DS 4440 Supervised Learning (Review)

Given Training data (X, y)

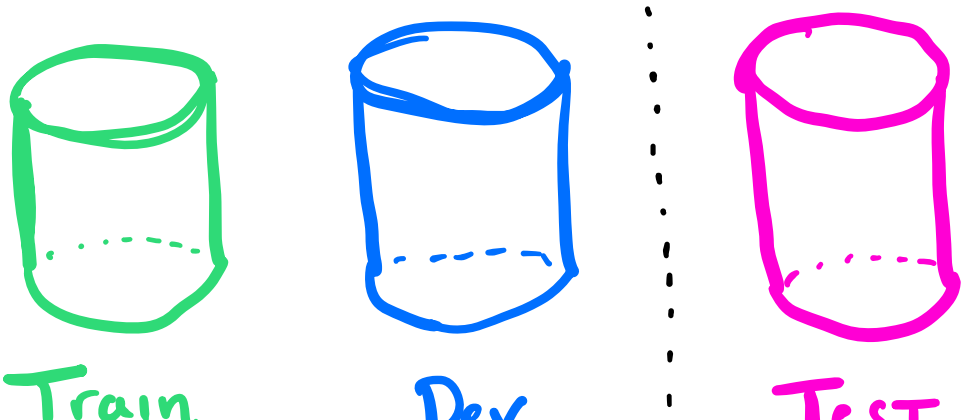
$$X = \{x_1 \dots x_N\}$$

$$= \begin{bmatrix} x_{11} & \dots & x_{1D} \\ \dots & \ddots & \dots \\ x_{N1} & & x_{ND} \end{bmatrix}_N^D \quad X \in \mathbb{R}^{N \times D}$$

$$y = [y_1 \dots y_N] \quad y \in \mathbb{R}^N$$

The goal: find Parameters θ for function f_θ s.t. $f_\theta(X) = y$

This should generalize beyond X.



Find $\hat{\theta}$

Initial eval

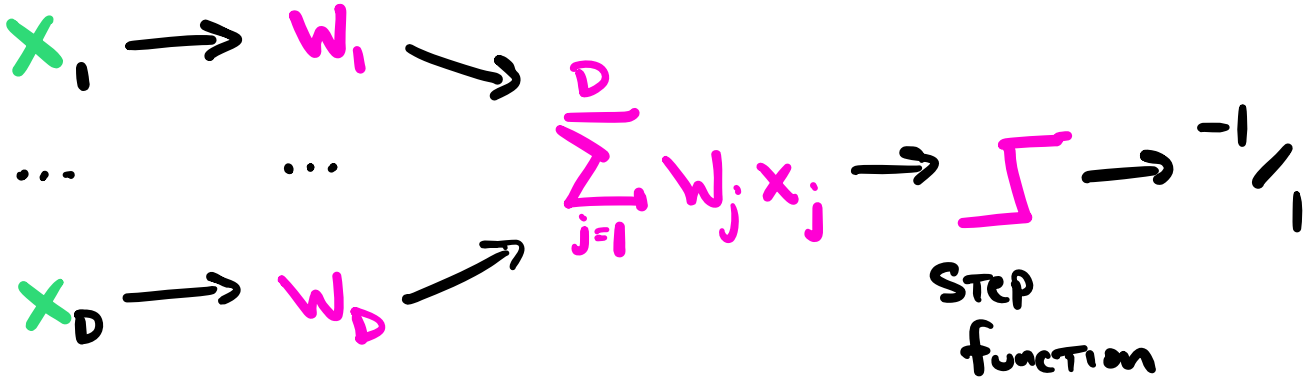
Does it generalize?

Learning means finding a $\hat{\theta}$ s.t. loss ($\mathcal{L}$) on (X, y) is minimized.

$$\arg\min_{\hat{\theta}} \sum_{i=1}^{N} \mathcal{L}(\underbrace{f_{\hat{\theta}}(x_i)}_{\hat{y}_i}, y_i)$$

Machine Learning

One choice The Perceptron



$$\hat{y}_i = \begin{cases} 1 & \text{if } w x_i > \tau \\ -1 & \text{otherwise} \end{cases}$$

The loss is 0/1.

$$\ell(\hat{y}_i, y_i) = \begin{cases} \emptyset & \text{if } \hat{y}_i = y_i \\ 1 & \text{otherwise} \end{cases}$$

Learning ($\theta = w$)

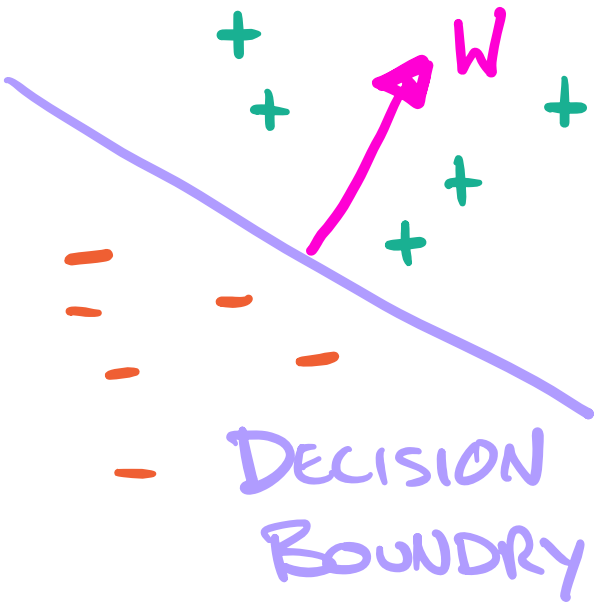
$w \leftarrow \emptyset$

for $(x_i, y_i) \in (x, y)$

if $\mathbb{1}(w x_i > \tau) \neq y_i$

$w \leftarrow w + y_i x_i$

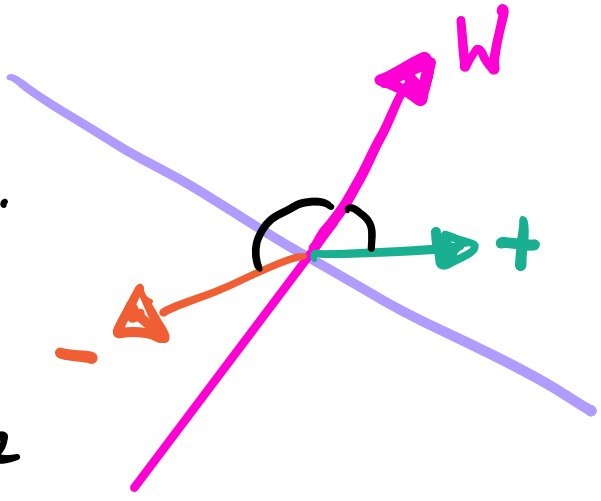
Why does this Work?!



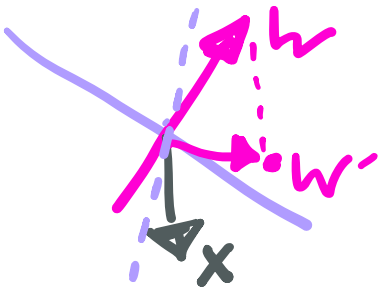
Weight vector is $\perp$ to the decision boundary.

Consider where $wx_i = \emptyset$.

Positive instances → acute angles.



Suppose we make an incorrect prediction for x that should be +



Add x to $w \rightarrow w'$

Reduces angle between w and x_i; Shifts —.

$$\cos\theta = \frac{w^T x}{\|w\| \|x\|}$$

$$\rightarrow w^T x = \|w\| \|x\| \cos\theta$$