \alpha_i (y_i (w x_i + b) - 1)$$

- too many variables \therefore change primal form to dual formation by setting $\frac{dL}{dw} \& \frac{dL}{db} = 0$

$$\frac{dL}{dw} = w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^n \alpha_i y_i x_i$$

w to substitute

$$\frac{\partial L}{\partial b} = - \sum_{i=1}^n \alpha_i y_i = 0 \rightarrow \sum_{i=1}^n \alpha_i y_i = 0$$

- we have a new w. replace in primal form:

$$\frac{\|w\|^2}{2} = \frac{1}{2} \left\| \sum_{i=1}^n \alpha_i y_i x_i \right\|^2$$

$$= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

- for the other term in $L(w, b, \alpha)$:

$$\sum_{i=1}^n \alpha_i \left(y_i (w x_i + b) - 1 \right) = \sum_{i=1}^n \underline{\alpha_i y_i (w x_i + b)} - \sum_{i=1}^n \alpha_i$$

has to be 0 from $\frac{\partial L}{\partial b}$ above

$$= - \sum_{i=1}^n \alpha_i$$

- finally:

$$L(\alpha) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^n \alpha_i$$

goal : $\max_{\alpha} L(\alpha)$

linear regression

$$y = X\beta + \varepsilon \quad \text{s.t.} \quad X \in \mathbb{R}^{n \times d} \quad \text{and} \quad y \in \mathbb{R}^{n \times 1}$$

goal: $\min_{\beta} \|y_i - X_i\beta\|_2^2$

Ridge : $+ \lambda \|\beta\|^2$ Lasso : $+ \lambda |\beta|$ elastic : $+ \lambda \|\beta\|_1$