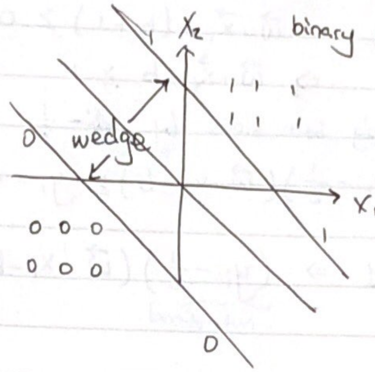
Support Vector Machine (SVM)

Assume Linear Separable

binary classification model



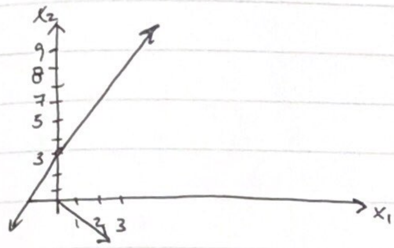
$\vec{w}\cdot\vec{x}-b=0$

$c\vec{w}\cdot\vec{x}-cb=0$

set $\delta=1$

$\forall\, c \in \mathbb{R}$

$\mathcal{H}=\{\mathbb{1}_{\underbrace{\vec{w}\cdot\vec{x}-b\geq 0}_{\text{Hesse Normal Form}}} : \vec{w}\in\mathbb{R}^p,\ b\in\mathbb{R}\}$



$l: x_2 = 2x_1+3$

$\Rightarrow 2x_1 - x_2 + 3 = 0$

$\underbrace{\begin{bmatrix}2\\-1\end{bmatrix}}_{\vec{w}}\cdot\underbrace{\begin{bmatrix}x_1\\x_2\end{bmatrix}}_{\vec{x}} = \underbrace{-3}_{b}$

where is $\vec{w}$ on the plot?

"Normal vector" it is ⊥ to the hipper plane.

$\|\vec{w}\| = \sqrt{\sum_{i=1}^{p} w_i^2} = \sqrt{\vec{w}\cdot\vec{w}}$

norm

$\vec{w}_0 = \dfrac{\vec{w}}{\|\vec{w}\|}$ the normal vector with unit length.

Let $\vec{z} = \alpha\vec{w}_0$

$\|\vec{z}\| = \sqrt{\sum_{j=1}^{p}(\alpha w_{0j})^2}$

$= \sqrt{\alpha^2 \sum_{j=1}^{p} w_{0j}^2}$

$= |\alpha|\,\|w_{0j}\| = |\alpha|$

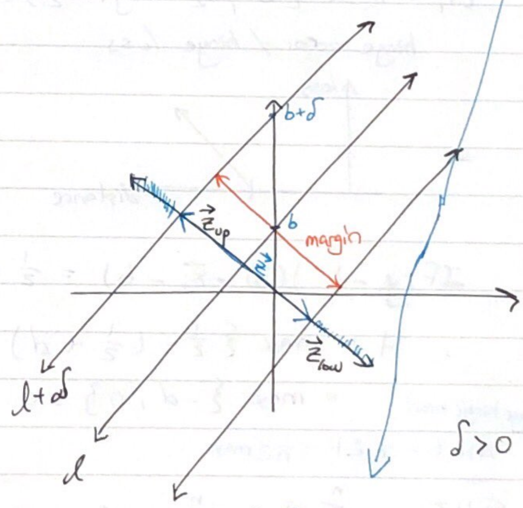
Find the $\vec{z}$ which from the origin to l.

$\vec{w}\cdot\vec{z}-b=0$

$\vec{w}\,(\alpha\vec{w}_0)-b=0$

$\alpha\,\dfrac{\vec{w}\cdot\vec{w}}{\|\vec{w}\|}=b$

$\Rightarrow \alpha\,\dfrac{\|\vec{w}\|^{\cancel{2}}}{\cancel{\|\vec{w}\|}}=b \Rightarrow \alpha=\dfrac{b}{\|w\|}$



$l_{(low)} = l - \cancel{\delta}\,1$

$\alpha_u: \vec{w}\cdot\alpha\vec{w}_0-(b+\delta)=0$

$\Rightarrow \alpha_u = \dfrac{b+\delta}{\|w\|}$

$\alpha_l \Rightarrow \alpha_l = \dfrac{b-\delta}{\|w\|}$

distance btw upper and low lines, "the margin"

fix $\delta=1$

margin $= \alpha_u - \alpha_l = \dfrac{\delta+f}{\|w\|} - \dfrac{b-f}{\|w\|} = \dfrac{2\delta}{\|w\|} = \dfrac{2}{\|w\|}$

maximize margin $\Longleftrightarrow$ minimize $\|w\|$

All $y_i = 1$ must have their x_i's above $l+1$

$\forall_i \quad y_i = 1 \qquad \vec{w} \cdot \vec{x}_i - (b+1) \geq 0$

$\Rightarrow \vec{w} \cdot \vec{x}_i - b \geq 1$

multiply both sides by $\cancel{y_i -} \frac{1}{2}$

$\Rightarrow (y_i - \frac{1}{2})(\vec{w} \cdot x_i - b) \geq y_i - \frac{1}{2}$

Since $y_i = 1 \Rightarrow \underbrace{(y_i - \frac{1}{2})}_{\text{intentional}}(\vec{w} \cdot x_i - b) \geq \frac{1}{2}$

All $y_i = 0$ must have their y_i's below $l - 1$.

$\vec{w} \cdot \vec{x}_i - (b-1) \geq 0$

$\Rightarrow \vec{w} \cdot \vec{x} - b + 1 \geq 0$

$\vec{w} \cdot \vec{x} - b \geq -1$

multiply both sides by $y_i - \frac{1}{2}$

$(y_i - \frac{1}{2})(\vec{w} \cdot \vec{x} - b) \geq -y_i + \frac{1}{2}$

Since $y_i = 0 \Rightarrow (y_i - \frac{1}{2})(\vec{w} \cdot \vec{x} - b) \geq \frac{1}{2}$

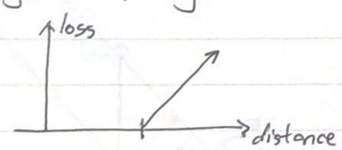
$\forall$ (for all) $i \quad (y_i - \frac{1}{2})(w \cdot \vec{x}_i - b) \geq \frac{1}{2}$

min s.t $\forall_i \; (y_i - \frac{1}{2})(\vec{w} \vec{x}_i - b) \geq 0$

solve $\vec{w}, b$

$H_i : \max \{0, \frac{1}{2} - (y_i - \frac{1}{2})(\vec{w} \cdot \vec{x}_i - b)\}$

hinge error / hinge loss



If $(y_i - \frac{1}{2})(\vec{w} \cdot \vec{x}_i - b) = \frac{1}{2} + d \geq \frac{1}{2}$

$H = \max \{\frac{1}{2} - (\frac{1}{2} + d), 0\}$

$= \max \{-d, 0\} = 0$

$d > 0$

If $(y_i - \frac{1}{2})(\vec{w} \cdot \vec{x}_i - b) = \frac{1}{2} - d \ngeq \frac{1}{2}$

$H = \max \{\frac{1}{2} - (\frac{1}{2} - d), 0\}$

$= \max \{+d, 0\} = +d$

(average hinge error)

$AHE = \frac{1}{n} \Sigma H_i = \frac{1}{n} \Sigma \max$

or

how far away

$SHE = \sum_{i=1}^{n} H_i = \sum_{i=1}^{n} \max \{0, \overbrace{\frac{1}{2} - (y_i - \frac{1}{2})(\vec{w} \cdot \vec{x}_i - b)}\}$.

sum of hinge error in $\mathbb{D}$

A : minimize SHE.

* Find $\vec{w}, b$ s.t. AHE is minimal

We want both minimal error (distance of the mistake from line) and maximum margin:
Vapnik (1963) proposal:

hyperparameter / tunning parameter.

$$\vec{w}, b \text{ in } g = \underset{\vec{w}, b}{\operatorname{argmin}} \{ \underbrace{AHE}_{\text{min aveg error.}} + \lambda \underbrace{\|w\|^2}_{\text{max wedge}} \}$$

$$g = \mathcal{A}(\mathbb{D}, \mathcal{H}, \lambda)$$

Building a model for ppl's favorite color.

$y = \{0, 1, \ldots, L\}, \quad L > 2$

nominal

{Red, green, blue, ... }

Null Model: $g = \text{Mode}[y]$

Consider the model

new data point, not inside of $\mathbb{D}$

a distance function: $d > 0$

$d(x,y) = d(y,x)$

$d(x,y) = 0$

$$g(x^*) = y_j \quad \text{s.t} \quad j = \underset{i=1,\ldots n}{\operatorname{argmin}} \{ d(\vec{x}^*, \vec{x}_i) \}$$

prediction of new data point.

Bob Bill

Nearest Neighbor Algorithm

Euclidean Metric

$$d(\vec{x}^*, \vec{x}_i) = \sqrt{\sum_{j=1}^{p} (x_j^* - x_{i,j})^2}$$

K-Nearess Neighbors (KNN)

$g(x^*) = \text{Mode}[y_1, \ldots, y_K]$ where $y_1 \ldots y_K$ correspond to the K closses observations in $\mathbb{D}$ to X_*.