Data Science and Artificial Intelligence

Machine Learning

Classification

Lecture No. 3

By- SIDDHARTH SABHARWAL SIR

Recap of Previous Lecture

Topic

Topic

Topic

Topic

Topic

Logistic Regression

S curve fit on data $\Rightarrow$ but the decision boundary, classifier is linear.

$$\frac{1}{1+e^{-x_i\beta}}, \quad x_i\beta = (\beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + ---)$$

Topics to be Covered

Topic

Topic

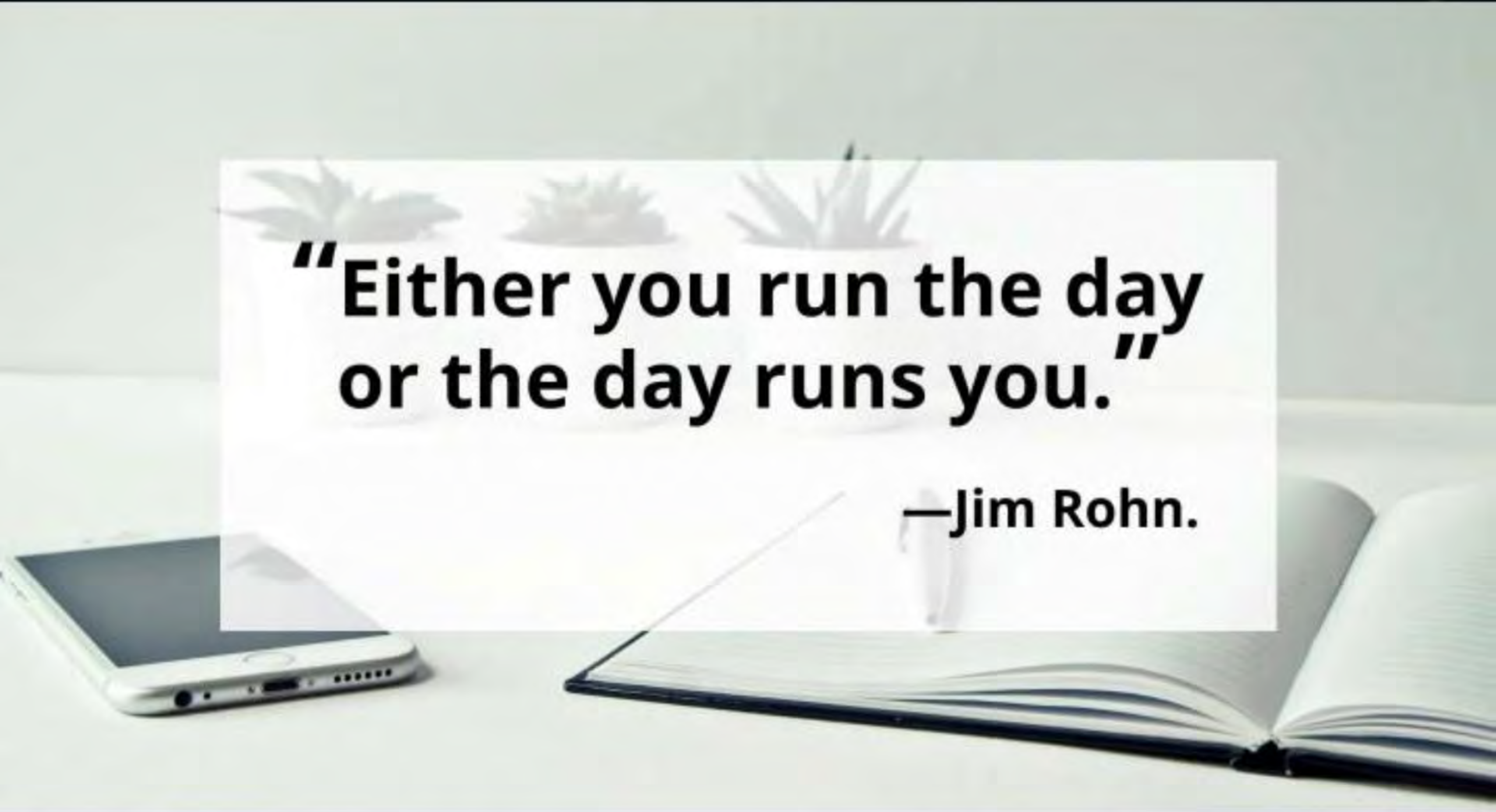
Topic

Topic

Topic

Logistic regression : Cost function,

- How the outlier effect is reduced
- Logit
- Question, H.W.


"Either you run the day
or the day runs you."
—Jim Rohn.

GATE 2026 FODENGE

About the Faculty

- AIR 1 GATE 2021, 2023 (ECE).
- AIR 3 ESE 2015 ECE.
- M.Tech from IIT Delhi in VLSI.
- Published 2 papers in field of AI-ML.
- Paper 1 : Feature Selection through Minimization of the VC dimension.
- Paper 2 : Learning a hyperplane regressor through a tight bound on the VC dimension.

By- SIDDHARTH SABHARWAL SIR

By- SIDDHARTH SABHARWAL SIR

REVISION

Linear Classification : Complete revision

done

$\boxed{\text{What is a parametric model.}}$ → KNN, DT } Non parametric

→ LR, LC, logistic Reg } Parametric model.

↓

Linear regression

⇓

- model equation Pehle se hi fix hai
- No. of Parameters fix hai
- and from data we train the model to find Parameters.

REVISION

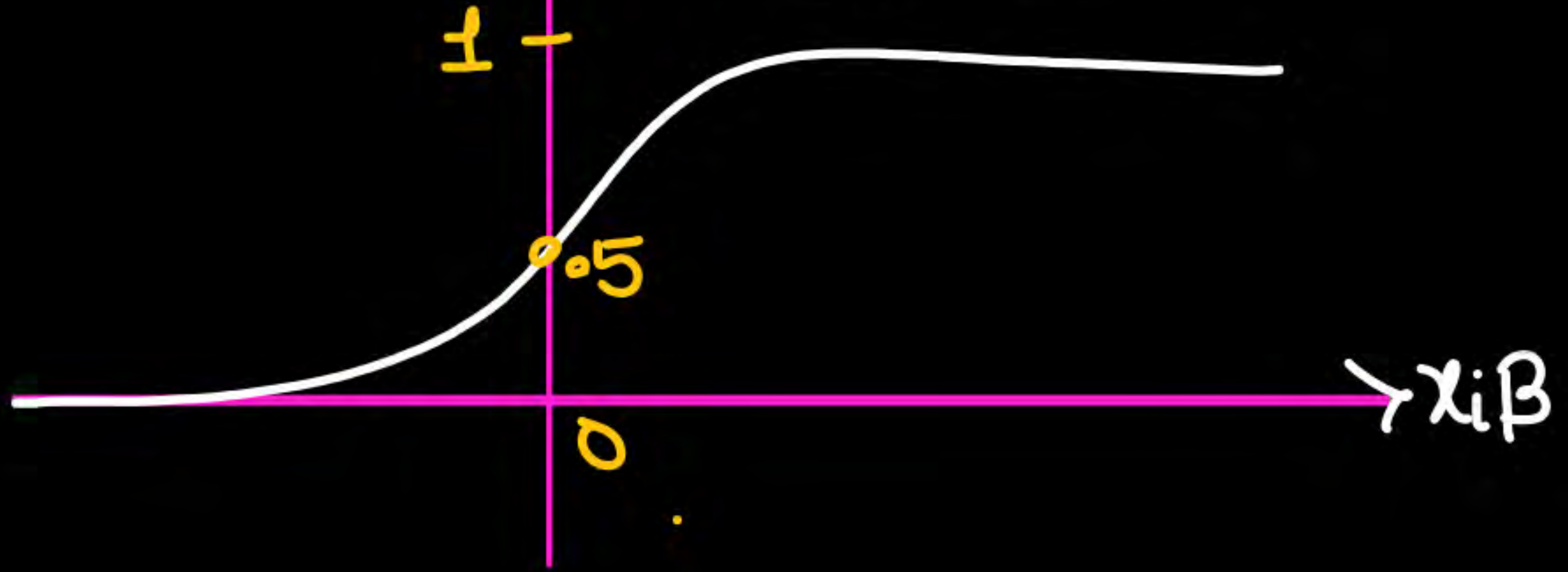
Linear Classification : The Loss function

Supewised learning

- Parametric models ⇒ The model is fixed,
  No of parameters fixed,
  we simply train model using data
  To find best parameters

Linear Classification

$$P_1 = \frac{1}{1+e^{-x_i\beta}}$$



How this solve the problem of outliers…??

Logistic Regression

Logistic Regression

Let us have a data with some classes 1 and 0, these are the Y values of the input.
In logistic Regression we actually try to fit a S curve on the data.

done

Logistic regression gives a linear classifier

but it is better than linear classification

becoz the algo is less affected by outlier.

The Sigmoid Function…

What is the value of the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$ when $x = 2$?

$\frac{1}{1+e^{-x}}$ @ $x=2$

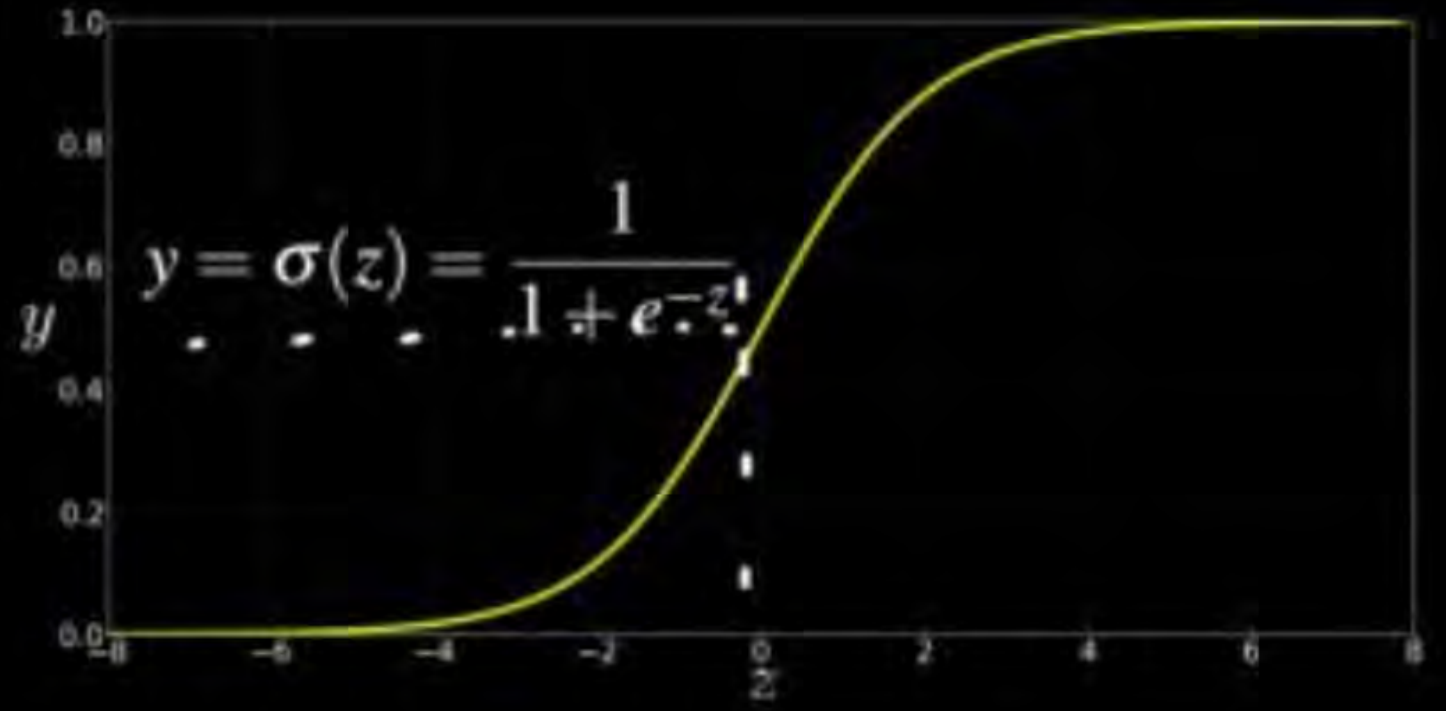
.88

(A) 0.119

(B) 0.268

(C) 0.5

(D) $\boxed{0.881}$ d



Sigmoid fxn.

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

Show sigmoid fxn.

σ

7.2 (5pt) Suppose that you have trained a logistic regression classifier $h_\theta(x) = \sigma(1-x)$ where $\sigma(\cdot)$ is the logistic/sigmoid function. What does its output on a new example $x = 2$ mean? Check all that apply. (*Hint:* $\sigma(-1) \approx 0.27$)

- ☐ Your estimate for $P(y = 1|x;\theta)$ is about 0.73.
- ☐ Your estimate for $P(y = 0|x;\theta)$ is about 0.27.
- ☑ Your estimate for $P(y = 1|x;\theta)$ is about 0.27.
- ☑ Your estimate for $P(y = 0|x;\theta)$ is about 0.73.

$\sigma(1-2) = \sigma(-1) =$

$P_1 = \sigma(-1) = \frac{1}{1+e^{-(-1)}} = \frac{1}{1+e^{1}} = .27$

$P_0 = 1 - P_1 = .73.$

$P(y=1/x) = .27$

$P(y=0/x) = .73.$

Logic regression model.

<u>Ex</u> $\beta = [\beta_0 = 1, \beta_1 = .5, \beta_2 = .8]$

data $x = [x_1 = 5, x_2 = 3]$

$$\boxed{\text{find } \hat{P}(y=1/x\,;\beta)} \Rightarrow \frac{1}{1+e^{-x\beta}} = \frac{1}{1+e^{-(\beta_0+\beta_1 x^1+\beta_2 x^2)}} = \frac{1}{1+e^{-(1+5\times .5+.8\times 3)}}$$

$$= .990$$

Find the Probability that for given x (5,3) the class of y is 1 using Parameters β

In logistic regression

Generally @ Testing $\Rightarrow$

$P(y=1/x) > .5 \longrightarrow 1$

$P(y=1|x) < .5 \longrightarrow 0$

But depending on data we add concept of threshold in logistic Reg:— If threshold is θ

then $P(Y=1/x) \geq \theta \Rightarrow 1$

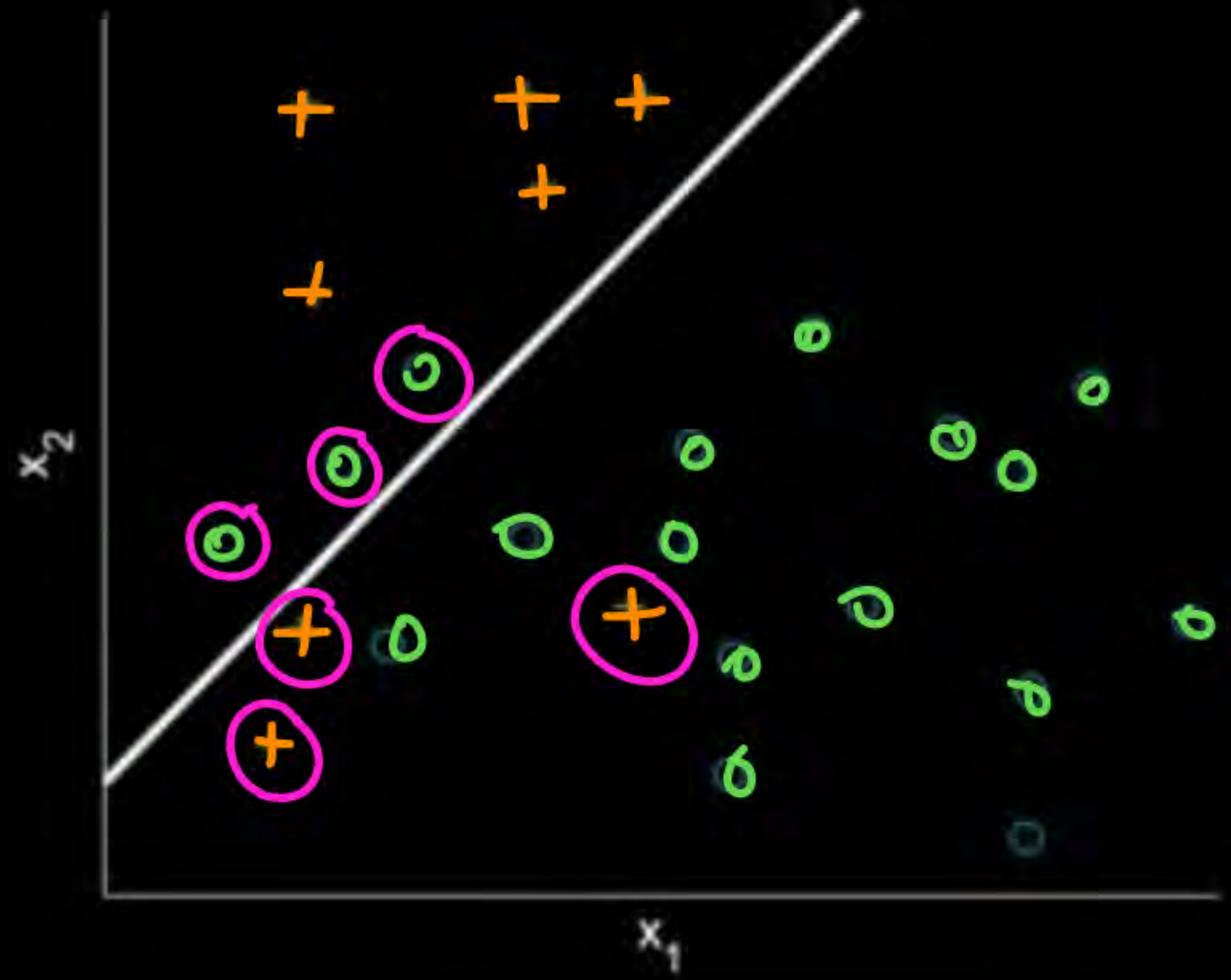
$P(Y=1/x) < \theta \Rightarrow 0.$

Consider the dataset shown in the diagram below with a logistic regression decision boundary.

Questions:

1. Count the number of mistakes the model is making. 6
2. Is the data linearly separable? → $\boxed{\text{No}}$

data not linearly Separable



Suppose we trained a logistic regression classifier for some binary classification task. The true labels y and predicted probabilities $P(Y=1|x)$ are given below.

$P(y=1/x)$

actualy

$P(Y=1/x)$

y	$P(Y=1\|x)$
1	0.9
1	0.6
0	0.6
0	0.55
1	0.4
0	0.3

1
1
1
1
0
0

(3error)

Threshold $\cdot 5$

- if $P(Y=1/x) \geq .5 \Rightarrow$ then we say data belong to Class 1
- if $P(Y=1|x) < .5$ data Class 0

How many mistakes model is doing considering threshold to be 0.5 ?

Suppose we trained a logistic regression classifier for some binary classification task. The true labels y and predicted probabilities $P(Y = 1|x)$ are given below.

y	$P(Y = 1\|x)$
1	0.9
1	0.6
0	0.6
0	0.55
1	0.4
0	0.3

H.W.

In order to make predictions, we need to threshold $P(Y = 1|x)$. We use our thresholds as follows:

$$\hat{Y} = \begin{cases} 1 & P(Y = 1|x) \geq T \\ 0 & P(Y = 1|x) < T \end{cases}$$

We want to decide between three possible thresholds for $P(Y = 1|x)$.

- Model A uses a threshold of $T = 0.25$ .25
- Model B uses a threshold of $T = 0.5$.   .5
- Model C uses a threshold of $T = 0.75$ .75

How many mistakes each model is making ?

Suppose we trained a logistic regression classifier for some binary classification task. The true labels y and predicted probabilities $P(Y = 1|x)$ are given below.

$P(y=1/x)$

actualy

$P(Y=1/x)$

y	$P(Y = 1\|x)$	
1 ✓	0.9	1
1 ✓	0.6	1
0 ×	0.6	1
0 ×	0.55	1
1 ✓	0.4	1
0 ×	0.3	1

Threshold .25

- if $P(Y=1/x) > .25$ Then we say data belong to Class 1
- if $P(Y=1/x) < .25$ data Class 0

3 mistake

How many mistakes model is doing considering threshold to be ~~0.5~~ .25 ?

Consider linear regression and logistic regression. Circle the correct answer for each statement below. If a statement is **false**, explain why in one sentence.

1. (1 pt) (True or False) $\boxed{\text{They both use linear functions.}}$

dinear fxn

dinear negression and logistic negnemion.

uses sigmoid.

$(\hat{y} = \beta_0 + \beta_1 x^1 + \ldots)$

$\hat{y} = \frac{1}{1+e^{-x_i\beta}}$

2. (1 pt) (True or False) They both can be used to solve regression problems.

No, linear negression → used for negnemion

I have numerical dat

logistic negnemion → used for clamification

No

3. (1 pt) (True or False) $\boxed{\text{They both use the logistic activation function.}}$

In which of the following situations can logistic regression be used? Select all that apply.

(a) Predicting whether an email is a spam email or not based on its contents.

Regression

(b) Predicting the rainfall depth for a given day in a certain city based on the city's historical weather data.

(c) Predicting the cost of a house based on features of the house.

Regression

(d) Predicting if a patient has a disease or not based on the patient's symptoms and medical history.

2 type of data

Categorical data

$y = 1/0$.

numerical data

$y =$ any number

x

Logistic Regression works as Regression

What is the purpose of the sigmoid function in logistic regression?

(a) It converts continuous input into categorical data.

(b) It standardizes the input to have zero mean and variance 1.

(c) It optimizes the weights to reduce loss.

(d) It transforms the output to a probability.

$$\left(\sigma(z) = \frac{1}{1+e^{-z}}\right)$$

- $P(Y=1|x) = \sigma(x_i\beta) = \frac{1}{1+e^{-x_i\beta}}$
- $P(Y=0|x) = 1 - P(Y=1|x) = 1 - \sigma(x_i\beta)$ ✓

- What is logit ⇒

in an experiement → Success / → fail

The odds of the success ⇒ $\dfrac{\text{Probability of succen}}{\text{Probability of failure}}$

Bag	10 W
	20 R
	10 B

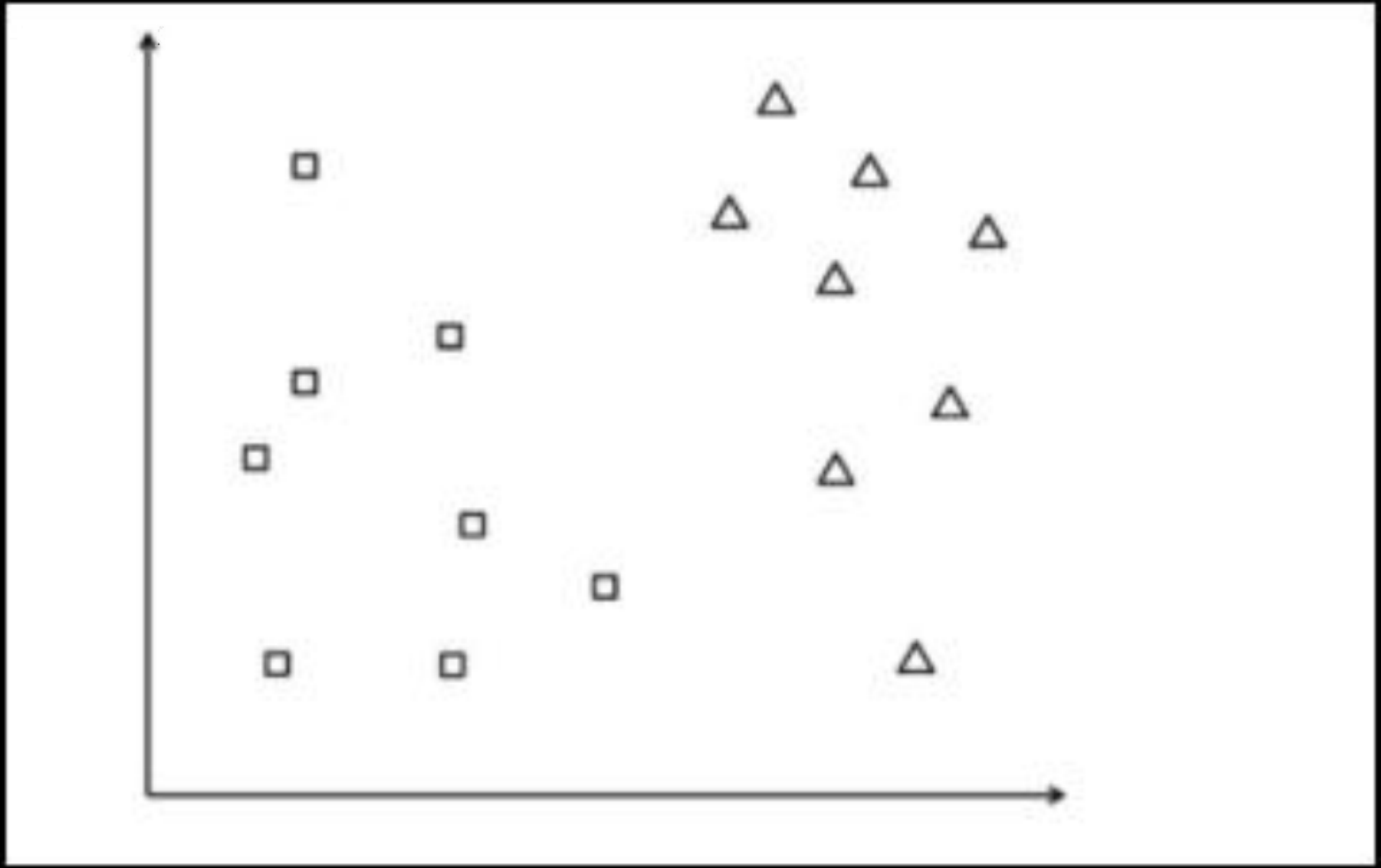
The odds of getting a Red Ball $= \dfrac{\text{Probab of getting a Red ball}}{\text{Probab of not getting Red ball}}$

$= \dfrac{20/40}{20/40} = 1$

(2 pt) Bubble the expression that describes the odds ratio $\frac{P(Y=1|X)}{P(Y=0|X)}$ of a logistic regression model.
Recall: $P(Y=0|X)+P(Y=1|X)=1$ for any X.

○ $X^T\beta$ ○ $-X^T\beta$ ○ $*\ \exp(X^T\beta)$ ○ $\sigma(X^T\beta)$ ○ None of these

Can Logistic regression and linear classification give 100% accuracy on this data.



• Logistic Regression

Now we have the concept of the threshold, how to find the best coefficients ?

• Logistic Regression

The concept of threshold

• Logistic Regression

Comparison of the linear classification and logistic Regression

In linear classification
we find a line and say
value <>0
but here we say value
<> some threshold

$\boxed{\text{Logit} \Rightarrow}$ Logit = log of odds.

In logistic regression $\log_e(\text{odd}) = \text{logit}$

- $P(Y=1|x) = \dfrac{1}{1+e^{-x_i\beta}}$
- $P(Y=0|x) = 1 - \dfrac{1}{1+e^{-x_i\beta}}$

Odds of Success =

- Odds of $P(Y=1|x) = \dfrac{P(Y=1|x)}{P(Y=0|x)}$

$\text{odds} = \dfrac{\dfrac{1}{\cancel{1+e^{-x_i\beta}}}}{\dfrac{e^{-x_i\beta}}{\cancel{1+e^{-x_i\beta}}}}$

$\text{Odds} = e^{+x_i\beta}$

$\text{logit} \Rightarrow \log_e e^{x_i\beta}$

$\Rightarrow (x_i\beta)$

- $\log_e \text{odd} = x_i\beta \Rightarrow$ classifier

$$\boxed{\log_e \frac{P(Y=1|x)}{P(Y=0|x)} = x_i\beta}$$

1D data

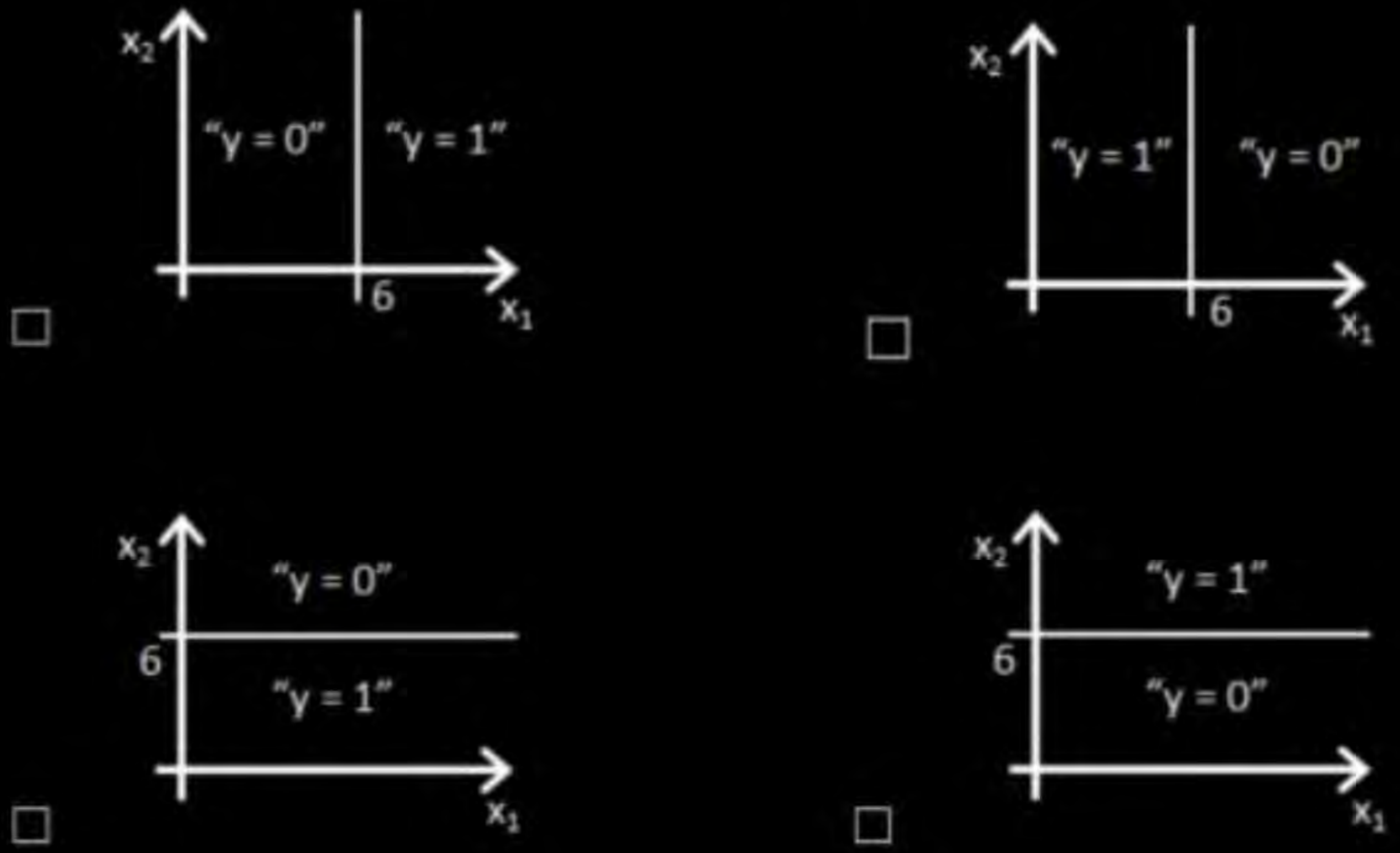
$$\log_e \frac{P(y=1|x)}{P(y=0|x)} = \beta_1 x + \beta_0$$

~~Take~~ x ki doalagalag value

$x = 2$ $\quad 2\beta_1 + \beta_0 =$ ✓

$x = 3$ $\quad 3\beta_1 + \beta_0 =$ ✓

7.1 (5pt) Suppose you train a linear classifier $h_\theta(x) = \text{sign}(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$. Suppose $\theta_0 = 6, \theta_1 = -1, \theta_2 = 0$. Which of the following figures represents the decision boundary found by your classifier?



7) Consider the data collected from 410 customers in a restaurant. It is observed that 40 of the 70 customers tipped the server who was wearing a black shirt and 130 of the 340 customers tipped the server who was wearing a different color. Compute the logit or log-odds of tipping a server wearing a black shirt.

- 0.2877
- 0.1249
- -0.7677
- -1.7677

410 Customers, 40 out of 70 Customers tipped server wearing Black shirt, 130 out of 340 Customers tipped server wearing Coloured shirt.

Find out log of odds of tipping a server with Black shirt

$$\log_e \frac{P(Y=1/Black)}{P(Y=0/Black)} = \boxed{\log_e \frac{40/70}{30/70}}$$

- let $x = 1$ if Server wear black shirt
  $x = 0$ if Server wear Coloured shirt

we Conducted the experiment and found 40 out of 70 Customers tip Black Server
" " " " " " 130 " " 340 " " Coloured Server

- $y = 1$ means Server get a tip
- $y = 0$ " " donot get a tip

find β's for logistic regression ⇒ $P_1 = \frac{1}{1+e^{-x_i\beta}}$ $P_0 = 1 - P_1$

Sol ⇒

$x \leftarrow$ Single dimension
Server ki shirt ka Color

- $\boxed{x_i\beta = \beta_0 + \beta_1 x}$

$\log_e \frac{P_1}{P_0} = x_i\beta$ ⇒ $\boxed{\log_e \frac{P(Y=1|x)}{P(Y=0|x)} = \beta_1 x + \beta_0}$

for $x = 0$

$\log_e \frac{P(Y=1|x=0)}{P(Y=0|x=0)} = \beta_0$ ⇒ $\boxed{\beta_0 = -.479}$

130/340

210/340

- let $x=1$ if Server wear black shirt
  $x=0$ if Server wear Coloured shirt

we Conducted the experiment and found 40 out of 70 Customers tip Black Server
" " " " " " 130 " " 340 " " Coloured Server

$y=1$ means Server get a tip
$y=0$ " " donot get a tip

Find β's for logistic regression ⇒

Sol ⇒

$\boxed{\begin{array}{l} x \leftarrow \text{Single dimension} \\ \text{Server ki shirt} \\ \text{Ka Color} \end{array}}$

- $x_i\beta = \beta_0 + \beta_1 x$

$\boxed{x=1}$

$$\log_e \frac{P(Y=1|x=1)}{P(Y=0|x=1)} = \beta_1 + \beta_0$$

40/70

30/70

$$\log_e \frac{4}{3} = \beta_0 + \beta_1$$

$\boxed{\beta_1 = .763}$

1D data Probab of Success

x	
3	$\xrightarrow{}$ .2
10	.9
11	.93
4	.25

Success $Y=1$

failure $Y=0$

β_1, β_0

$$\boxed{\log_e \frac{P(Y=1|x)}{P(Y=0|x)} = \beta_1 x + \beta_0}$$

@ $x=3$, $\log_e \frac{.2}{.8} = 3\beta_1 + \beta_0$

@ $x=10$, $\log_e \frac{.9}{.1} = 10\beta_1 + \beta_0$

$3\beta_1 + \beta_0 = -1.38$

$10\beta_1 + \beta_0 = 2.19$

$\beta_1 = .511$

$\beta_0 = -2.92.$

Linear Classification

y.

What type of dependent variable is suitable for logistic regression?

A) Continuous variable
B) Categorical variable with multiple categories
C) Binary or dichotomous variable $\boxed{y=1,0}$
D) Ordinal variable

In logistic regression, what is the role of the logistic function (sigmoid function)?

A) It transforms the independent variables.
B) It models the relationship between the dependent and independent variables.
C) It converts the log-odds into probabilities.
D) It calculates the likelihood of the data.

$y = P(y=1|x) = \sigma(x_i\beta)$

$\log_e(\text{odds}) = x_i\beta$ ✓

$x_i\beta$

Sigmoid Convert $x_i\beta$ into Probab

$\sigma(x_i\beta) = \text{Probab}$

Which term represents the natural logarithm of the odds of an event occurring in logistic regression?

A) Odds ratio
B) Probability
C) Log-odds or logit
D) Coefficient

Linear Classification

Kal.

What is the likelihood function used for in logistic regression?

A) To estimate the coefficients of the model.
B) To calculate the odds ratio.
C) To find the best threshold for classification.
D) To assess the fit of the model by maximizing the likelihood of the observed outcomes.

1. What kind of algorithm is logistic regression?

a) Cost function minimization

b) Ranking

c) Regression

d) Classification

6. Probability of an event occurring is 0.9. What is odds ratio?

a) 0.9:1 $\frac{.9}{.1}$

b) 9:1

c) 1:9

d) 1:0.9

#Q. The following table gives the binary labels ($y^{(i)}$) for four points $\left(x_1^{(i)}, x_2^{(i)}\right)$ where $i = 1, 2, 3, 4$. Among the given options, which set of parameter values $\beta_0, \beta_1, \beta_2$ of a standard logistic regression model $\boxed{p(x_i) = \frac{1}{1+e^{-(\beta_0+\beta_1 x+\beta_2 x)}}}$ results in the highest likelihood for this data?

• $P(x_i) > .5 \Rightarrow 1$ $\quad \Rightarrow \sum y_i \cancel{(x_i \beta)}$

• $P(x_i) < .5 \Rightarrow 0$

(a) $\beta_0 = 0.5, \beta_1 = 1.0, \beta_2 = 2.0$

(b) $\beta_0 = -0.5, \beta_1 = -1.0, \beta_2 = 2.0$

(c) $\beta_0 = 0.5, \beta_1 = 1.0, \beta_2 = -2.0$

(d) $\beta_0 = -0.5, \beta_1 = 1.0, \beta_2 = 2.0$

x_1	x_2	y
0.4	−0.2	1
0.6	−0.5	1
−0.3	0.8	0
−0.7	0.5	0

Kal

$x_i\beta$

$P(x_i) > .5 \Leftarrow x_i\beta > 0 \rightarrow$ Clan 1

$P(x_i) < .5 \Leftarrow x_i\beta < 0 \rightarrow$ Clan 0

Rule — Rule — Kal

Same Result

$\beta_0 = 1$

$\beta_2 = 2$

$\beta_1 = .5$

linear clan Clanifier

$1 + .5x^1 + 2x_2$

$x_i\beta > 0 \rightarrow 1$

$x_i\beta < 0 \rightarrow 0$

logistic

$$P_1 = \frac{1}{1 + e^{-(1 + .5x^1 + 2x_2)}}$$

$1 \leftarrow P_1 > .5 \Rightarrow x_i\beta > 0$

$0 \leftarrow P_1 < .5 \Rightarrow x_i\beta < 0$

Logistic Regression

- **The Loss function**

Logistic Regression

- **The Loss function**

How can we use log into this function

Linear Classification

Logistic Regression

- **The Loss function**

How can we use log into this function

Logistic Regression Objective Function

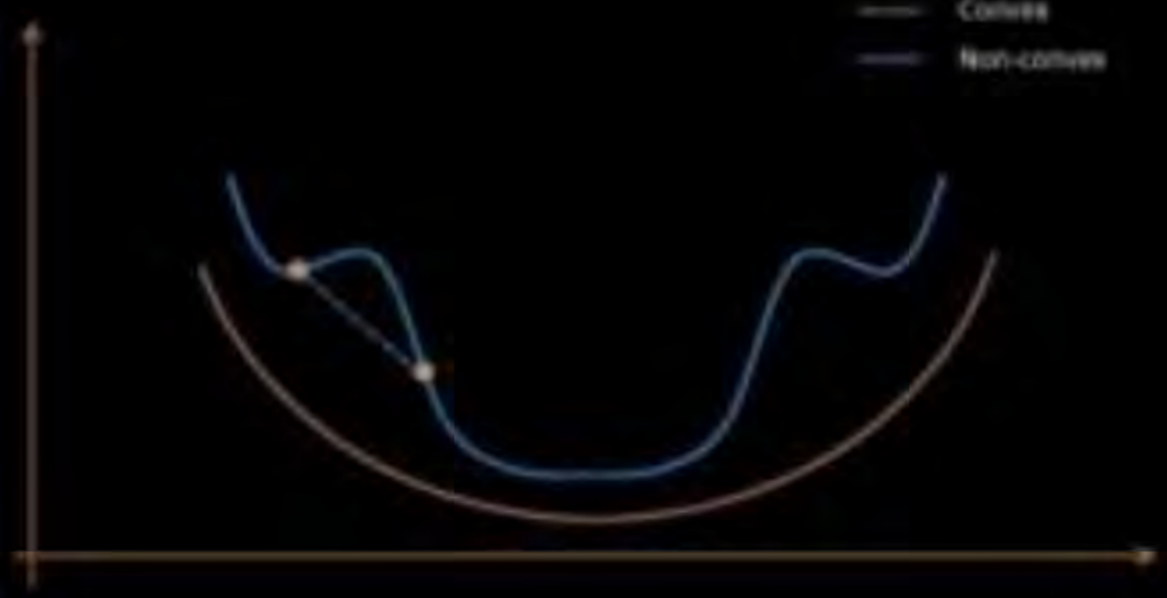
- Can't just use squared loss as in linear regression:

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

- Using the logistic regression model

$$h_{\boldsymbol{\theta}}(\boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{\theta}^\top \boldsymbol{x}}}$$

results in a non-convex optimization



i. [2 Pts] Suppose our true labels are $\vec{y} = [0, 0, 1]$, our predicted probabilities of being in class 1 are $[0.1, 0.6, 0.9]$, and our threshold is $T = 0.5$. Give the total (not average) cross-entropy loss. Do not simplify your answer.

Total CE Loss = $\boxed{}$

ii. [2 Pts] For the same values as above, give the total squared loss. Do not simplify your answer.

Squared Loss = $\boxed{}$

Suppose after training our model we get $\vec{\beta} = \begin{bmatrix} -1.2 & -0.005 & 2.5 \end{bmatrix}^T$, where -1.2 is an intercept term, -0.005 is the parameter corresponding to passenger's age, and 2.5 is the parameter corresponding to sex.

i. [3 Pts] Consider Sīlānah Iskandar Nāsīf Abī Dāghir Yazbak, a 20 year old female. What chance did she have to survive the sinking of the Titanic according to our model? Give your answer as a probability in terms of σ. If there is not enough information, write "not enough information".

H.W.

$$P(Y = 1 | \text{age} = 20, \text{female} = 1) = \boxed{}$$

ii. [3 Pts] Sīlānah Iskandar Nāsīf Abī Dāghir Yazbak actually survived. What is the cross-entropy loss for our prediction in part i? If there is not enough information, write "not enough information."

cross entropy loss = $\boxed{}$

THANK - YOU