Perceptron training $w \in \mathbb{R}^{D \times 1}$

* find hyperplane to separate positive and negative examples in a linearly separable dataset

for (x,y)
  $a \leftarrow (\sum_{d=1}^{D} w_d x_d) + b$
  if $ya \leq 0$:
    $w_d \leftarrow w_d + y x_d$ for all $d = 1 \dots D$
    $b \leftarrow b + y$
  end if

this is gradient descent

$w_d \leftarrow w_d - ((-1) y x_d)$ $\nabla_w L$

$b \leftarrow b - ((-1)y)$ $\frac{\partial L}{\partial b}$

Perceptron uses the hinge loss variation

$\max\{0, 1 - y_n(w^T x_n)\}$ is the hinge loss function

we will use:

$\max\{0, -y_n(w^T x_n + b)\}$ implement the max function in code with an if statement

Objective:

$$\min_{w,b} L(w,b) = \min_{w,b} \sum_{n=1}^{N} \max\{0, -y_n(w^T x_n + b)\}$$

→ use gradient descent

$$\frac{\partial L}{\partial b} = \frac{\partial}{\partial b} \max\{0, -y_n(w^T x_n + b)\}$$

$$\frac{\partial L}{\partial b} = \begin{cases} 0 & \text{if } y_n(w^T x_n + b) > 1 \\ -y_n & \text{otherwise} \end{cases}$$

$$\nabla_w L = \nabla_w \max\{0, -y_n(w^T x_n + b)\}$$

$$\nabla_w L = \begin{cases} 0 & \text{if } y_n(w^T x_n + b) > 1 \\ -y_n x_n & \text{otherwise} \end{cases}$$

Testing

$a \leftarrow \sum_{d=1}^{D} [w_d x_d] + b$

$\text{sign}(a)$