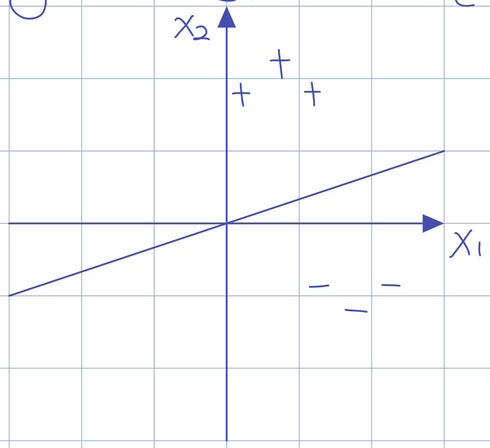
Binary Learning Classifier

$D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$

x is feature vector

y is class label $\in \{-1, 1\}$



1) choose fn from space of fns
2) metric that check if fn is good
3) method to improve fn

1) $h(x) = \text{sign}(wx + b) = \hat{y}$ (Predicted Output)

w is weight parameter $\qquad x \in \mathbb{R}^d$

b is the bias parameter

$w = [b \;\; w_1 \;\; w_2 \;\; \cdot \;\; \cdot \;\; \cdot \;\; w_d] \in \mathbb{R}^{d+1}$

$x = [1 \;\; x_1 \;\; x_2 \;\; \cdot \;\; \cdot \;\; \cdot \;\; x_d] \in \mathbb{R}^{d+1}$

d (number of dimension of the feature vector)

2D (Line) 3D (Hyperplane)

2) Zero-One Loss

$$0-1\ Loss = \begin{cases} 0, & \hat{y} = y \\ 1, & \hat{y} \neq y \end{cases}$$

w^* , b^*

$$E_{in} = \frac{1}{N} \sum_{n=1}^{N} L_{0-1}(h(x), y)$$

3) Perceptron Learning Algorithm (PLA)

while (any $y_n \neq sign(w^T x_n)$):

  for (n from 1 to N):

    $\hat{y} = sign(w_t^T x_n)$ Current Weight Vector

    if ($\hat{y} \neq y$):

New Weight Vector → $w_{t+1} = w_t + y_n x_n$ (PLA Update)

    break

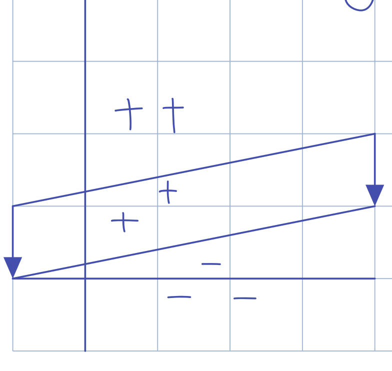
$W = W_{t+1}$

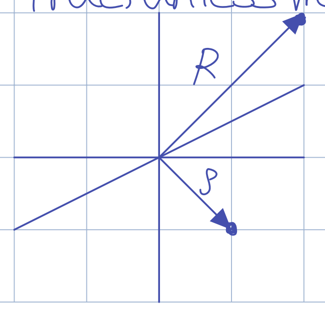
$\downarrow$

W^*

Perceptron Convergence Theorem

True, unless we know D is linear separable





Misclassified

$y_n(w^T x_n) < 0$

$w_{k+1} = w_k + y_n x_n$

$$
\begin{aligned}
y_n(w_{k+1}^T x_n) &= y_n((w_k + y_n x_n)^T x_n) \\
&= y_n(w_k^T x_n + y_n x_n^T x_n) \\
&= y_n(w_k^T x_n + y_n \|x_n\|_2^2) \\
&= y_n x_k^T x_n + y_n^2 \|x_n\|_2^2 \\
&= y_n x_k^T x_n + \|x_n\|_2^2 > y_n w_k^T x_n
\end{aligned}
$$

$$\|v\|_2 = \sqrt{v_1^2 + v_2^2 + \dots + v_d^2}$$

1. $\rho = \min_{1 \le n \le N} y_n(w^{*T} x_n) > 0$

2.
$$
\begin{aligned}
w(t)^T w^* &= (w(t-1) + y_n x_n)^T w^* \\
&= w(t-1)^T w^* + y_n x_n^T w^* \\
&\geq w(t-1)^T w^* + \min_{1 \le n \le N} y_n x_n^T w^* \\
&= w(t-1)^T w^* + \rho
\end{aligned}
$$

Using $w(0) = 0$

$w(1)^T w^* \geq \rho$

$w(2)^T w^* \geq w(1)^T w^* + \rho = 2\rho$

3. $\|w(t)\|^2 \leq \|w(t-1)\|^2 + \|x(t-1)\|^2$