Saturday, May 11, 2024 9:33 AM

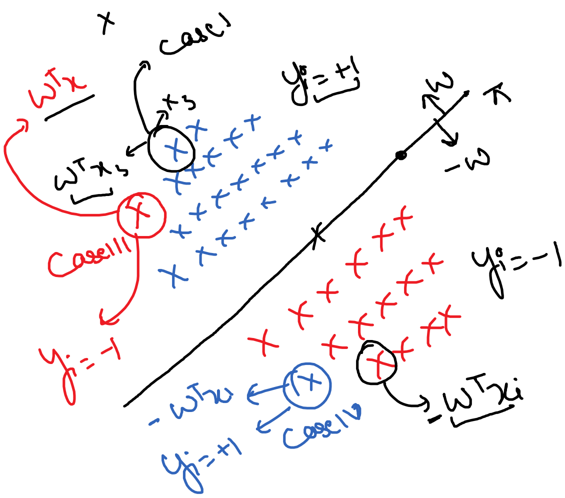
Logistic Regression

→ Binary classification

Geometric | Probabilistic | Loss minimization

Assume: data is linearly seperable

Geometric Intuition [$y_i = -1$ or $+1$]



eq^n of hyperplane ⇒

$$w_0 + \sum_{i=1}^{n} w_i^T x_i = 0$$

passing through origin,

$$w_0 = 0 \qquad y_1 = w^T x_i$$

eq^n ⇒ $w_i^T x_i = 0$

(I) $w^T x_i > 0$ → for the class

(II) $-w^T x_i < 0$ → for -ve class

Lets multiply y with $w^T x$ →

$$y_i w^T x_i$$

Case1: $y_i = +1$, $w^T x_i > 0$ Case2: $y_i = -1$, $\underbrace{w^T x_i < 0}_{\text{negative}}$

$y_i w^T x_i > 0$ $y_i w^T x_i > 0$

correct classification

Case IV: $y_i = +1$, $w^T x_i < 0$

$w^T x_i > 0$

Case 3: $y_i = -1, \quad w^T x_i > 0$

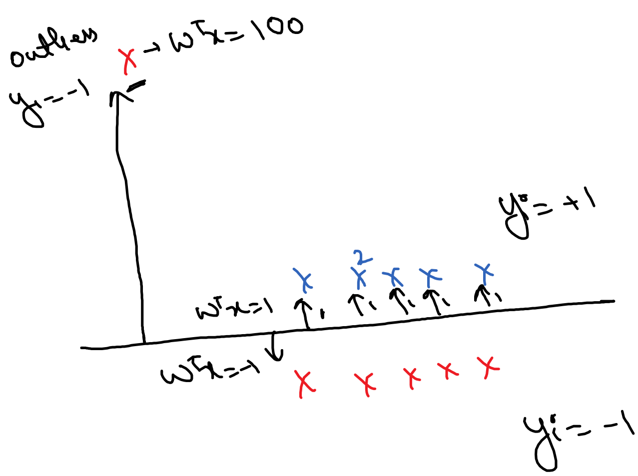
$y_i w^T x_i < 0$ → incorrect classification ← $y_i w^T x_i < 0$

$$\begin{cases} y_i w^T x_i > 0 & : \text{correct classification} \\ y_i w^T x_i < 0 & : \text{incorrect classification} \end{cases}$$

"Mathematical objective f^n"

$$\text{MOF} \Rightarrow \underset{w}{\operatorname{argmax}} \left(y_i w^T x_i \right) \Rightarrow w^*$$

effects of outliers on MOF:

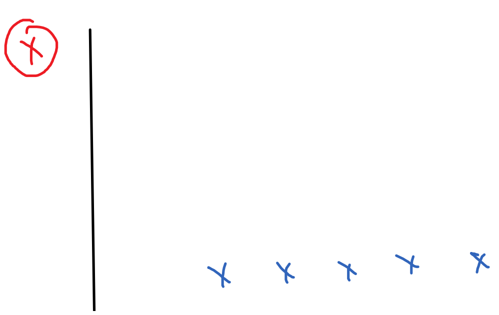


Not a good model

$$\text{MOF} = \sum y_i w^T x_i$$

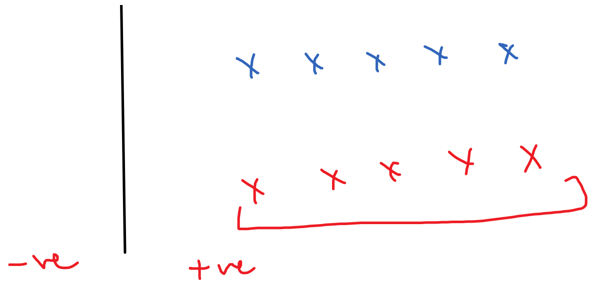
$$= 1+1+1+1+1+1+1+1+1+1 - 100$$

$$= -90$$



$\text{MOF} = +1$

Not a good model



Not a good model

Sigmoid function

- Probabilistic
- Squashes value b/w 0 & 1
- Robust to outliers

$$\sigma(x) = \frac{1}{1+e^{-x}}$$ ⇒ expression for sigmoid f^n

$$\sigma(y_i w^T x_i) = \frac{1}{1+e^{-y_i w^T x_i}}$$

$$\text{MOF} = \text{argmax}\left(\sigma(y_i w^T x_i)\right) \Rightarrow \text{argmax}\left(\frac{1}{1+e^{-y_i w^T x_i}}\right)$$

Take log

$$\Rightarrow \text{argmax}\left[\log\left(\frac{1}{1+e^{-y_i w^T x_i}}\right)\right]$$

$$\log\frac{1}{a} = -\log a$$

$$\Rightarrow \text{argmax}\left[-\log(1+e^{-y_i w^T x_i})\right]$$

max = (100)

$$\Rightarrow \text{argmax}\left[-\log(1+e^{-y_i w^T x_i})\right]$$

$$\Rightarrow \arg\max \left[-\log\left(1+e^{-y_i w^T x_i}\right) \right]$$

max = $-(100)$
$\downarrow$
min

MOF $\Rightarrow \arg\min \left[\log\left(1+e^{-y_i w^T x_i}\right) \right] \Rightarrow$ Logistic loss

for fun $\Rightarrow$ $\Rightarrow \log\left(e^{-y_i w^T x_i}\right)$

$\Rightarrow \arg\min\left(-y_i w^T x_i\right) \Rightarrow \arg\max\left(y_i w^T x_i\right)$

Squashes values b/w 0 & 1

$\downarrow$

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

$x = -\infty$:

$$\frac{1}{1+e^{-(-\infty)}} \rightarrow \frac{1}{1+e^{\infty}} \rightarrow \frac{1}{1+\infty} \rightarrow \frac{1}{\infty} = 0$$

$x = +\infty$:

$$\frac{1}{1+e^{-\infty}} \rightarrow \frac{1}{1+0} \rightarrow \frac{1}{1} \Rightarrow 1$$

$e^{-\infty} = 0$

Probabilistic Intuition [o/p = +1, 0]

$$P = \frac{1}{1+e^{-\gamma}}$$

$$p(1+e^{-y}) = 1$$

$$p + pe^{-y} = 1$$

$$pe^{-y} = 1-p$$

$$e^{-y} = \frac{1-p}{p}$$

$$\frac{1}{e^{y}} = \frac{1-p}{p}$$

$$\boxed{e^{y} = \left(\frac{p}{1-p}\right)} \rightarrow \text{odd's ratio}$$

<u>Take ln</u> on <u>both sides</u>

$$\ln e^{y} = \ln\left(\frac{p}{1-p}\right)$$

$$\boxed{y = \ln\left(\frac{p}{1-p}\right)} \rightarrow \text{Logit } f^n$$

$$\text{log loss} = -\left[\overbrace{\underset{\underset{y_i=1}{\downarrow}}{y_i} \log\{P(y_i)\}}^{\rightarrow 0} + \underbrace{\overset{\rightarrow y_i=0}{(1-y_i)} \log\{P(1-y_i)\}}_{\rightarrow 0}\right]$$

if $y_i=1 \Rightarrow \text{log loss} = -y_i \log(P(y_i))$

if $y_i=0 \Rightarrow \text{log loss} = -[(1-y_i)\log(P(1-y_i))]$

lets say, $y_i \to$ prob. $\Rightarrow 1$
$\quad \hookrightarrow$ geo $\Rightarrow 1$

geo: $\log(1+e^{-y_i w^T x_i}) \Rightarrow \log(1+e^{-w^T x_i})$

$y_i = w^T x_i$

prob: $-[y_i \log p(y_i) + \cancel{(1-y_i)} \log p(1-y_i)]$ $\to 0$

$$-\log p(y_i) = -\log\left(\frac{1}{1+e^{-y_i}}\right)$$

$$\Rightarrow \log(1+e^{-y_i})$$

$$\Rightarrow \log(1+e^{-w^T x_i})$$

Derive Sigmoids from MoF:

$$\ln\left(\frac{1}{1-p}\right) = y_i w^T x_i$$

Take e on both sides

$$\exp\left[\ln\left(\frac{p}{1-p}\right)\right] = e^{y_i w^T x_i}$$

$$\frac{p}{1-p} = e^{y_i w^T x_i}$$

$$p = (1-p)e^{y_i w^T x_i}$$

$$p = e^{y_i w^T x_i} - p e^{y_i w^T x_i}$$

$$p + p e^{y_i w^T x_i} = e^{y_i w^T x_i}$$

11:15am

$$P + P e^{y_i w^T x_i} = e^{y_i w^T x_i}$$

$$P(1 + e^{y_i w^T x_i}) = e^{y_i w^T x_i}$$

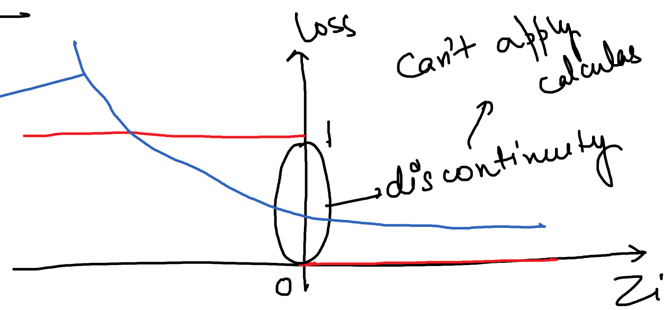
$$P = \frac{e^{y_i w^T x_i}}{1 + e^{y_i w^T x_i}} \Rightarrow \frac{\frac{e^{y_i w^T x_i}}{\cancel{e^{y_i w^T x_i}}}}{\frac{1 + e^{y_i w^T x_i}}{e^{y_i w^T x_i}}}$$

$$P = \frac{1}{\frac{1}{e^{y_i w^T x_i}} + \frac{\cancel{e^{y_i w^T x_i}}}{\cancel{e^{y_i w^T x_i}}}} = \frac{1}{1 + e^{-y_i w^T x_i}}$$

Probability of something happening

$$\boxed{P = \sigma(e^{-y_i w^T x_i})}$$

Loss Minimization (0-1 Loss)



logistic loss

continous
↳ apply calculas

$z_i = y_i w^T x_i$

$y_i w^T x_i < 0$
↳ incorrect

$y_i w^T x > 0$
↳ correct

Overfitting & Underfitting

$$w^* = \underset{w}{\text{argmin}} \sum_{i=1}^{n} \underbrace{\log(1 + e^{-y_i w^T x_i})}$$

$w \uparrow \rightarrow y_i w^T x_i \uparrow \rightarrow$ all correct classifications

$\downarrow$

$e^{-y_i w^T x_i} \downarrow$

$\downarrow$

- ations

$e^{-y_i w^T x_i}$ ↓

↓

loss will be 0

Regularization

① RIDGE REGULARIZER

np. linspace (0.001, 1) 20

$$\text{Loss} = \underset{w}{\operatorname{argmin}} \sum_{i=1}^{n} \log\left(1 + e^{-y_i w^T x_i}\right) + \lambda \left[w^T w \right] \uparrow$$

$w^T w$ → L^2 Norm

λ → hyperparameter (small value)

② LASSO Regularizer

$$\text{Loss} = \underset{w}{\operatorname{argmin}} \sum_{i=1}^{n} \log\left(1 + e^{-y_i w^T x_i}\right) + \lambda \|w\|$$

LASSO creates sparsity → Sparse Vector [1, 0, 0, 0, 0, 1, 0, 0, 0]

0.5 1 2 3 4

$(0.5)^2 (0.25)^2 (0.625)^2$

0.0 0 6 25 → 0

very close

$y = m_1 x_1 + 0\, x_2 + m_3 x_3 + 0\, x_4$

LASSO pulls out important columns

$\lambda \Rightarrow$ hyperparameter $\Rightarrow$ $\lambda = 0 \rightarrow$ overfitting

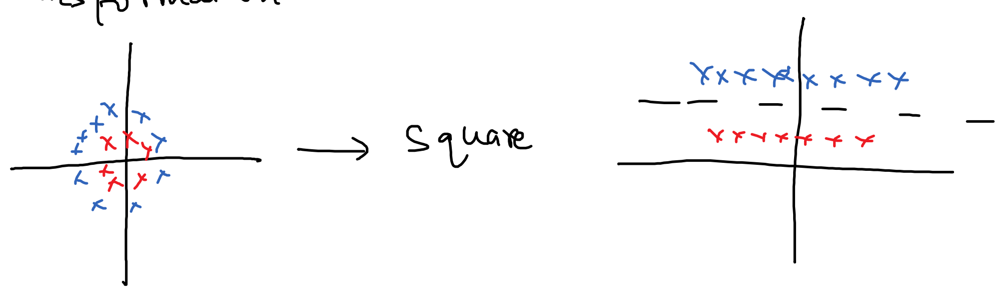
$\lambda = \text{high} \rightarrow$ underfitting

Features: → Always standardize → explore!

↳ linearly seperable

Feature transformation:

(1)



(2)