$\boxed{\text{Logistic Regression-2}}$

L $\boxed{\text{MLE}}$ - Maximum Likelihood Estimation

Loss $\rightarrow$ $-\log loss + \lambda \sum_{j=1}^{d} |w_d|$ - lasso

$\hat{y} = \sigma(\underline{w}^T . x + \underline{w_0})$

$-\log loss + \lambda \sum_{j=1}^{d} w_d^2$ - Ridge.

λ ✓

Regularization + logistic Regression

Validation $\rightarrow$ Hyperparameter tuning

SD - $\dfrac{x_i - \mu}{\sigma}$ ✓ ✓

μ - ? ✓

σ - ✓

<u>Accuracy</u> y $\hat{y}$ if $y == \hat{y}$ # correct prediction

$y \neq \hat{y}$ wrong prediction

y	$\hat{y}$	
→ 1	1 ✓	— Corrected
1	0	- wrong
0	0 ✓	
1	0 ✗	

$$\text{Accuracy} = \frac{\#\text{ of Correct Predictions}}{\text{Total }\#\text{ of Predictions}} = \frac{80}{100} = 0.8$$

80%

Regularization, Hyperparameter tuning.

Feature importance

$$\hat{y} = \frac{1}{1+e^{-(w_0+w_1x_1+w_2x_2 - \cdots w_d x_d)}} \quad w x_1$$

odds $\frac{\text{Prob of Success}}{\text{Prob of failure}}$

$$\frac{0.5}{0.5} \quad ? \quad 1:1$$

$\frac{\text{Prob of Success}}{\text{Prob of failure}} \quad \frac{4/5}{1/5} \quad \underline{4} : \underline{1}$

feature imp.

Linear Reg.

$$\hat{y} = w_0 + w_1 x_1$$

$\hat{y}\uparrow \quad w_1 \quad x_1\uparrow$ 1unit

$$z = w_0 + w_1x_1 + \cdots w_d x_d$$

$\therefore$

$p(y=1|x)$ P

$$p = \frac{1}{1+e^{-(z)}}$$

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

$$p + p\cdot e^{-z} = 1$$

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

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

$$-z = \log\left(\frac{1-p}{p}\right)$$

$$z = \log\left(\frac{p}{1-p}\right)$$

$$\boxed{\log\left(\frac{p}{1-p}\right) = w_0 + w_1x_1 + \cdots w_d x_d}$$

$$\frac{p}{1-p} - \frac{\text{Prob of } y=1 \text{ (succm)}}{\text{prob of } y=0 \text{ (failure}} = \text{odds?}$$

prob $y=1\uparrow$ w+ve

or $y=1\downarrow$ w-ve

$$\log(\text{odds})\uparrow = w_0 + \overset{2.6}{w_1} x_1\uparrow + \overset{1.2}{w_2} x_2 \cdots \overset{-0.9}{w_d} x_d$$

w_1 $x_1\uparrow$

Predict heart failure

$P(y=1 \mid \text{heartfile}) \uparrow = w_0 + w_1 \text{ age} \uparrow + w_2 \text{ BMI} \uparrow - w_3 \text{ good col (HDL)} \uparrow$

$p(y=0 \mid \text{No heart})$

$\boxed{\log(\text{odds})}$ — w ?

w high value - very imp

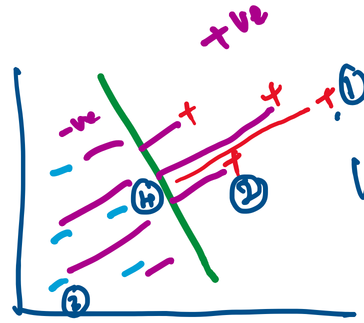
w - small - less impt

w +ve - increase odds, increases churn risk

w -ve - decreases churn risk

$\boxed{10:11}$

Geometrical intution.



high prob $y=1$

$$Z = \frac{w^T \cdot x + w_0}{\|w\|}$$

$\boxed{\|w\| = 1}$

$$\log\left(\frac{p}{1-p}\right) = Z \uparrow$$

$$\boxed{\frac{\text{prob of } y=1}{\text{Prob of } y=0}}$$

$$\log\left(\frac{p}{1-p}\right) = Z$$

$$p = \boxed{\frac{1}{1+e^{-z}}}$$ -? Sigmoid.

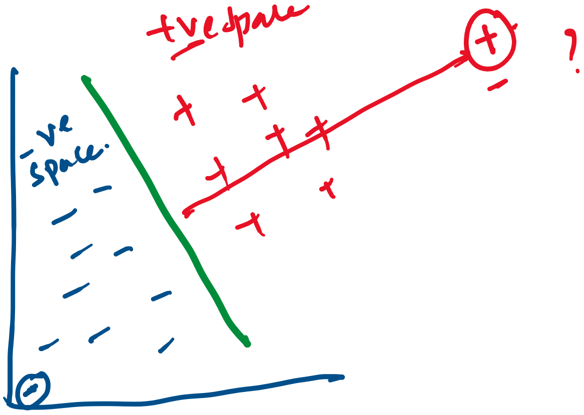
Assumptions ?

EDA ✓

① NO multicollinearity

② outlier effect.

outlier is on the same half space.

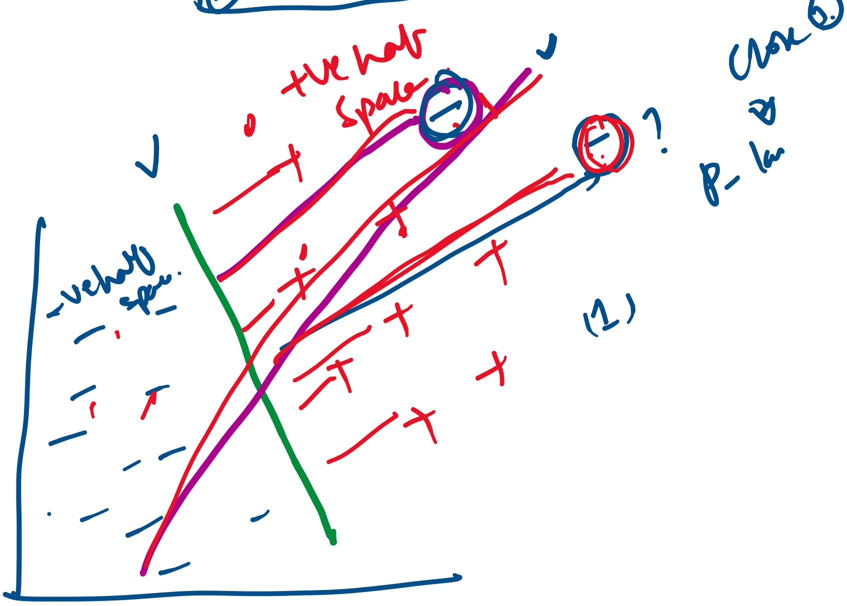


+ve   +ve half

-ve   -ve half

— (0)

Metric?   ~~log(p_i)~~



log loss

$$-\left[y_i \log p_i + (1-y_i)\log(1-p_i) \right]$$

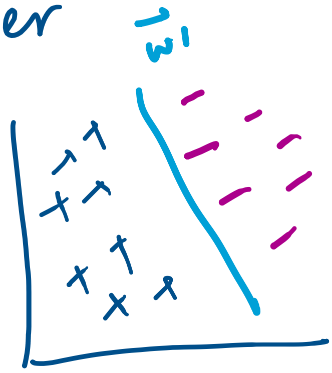
0 ~~log(p_i)~~   1 log(1−0.99)

log(0.01)

$$-\left[\boxed{4.6} \ ? \right]$$

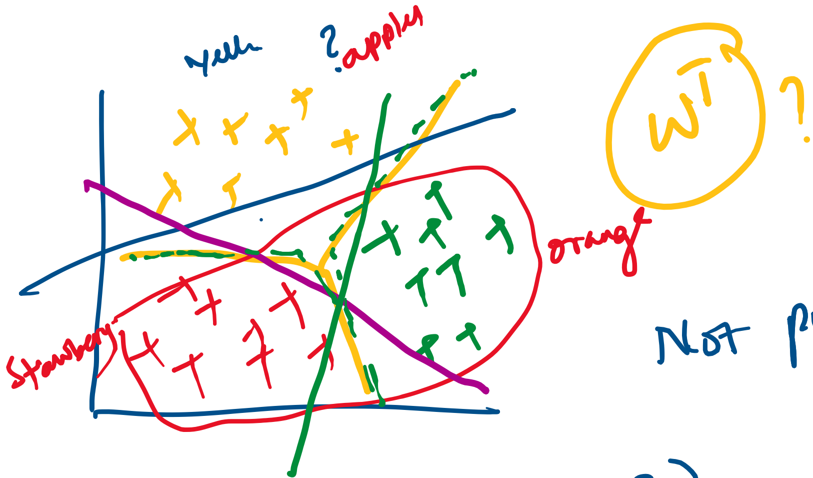
① if outlier are on same half space
No impact

② if outlier are on opp half space
hyperplane will twisted to reduce error of outlier



Binary Classification

Multiclass Classification   logistic Regression?



Not possible

Hack : One vs rest (OVR)

X

y	Y_Orange	Y_Apple	Y_Stawberry
Apple	0	1	0
orange	1	0	0
strawberry	0	0	1
apple	0	1	0
orange	1	0	0

$x \rightarrow \hat{y}$

x, Y_orange

$M_1 \rightarrow P(Y=\text{orange})$ $\quad \hat{y}_{\text{orang}}$ $\quad \boxed{0.9}$ ✓

x, Y_Apple

$M_2 \rightarrow P(\text{Apple})$ $\quad \hat{y}_{\text{apple}}$ $\quad 0.1$

x, Y_Straw

$M_3 \rightarrow P(Y = \text{Strawberry}$ $\quad \hat{y}_{\text{strawberry}}$ $\quad 0.15$

Arg Max

0.1   0.22
0.15   0.25
0.12   0.24
0.17   0.26
$\boxed{0.5}$?   ⋮
0.18

one

P(orang) – 0.7
P(Apl) 0.3
P(straw) = 0.25

Soft Max :

$$P(\text{oran}) \quad \frac{0.7}{0.7+0.3+0.25} \qquad \frac{0.3}{0.7+0.3+0.25} \qquad \frac{0.25}{0.7+0.3+0.25}$$

$\boxed{0.69}$ 0.2 $\boxed{0.11}$

x	y	A	st
	0.69	0.2	0.11