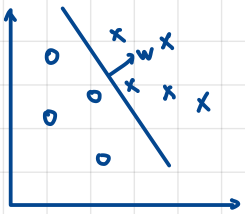
2. The Perceptron (Binary Classification)

Assumption: There exists a hyperplane such that it seperates two classes of data



In lower dimensional spaces, this condition is restrictive

However in higher dimensional spaces, points are more sparse

Dimensionality aids perceptron

Hyperplane

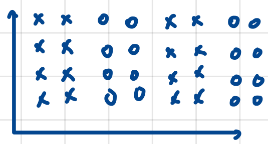
$H = \{x : w^T x + b = 0\}$

On opposite sides of the hyperplane, sign of $w^T x + b$ is opposite.

sign $(w^T x + b)$ is classification result

Collecting more data does not always improve performance

Sinusoidal distr



No hyperplane can divide the two classes

Adding more data does not help

Formalization

Labels: $y = \{-1, +1\}$

Need to lear w, b. Can we learn only w?

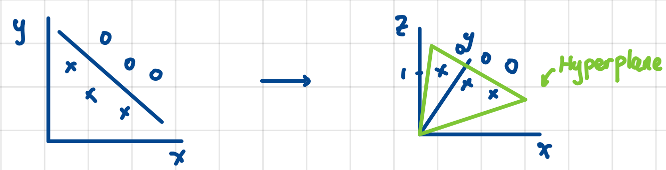
Perform transformation

$\vec{x}_i \to \begin{pmatrix} \vec{x}_i \\ 1 \end{pmatrix} = \vec{x}'$ $\quad \vec{w} \to \begin{pmatrix} \vec{w} \\ b \end{pmatrix} = \vec{w}'$

$\vec{w}' \vec{x}' = w^T x + b$

$H = \{x : w'^T x' = 0\}$

Geometrically



With additional dimension

Hyperplane is coerced to go thru origin

Algorithm

$\vec{w} = \vec{0}$

while true:

m = 0 # counter for # of misclassified pts

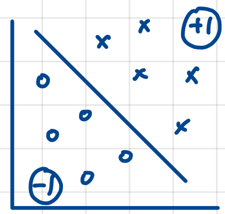
for x, y ∈ D

if $y w^T x \leq 0$: #makes an error

$w' = w + yx$

if m = 0:

STOP



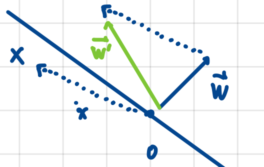
$y w'^T x = y(w^T + y x^T) x$

$= y w^T x + y \underbrace{x^T x}_{>0}$

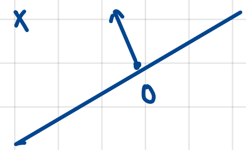
If $y > 0$, product will be bigger next time

$y < 0$, product smaller next time

Visualizing the Update Algorithm



Misclassified



Correctly Classified

If we are exposed to the same pt many times, max step before it gets it right?

$w^T x \leq 0$

$(w+x)^T x \leq 0 \quad (y=1)$

$\vdots$ k update

$(w+kx)^T x \leq 0$

$wx + kx^T x \leq 0$

$k \leq \frac{-w^T w}{x^T x}$ Can only get a fixed pt wrong finite times

Convergence Proof next time