Recap

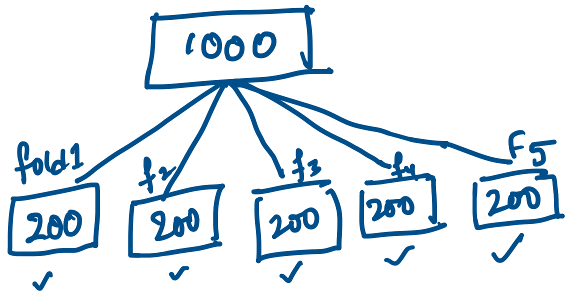
① Cross Validation ✓

K-fold Cross Validation

$m = 1000$

$K = 5$ ✓ $\boxed{\text{5-folds}}$ ✓

if $K = 10$. $\boxed{\text{10-folds}}$ ✓



polynomial(1)

Training set		Validation set	Training R^2	Validation R^2
[F_1 F_2 F_3 F_4] $K-1$	m_1	F_5 ✓ K^{th}	0.91	0.89
F_2 F_3 F_4 F_5		F_1	0.92	0.9
F_3 F_4 F_5 F_1		F_2	0.91	0.88
F_4 F_5 F_1 F_2		F_3	0.93	0.9
F_5 F_1 F_2 F_3		F_4	0.91	0.9

mean R^2 0.92 ✓ — 0.895

	Train average	Val average
poly(1)	–	–
(2)	– –	–
(3)	–	

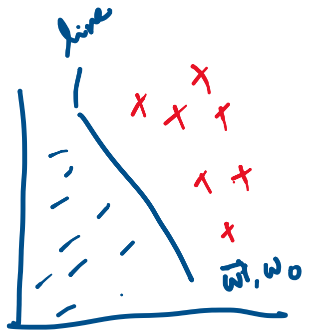
Logistic Regression ✓

→ classification ✓

Target – Binary ✓, multiclass
$\{0, 1\}$ $\{1, 2, 3\}$

Binary Classification ✓

$$y \begin{cases} 0 \\ 1 \end{cases}$$



Linear Regression ?

$y \in R^1$
$-\infty$ to ∞

$$\boxed{\hat{y} = w_0 + w_1 x_1 + \cdots w_d x_d}$$

Loss

$$\boxed{\frac{1}{m} \sum_{i=1}^{m} (y^i - \hat{y}^i)^2} \ ?$$

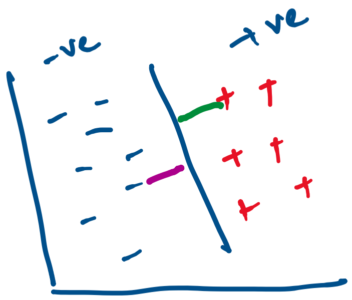
ML

Training

$$\left[X : x \in R^d, \quad y \begin{cases} 0 \\ 1 \end{cases} \right] \rightarrow \boxed{\text{mapping fn} \ ?} \rightarrow \text{Loss fn} \ ?$$

→ optimization (GD) ✓

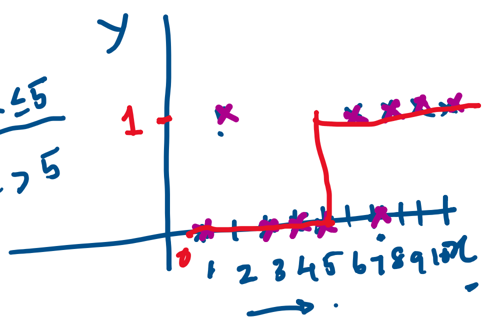
Logistic Regression : –



$$\frac{W^T \cdot x + w_0}{\|W\|} ✓$$

One variable

map $\hat{y} = f(x) = 0$ if $x \le 5$
$= 1$ if $x > 5$



Years of service (~~Account length~~) ✓	Churn
	0
1 ✓	1 ×
2 ✓	0
3 ✓	0
4 ✓	0
5 ✓	1
6	1
7	1
8	

Step function

Problems :- 1. Hard boundary

2. Not differentiable

e^{999999} - large value

e^{-x} $\quad x = \infty \quad \frac{1}{e^{\infty}}$

Sigmoid function ✓

Logistic function

$$y = \frac{1}{1 + e^{-x}} \qquad y = 1$$

$x = 0 \quad y = 0.5$

$x = -\infty \quad y \sim 0$

$0.0000000.1$

x

$(-\infty, 0, \infty) \; - \; y \sim (0, 1)$

$$f(x) = \frac{1}{1 + e^{-x}} \qquad f(x) \; (0, 1)$$

x

$(-\infty, \infty)$

$x \to \infty \quad \frac{1}{1 + e^{-1000}} = \frac{1}{1 + \frac{1}{e^{1000}}} = \frac{1}{1} = 1$

~ 0

$x \to 0 \quad \frac{1}{1 + e^{-0}} = \frac{1}{1+1} = 0.5$

$x \to \infty \quad \frac{1}{1 + e^{-(-1000)}} = \frac{1}{1 + e^{1000}} = \frac{1}{9999999} \sim 0$

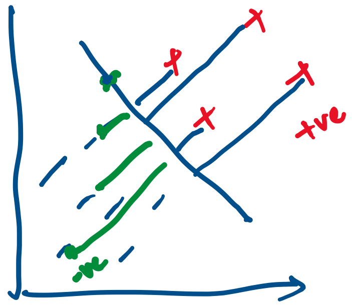
x

$(-\infty, \infty)$ Sigmoid fcn $\sim$ output $(0, 1)$

Treat Probability ?

Sigmoid $\Rightarrow$ $\boxed{f(x) = \dfrac{1}{1+e^{-x}}}$

$$\frac{W^T . x + w_0}{\|W\|} \checkmark$$



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

$$f(x) = \frac{1}{1+e^{-(W^T . x + w_0)}}$$

$$\hat{y} \text{ (Prob)} = \frac{1}{1+e^{-(W^T . x + w_0)}}$$

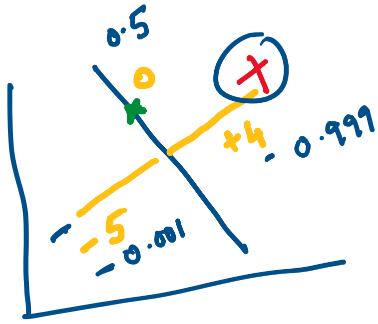
$f(x) \sim 1$

$f(x) = 0.5$

$f(x) = \sim 0$

$d = W^T . x + w_0$

$\|W\| = 1$



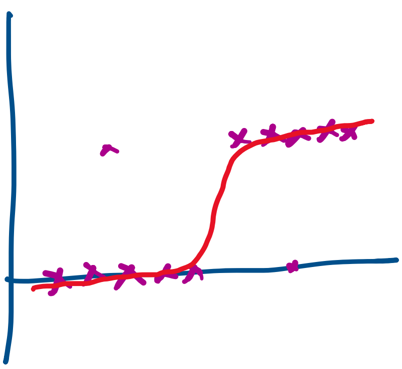
$$f(x) = \frac{1}{1+e^{-4}} = \frac{1}{1+\frac{1}{81}}$$

$$\frac{1}{1+0.001} \sim \underline{0.9999}$$

$$\frac{1}{1+e^{-(-5)}} \quad \frac{1}{1+e^{5}} \quad \frac{1}{2004} \sim 0.001$$

$$\frac{1}{1+e^{-0}} = \frac{1}{1+1} = \underline{0.5}$$

$$2.7^{-4} = \frac{1}{\underline{2.7}^4} = \frac{1}{3^4} = \frac{1}{81}$$



Sigmoid function

$$\boxed{z = w^T \cdot x + w_0}$$

$$p(y=1 \mid \underline{x}) = \frac{1}{1+e^{-(z)}}$$

σ - Sigmoid.

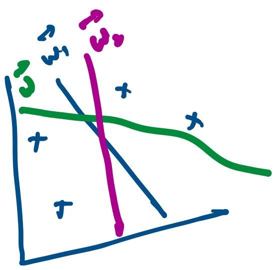
$$p = p(y=1 \mid x^i) = \sigma(z)$$

$$1-p = p(y=0 \mid x^i) = \underline{1 - \sigma(z)}$$

10:14

Loss fn

$$P(y^{(i)}=1 \mid x^{(i)}) = \begin{cases} P & \text{if } y^{(i)} == 1 \\ 1-P & \text{if } y^{(i)} == 0 \end{cases}$$



y^i	P $p(y=1)$	1-P $p(y=0)$
1 ✓	0.9 ✓	0.1
0	0.1	0.9
1	0.8	0.2
0	0.2	0.8

likelihood.

$$p^i y^i + (1-p^i)(1-y^i)$$

Bernouli

p.d.f

y	$\vec{w}_1$ p $\hat{y}$		$\vec{w}_2$ $\hat{p}$ $\hat{y}$	$\vec{w}_3$ p $\hat{y}$
1	0.9 ✓	0.9	0.8 ✓	0.7 ✓
0	0.1 ✓	0.9	0.2	0.3
1	0.9	0.9	0.8	0.7
0	0.1	0.9	0.2	0.3
✓	✓ 2		2	2

$$= \frac{p^{i\,y^i}(1-p^i)^{1-y^i}}{(1-p^i)^{1-y^i}}$$

$y^i = 0$ $\quad$ $p^{i\,0}$ $\quad$ $1-y^i=0$

$y^i = 0$

$y^i = 1$ $\quad$ $p^{i\,y^i}$ $\quad$ $1\cdot y=0$

	y^i	p	$1-p$	$(p^i)^{y^i}\cdot(1-p^i)^{1-y^i}$			
$P(A)$ -	1	0.9	0.1	0.9×1	0.9	0.8	0.9
$P(B)$ -	0	0.1	0.9	1×0.9	0.9	0.8	0.7
	1	0.9	0.1	0.9×1	0.9	0.8	0.7
	0	0.1	0.9	1×0.9	0.9	0.8	0.7

likelihood $(p^i)^{y^i}(1-p^i)^{1-y^i}$ ✓

$p(A \text{ and } B) = p(A)\cdot p(B)$

$$\prod_{i=1}^{m}(p^i)^{y^i}\cdot(1-p^i)^{1-y^i}$$

likelihood large ✓

Σ - addition $\quad$ $\prod_{i=1}^{m}$ → multiplication

Maximum likelihood

$$\arg\max_{\vec{w}} \prod_{i=1}^{m}(p^i)^{y^i}(1-p^i)^{1-y^i} \checkmark$$

$\therefore\ p^i = \sigma(x)$

$$p^i = \frac{1}{1+e^{-(w^T\cdot x_{\dots})}}$$

→ $\boxed{\text{Sigmod + MSE}}$ (map, loss) → $\boxed{\text{Non - Convex}}$ -

Maximization → minimization

Max. likelihood $\prod_i^m (p^i)^{y^i} (1-p^i)^{1-y^i}$

min $\boxed{-\prod_i^m (p^i)^{y^i} (1-p^i)^{1-y^i}}$

log likelihood

$$\log\left((p^i)^{y^i} \cdot (1-p^i)^{1-y^i}\right)$$

$\log(a \cdot b) \rightarrow \log a + \log b$

$\log a^2$ $\qquad \log (p^i)^{y^i} + \log\left((1-p^i)^{1-y^i}\right)$

log likihood $\quad y^i \log(p^i) + (1-y^i) \log(1-p^i)$

$$\boxed{-\text{log likihood} \qquad -\prod_i^m y^i \log(p^i) + (1-y^i)\log(1-p^i)}$$

$\boxed{-\text{ve Log Loss}}$ ✓ $\qquad \boxed{\text{Negative log loss}}$ ✓

$y=0 \leftarrow \boxed{\hat{y}=0.01}$ log loss

$\boxed{\text{Data}} \longrightarrow \boxed{\text{mapping}} \longrightarrow$ loss fn $\longrightarrow$ optimization

$\boxed{-\text{log loss}}$

$(X : x \in R^d . y \in {}^0_1)$ $\sigma(w^T.x+w_0)$

$$\arg\min_{\vec{w}} -\prod_{i=1}^{m} y_i \log(p^i) + (1-y^i)\log(1-p^i)$$

$$p^i = \frac{1}{1+e^{-(w^T.x+w_0)}}$$

$$w^1 = w^0 - \eta \cdot \frac{\partial L}{\partial w_d}$$

$$\boxed{\frac{\partial L}{\partial w_d}}$$

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

$$\boxed{z = w^T.x+w_0}$$

$d(\hat{y})$ $\sigma'(z) = \frac{\partial \sigma(z)}{\partial z}$

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

$$\frac{df(x)}{\partial x} = \frac{1}{1+e^{-x}}\left(1 - \frac{1}{1+e^{-x}}\right) = f(x)(1-f(x))$$

$$= \sigma(z)(1-\sigma(z))$$

derivative sigmoid) $\boxed{\sigma'(x) = \sigma(x)\left(1-\sigma(x)\right)}$

$f(g(h(x)))$
$f(g(h(w)))$

$$L = -\left[\underbrace{\overbrace{y^i \cdot \log(p^i)}^{A} + \overbrace{(1-y^i)\log(1-p^i)}^{B}}\right]$$

$z = w^T \cdot x + w_0$

$p^i = \sigma(z)$

$w_1 x_1 + w_2 x_2 + \ldots$

$$\frac{\partial L_A}{\partial w} = \frac{\partial A}{\partial p^i} \cdot \frac{\partial p^i}{\partial z} \cdot \frac{\partial z}{\partial w}$$

$$= -y^i \cdot \frac{1}{\cancel{p^i}} \cancel{p^i}(1-p^i) \cdot x_d$$

$$= -y^i (1-p^i) \cdot x_d$$

$$\frac{\partial L_B}{\partial w} = \frac{\partial B}{\partial (1-p^i)} \cdot \frac{\partial (1-p^i)}{\partial p^i} \cdot \frac{\partial p^i}{\partial z} \cdot \frac{\partial z}{\partial w_d}$$

$$= \frac{1-y}{\cancel{1-p^i}} \cdot (-1)\; p^i \left(\cancel{1-p^i}\right) \cdot x_d$$

$$= (-1)(1-y) \cdot p^i \cdot x_d.$$

$$= -(1-y) \cdot p^i \cdot x_d.$$

$$\partial L_A + \partial L_B = y(1-p^i)x_d - p^i(1-y) \cdot x_d$$

$$= (p^i - y)\, x_d.$$

$p^i = \hat{y}$

$$\frac{\partial L}{\partial w_d} = (\hat{y}^i - y^i) \cdot x_d \qquad (y - \hat{y}) \cdot x_d$$

derivative - ~ linear Regression

loss fn MSE

$$\boxed{\frac{\partial L}{\partial w_d} = (\hat{y}^i - y) \cdot x_d}$$

— log loss

$\frac{\partial L}{\partial w_d}$

Optimization G.D

η - learning rate

$$w^1 = w^0 - \eta \cdot \underbrace{(\hat{y} - y) \cdot x_d}$$

$$MSE = \frac{\partial L}{\partial w_d} \quad (y - \hat{y}) \cdot x_d$$

$$\boxed{\text{Logloss} - \frac{\partial L}{\partial w_d} \quad (\hat{y}^i - y^i)\, x_d}$$