Learning

Supervised

unsupervised

cont

cat

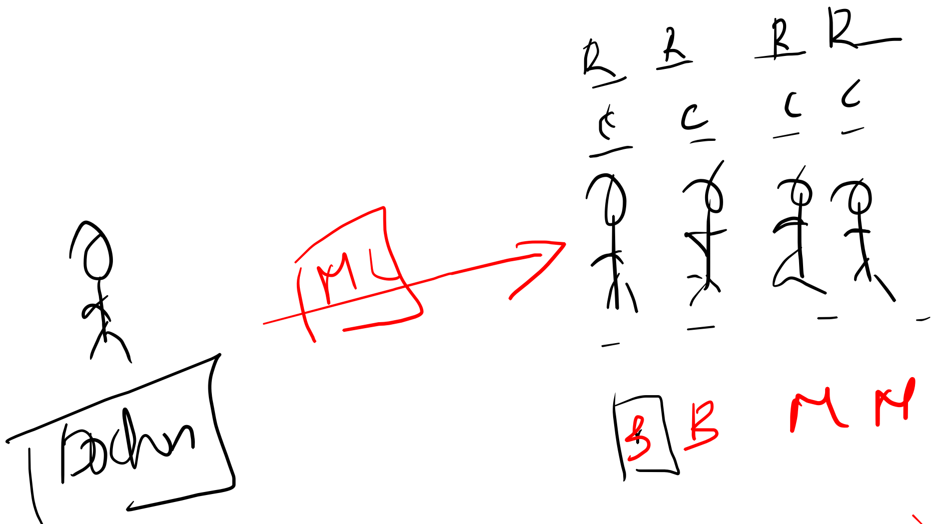
classification

✓ regression

classification

$x_1 \rightarrow \boxed{ML} \rightarrow y$ [Yes / No]

① i/p x_1, x_2, x_3 . . . → ML → cat / classify



classify the patient into b cat

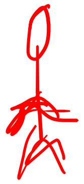
C < M →, B →

i/p, i/p, i/p → ML → o/p, o/p, o/p

↳ f(x) < B, M

i/p → $\boxed{ML}$ → predict < B, M

↳ classification



online shop

BM

cont ~~userid~~

cat ~~Gender~~

cont ✓ Age

cont ✓ Estimated Salary

X

chisquare

t-test

t-test

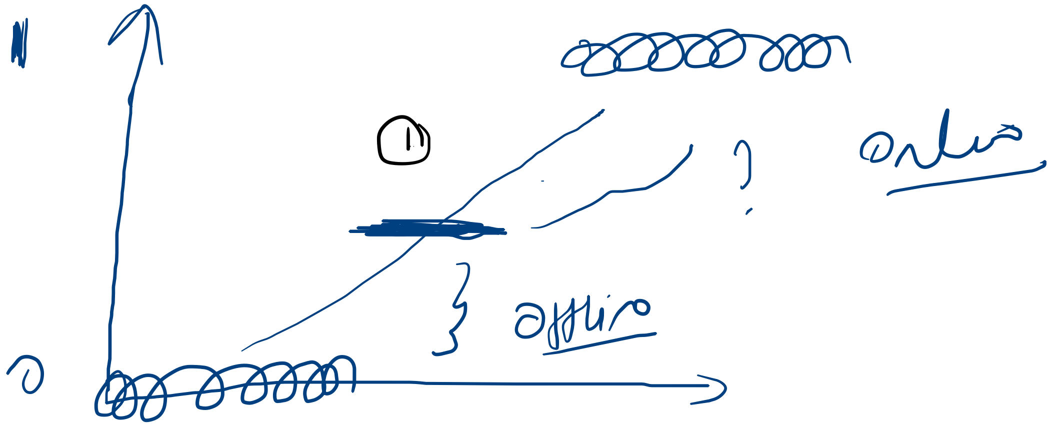
purchased [cat] $\begin{cases} 0 \\ 1 \end{cases}$

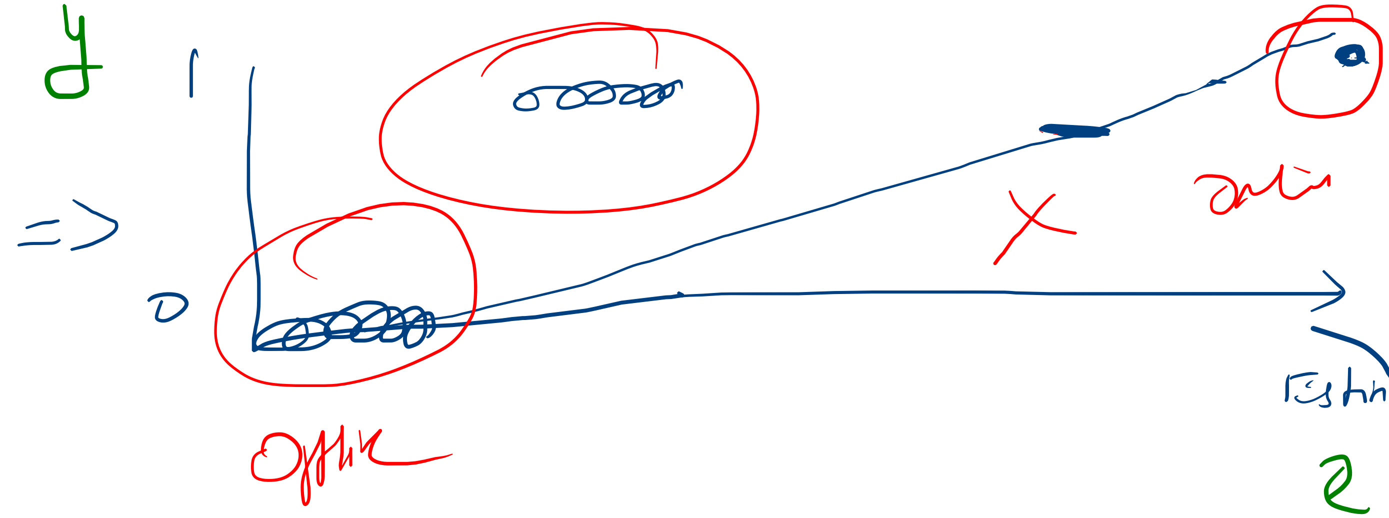
chisquare

↳ two freq table

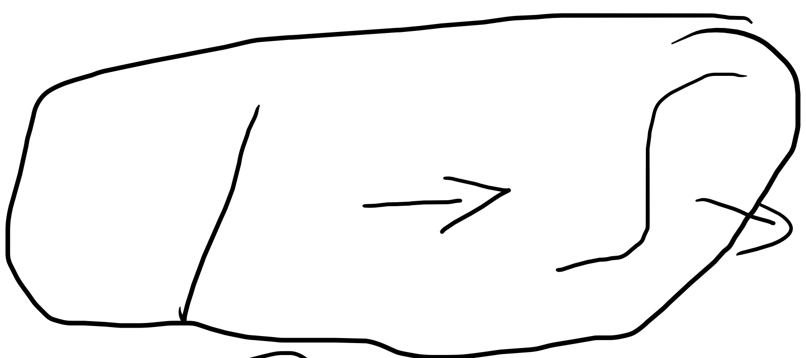
↳ chi squ contig

↳ p-value





predict
↳ 0.5 > online
< offline



S-shaped

logistic / logit

$$y = \theta_0 + \theta_1 x$$

$$g(x) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x)}}$$

↳ lin

Logistic regression

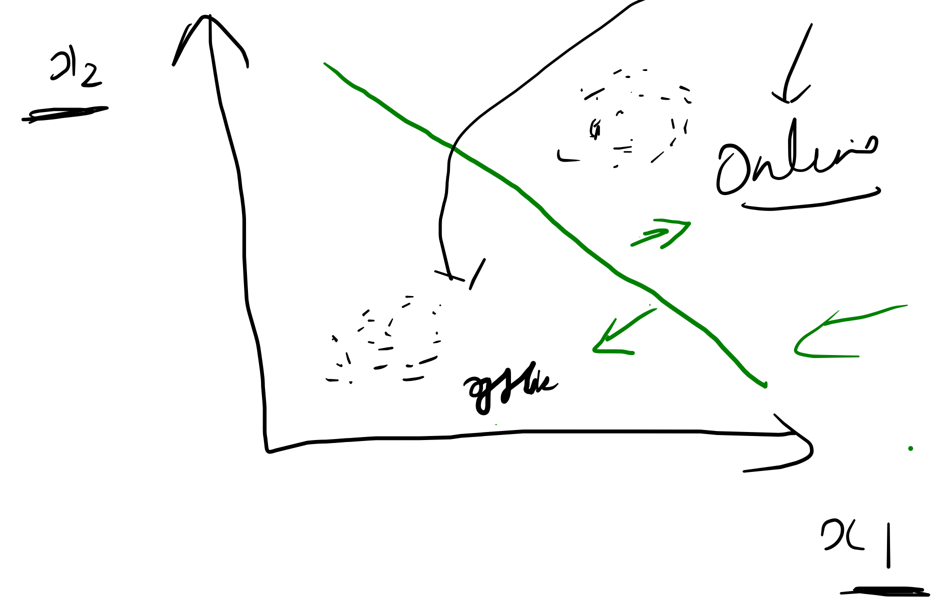
MLE

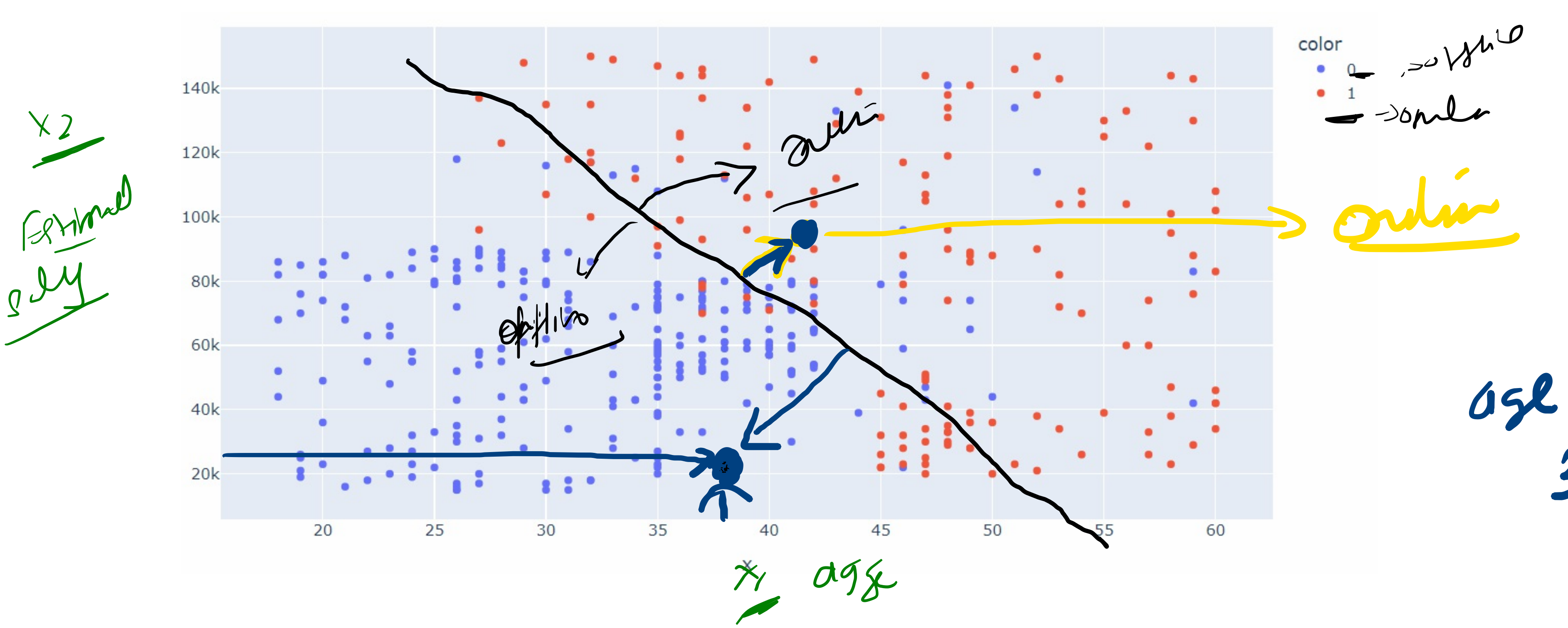
Maximum likelihood estimation

1. gradient descent
2. SGD
3. BFGS

⋮

decision boundary





age 37 | Esalary 25k → 0 (no) ?

predicted = 0

odd ratio

$$\frac{P}{q} = \frac{P}{1-P}$$

$P \rightarrow$ online purch

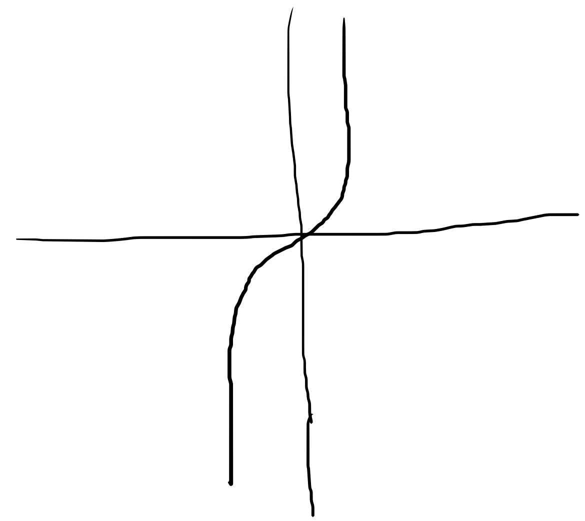
$q \rightarrow$ offline purcha

$\frac{60}{40}$

Suc

$\frac{0.60}{0.40}$

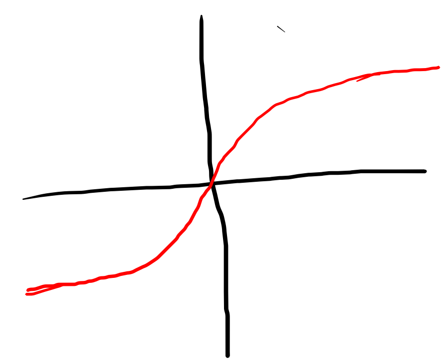
fail



$y = \theta_0 + \theta_1 x$

$$\ln \frac{P}{q} = \ln\left(\frac{P}{1-P}\right)$$

ln



$$\boxed{\ln\left(\frac{P}{1-P}\right) = y = \theta_0 + \theta_1 x}$$

$$\boxed{\ln\left(\frac{p}{1-p}\right) = y} = \underline{\theta_0 + \theta_1 x}$$

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

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

$$\underline{p} = e^{y} - \underline{e^{y}p}$$

$$p + pe^{y} = e^{y}$$

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

$$p = \frac{e^{y}}{1+e^{y}} \quad \begin{matrix}\div e^{y}\\ \div e^{y}\end{matrix}$$

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

$$p = \frac{1}{e^{-y} + 1} = \frac{1}{1+e^{-y}} = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}$$

$$\boxed{p = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}}$$

$$P = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}$$

→ decision boundary

?

MLE

→ [θ] = find coefficient

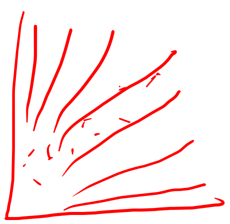
$\boxed{\text{gradient descent}}$

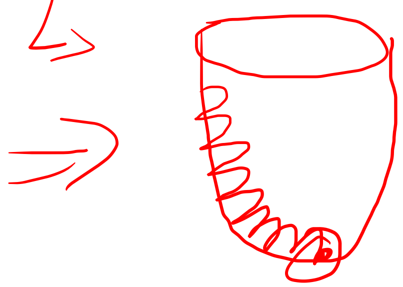
→ coeff

← loss function / cost function

Logistic regression

Linear regression





$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x) - y\right)^2$$

Logistic regression

↳ cost function / loss function

$$h_\theta(x) = \frac{1}{1+e^{-y}} = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}$$

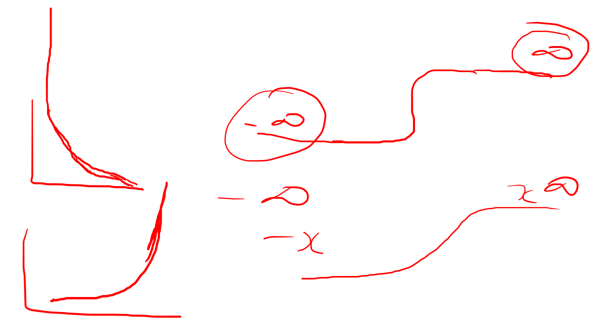
GD : $\{$

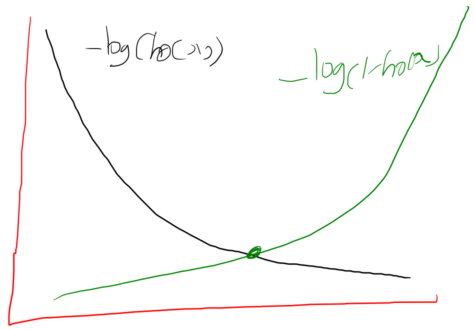
$$\theta_{new} = \theta_{old} - \alpha \frac{\partial}{\partial \theta} \left(J(\theta)\right)$$

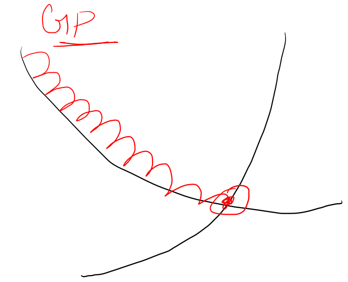
$\}$

Cost function logistic regression

$$\text{cost} = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1-h_\theta(x)) & \text{if } y = 0 \end{cases}$$







$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \text{cost}(h_\theta(x), y)$$

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left[-y \log(h_\theta(x)) - (1-y) \log(1-h_\theta(x)) \right]$$

y = 0

y = 1

0%

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \left[y \log(h_\theta(x)) + (1-y) \log(1-h_\theta(x)) \right]$$

coefficient → log

GP

actual | pred

cat y_test	y_pred cat	
0	1	✓
1	1	
1	0	✓
0	0	
0	1	
1	1	
0	1	✓
1	0	✓
1	1	
1	1	

✓ → misclassification error

→ **confusion matrix**

pred

actual	1	0
1	TP a 4 ✓	FN b 2 ✗
0	FP c 3 ✗	TN d ✓

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} = \frac{TP+TN}{TP+TN+FP+FN}$$

predict

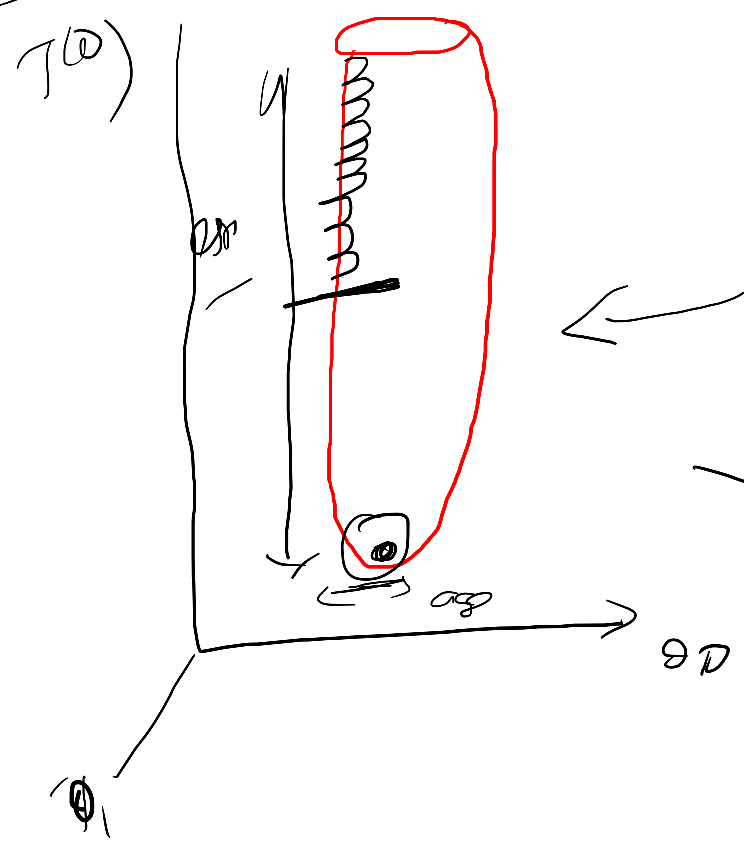
actual

	0	1
0	52	0
1	28	0

Same problem

model

max+r·100



estimated salary → Lac 0 - 1000000

age = 0-100

Standardization / z normalization

→ -3 to +3

Continuous

$$z = \frac{x - \mu}{\sigma}$$ ⇒