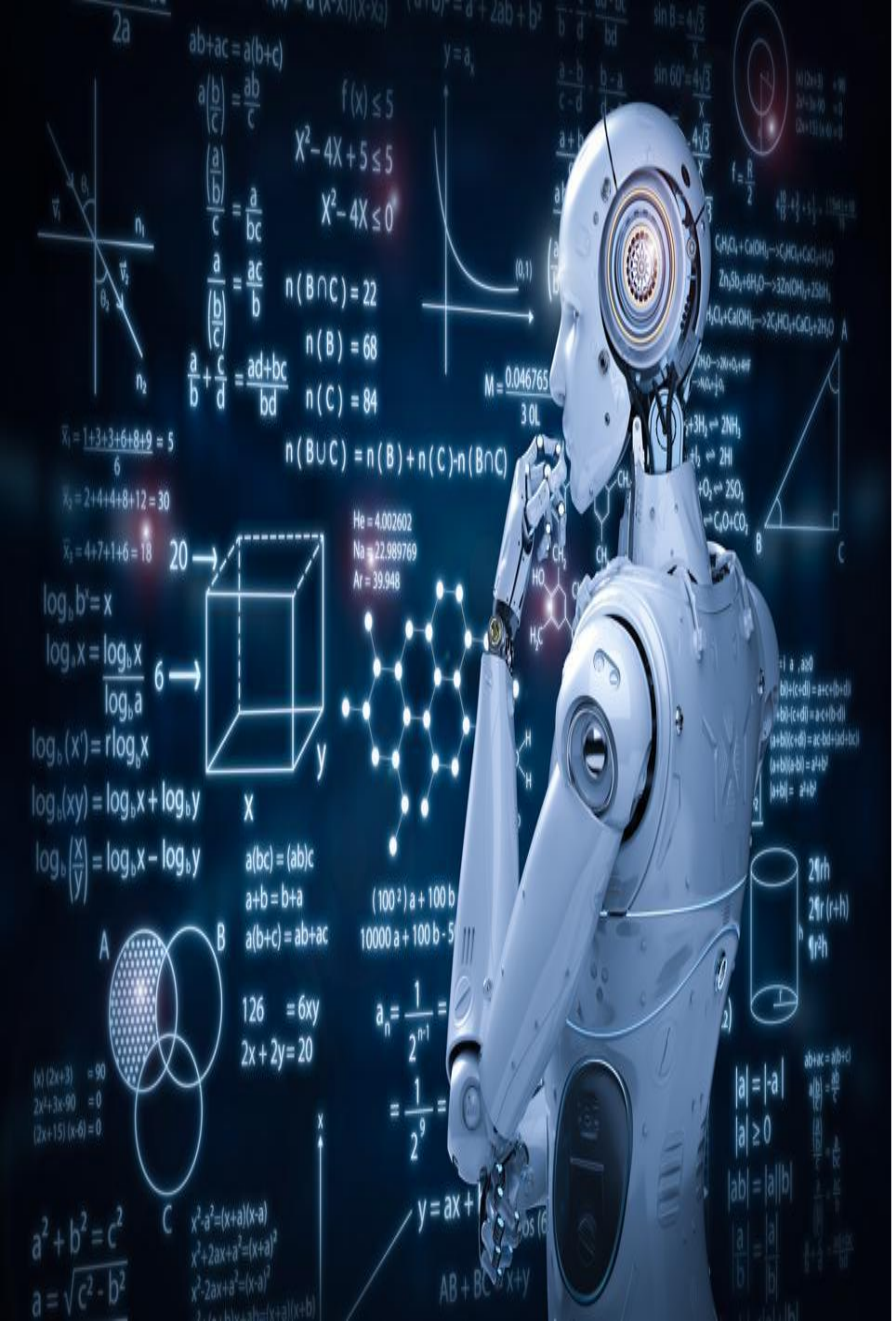


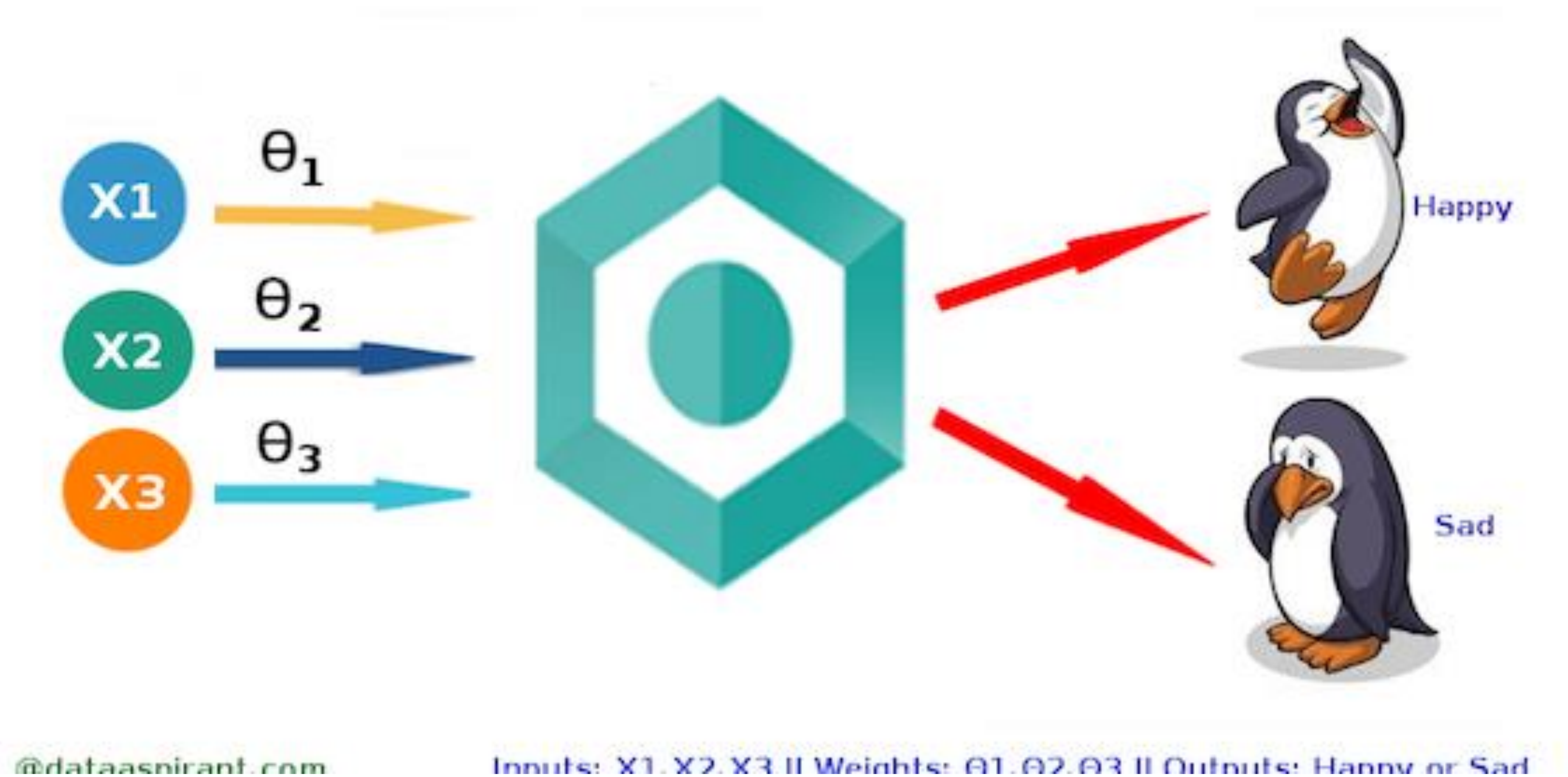
Practical Machine Learning

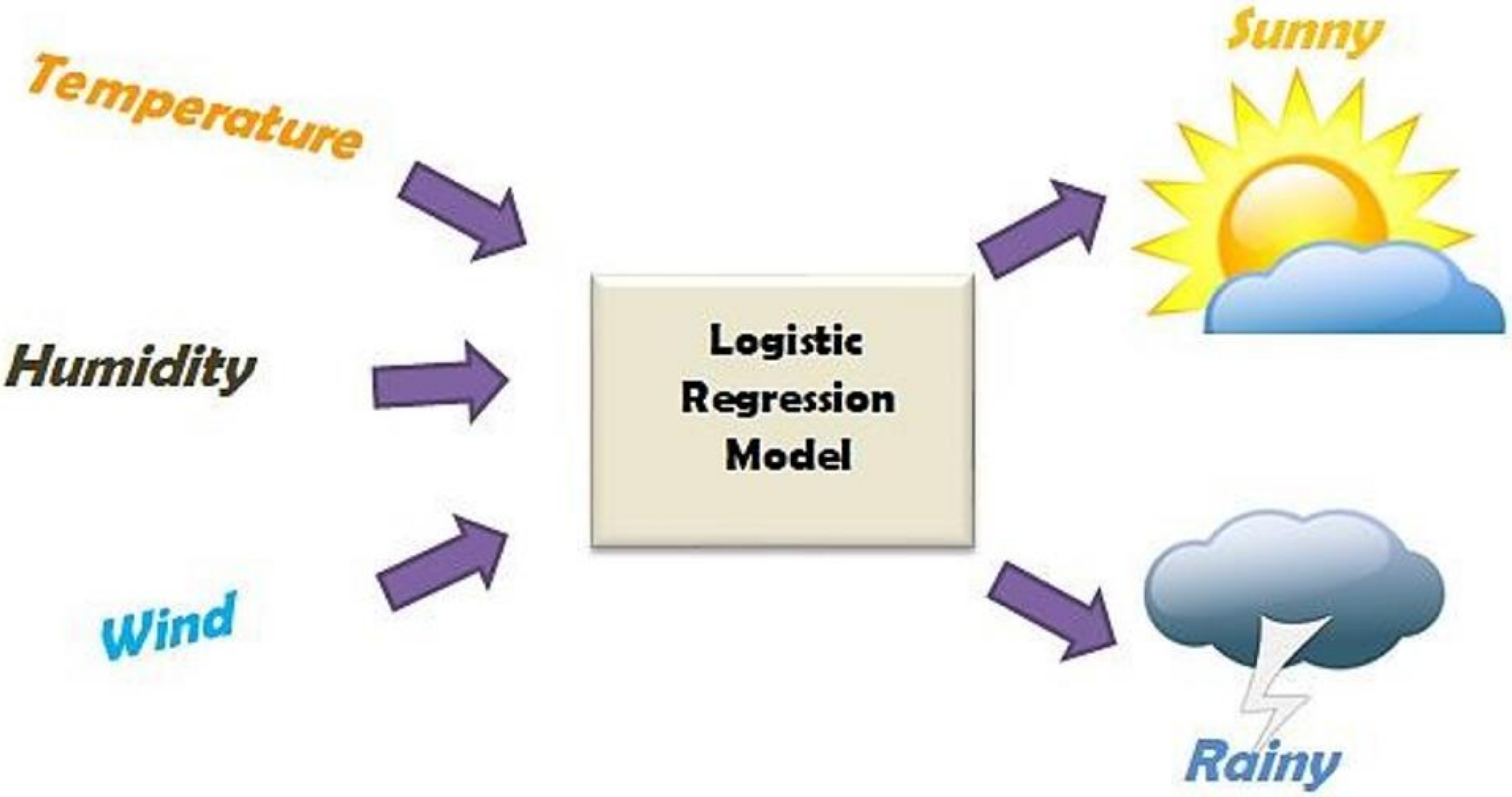
Day 6: SEP23 DBDA

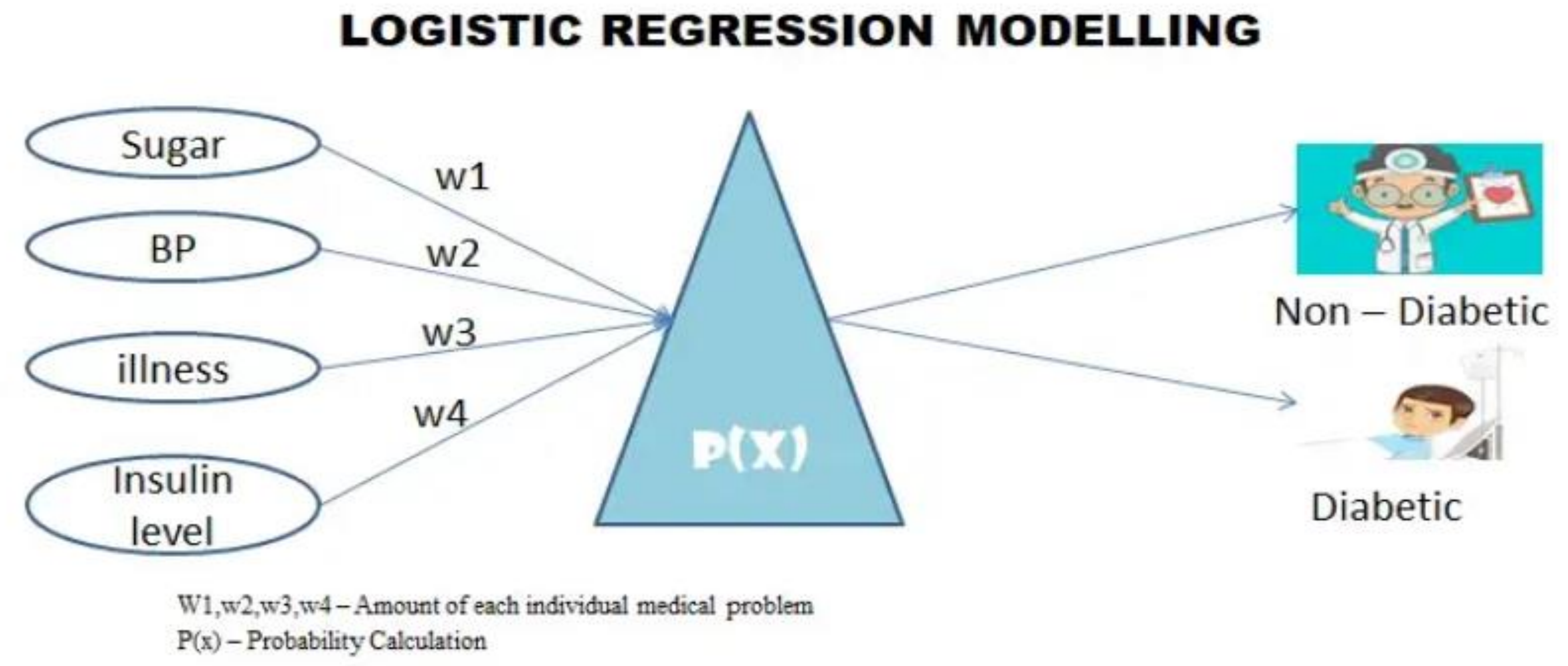
Kiran Waghmare

Agenda

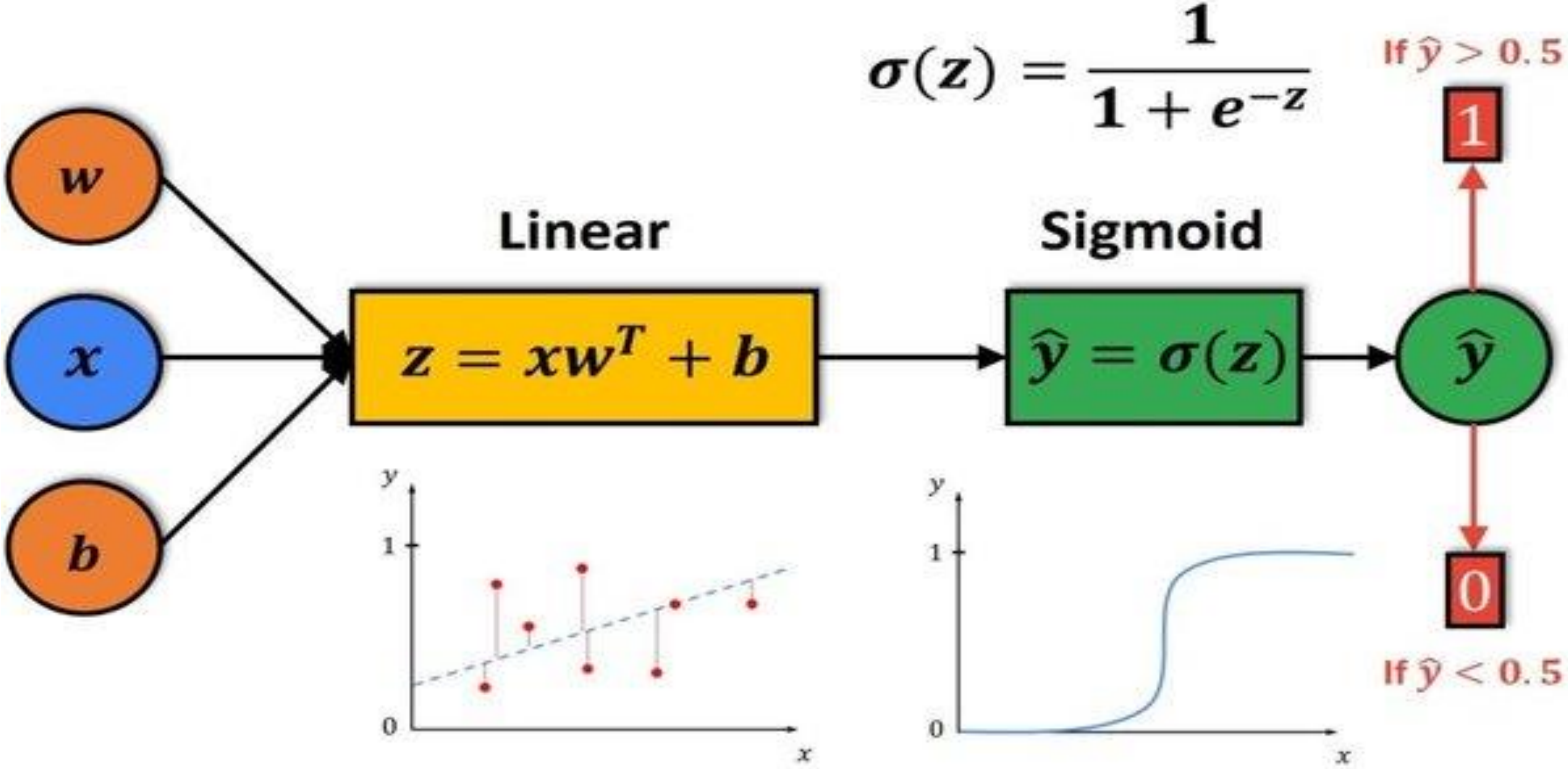
- Logistic Regression
- Classification
- Measures for classification


X1
X2
X3
θ_1
θ_2
θ_3
Happy
Sad
@dataaspirant.com
Inputs: X1,X2,X3 || Weights: Θ1,Θ2,Θ3 || Outputs: Happy or Sad

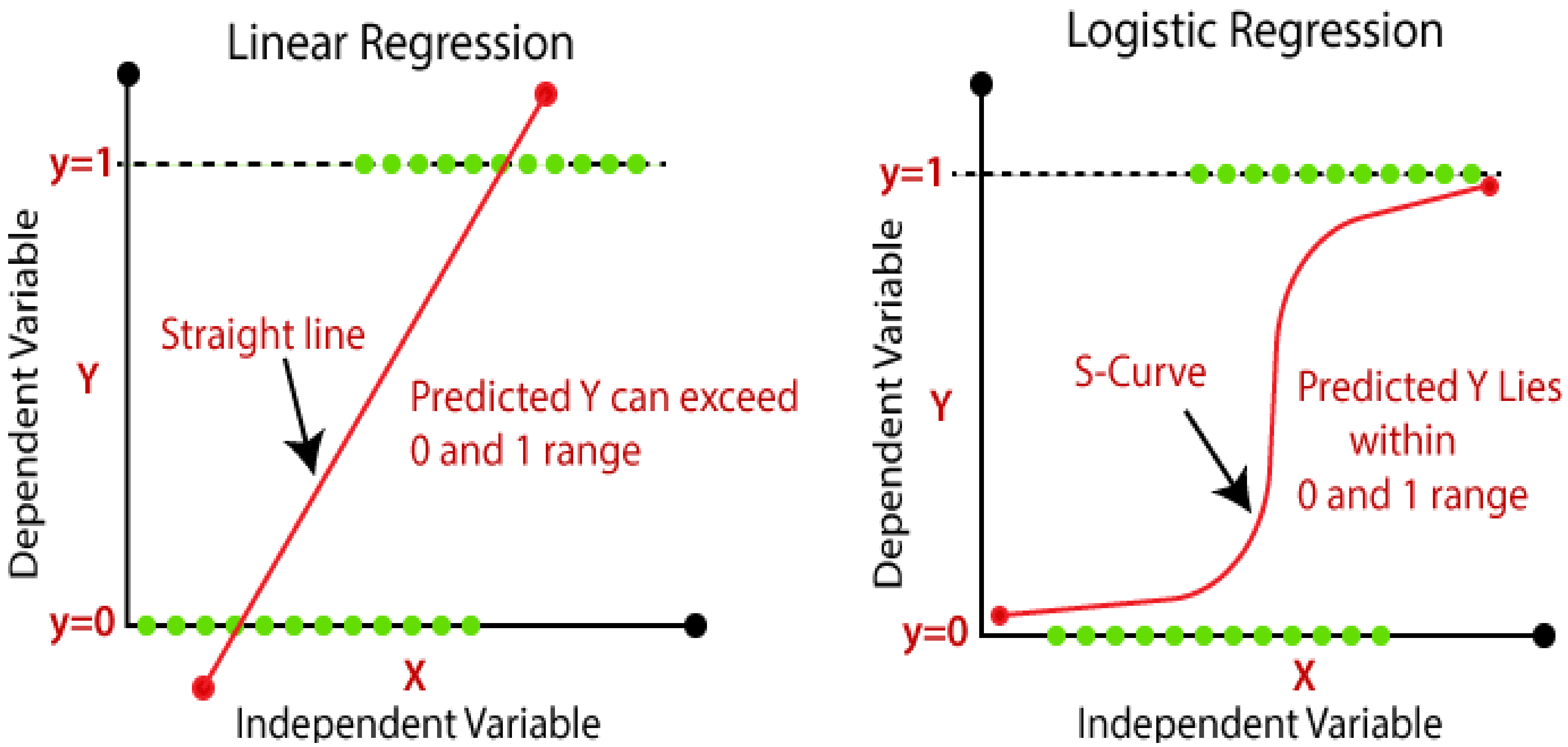

Temperature
Humidity
Wind
Logistic Regression Model
Sunny
Rainy



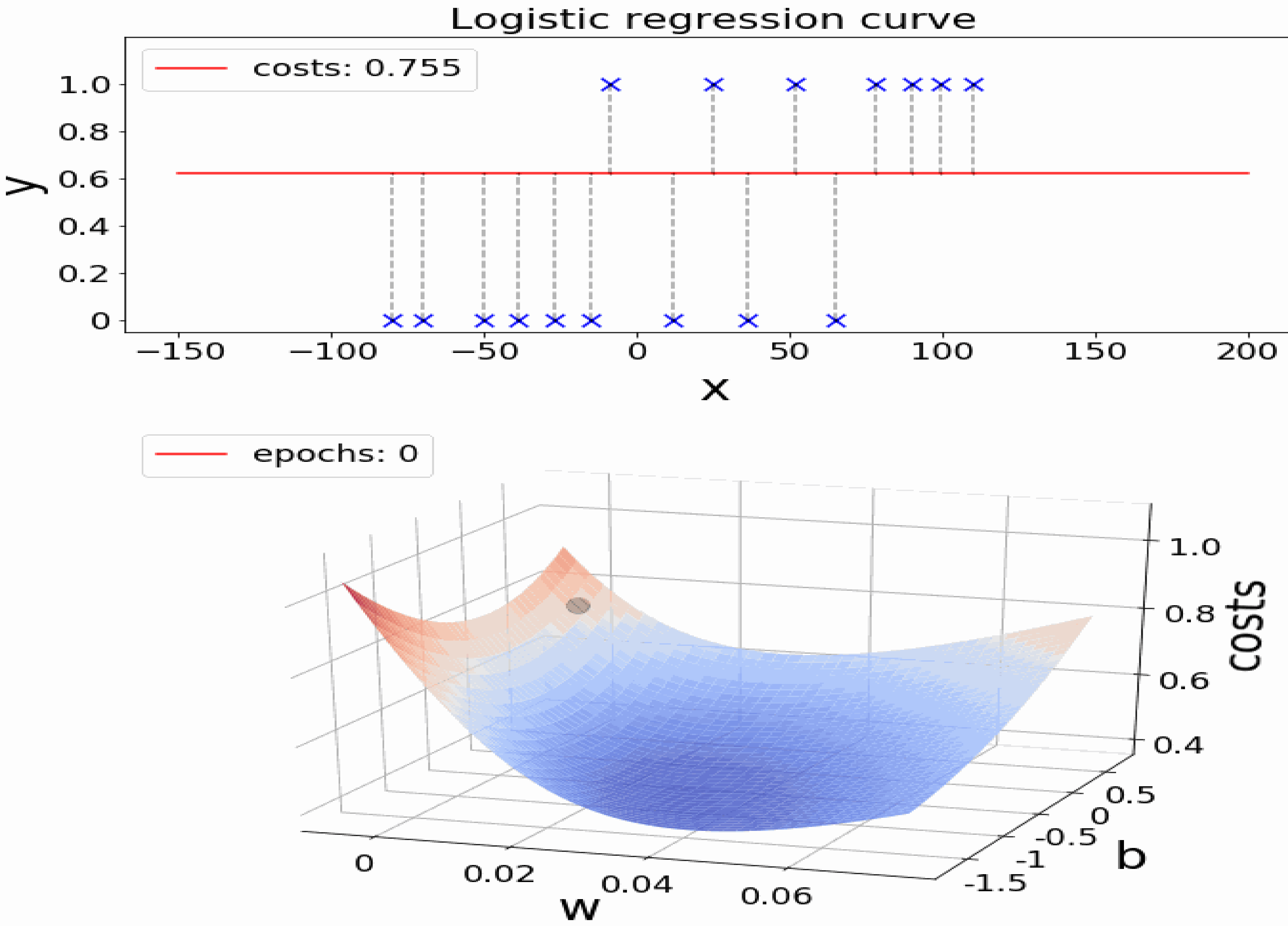
Logistic Regression



$\sigma(z) = \frac{1}{1+e^{-z}}$
If $\hat{y} > 0.5$
1
w
x
b
Linear
$z = xw^T + b$
Sigmoid
$\hat{y} = \sigma(z)$
$\hat{y}$
0
If $\hat{y} < 0.5$
y
1
0
x
y
1
0
x



Linear Regression
y=1
y=0
Dependent Variable
Y
Straight line
Predicted Y can exceed
0 and 1 range
X
Independent Variable
Logistic Regression
y=1
y=0
Dependent Variable
Y
S-Curve
Predicted Y Lies
within
0 and 1 range
X
Independent Variable



Logistic regression curve
costs: 0.755
y
x
epochs: 0
costs
w
b

Problem Statement

- **Titanic dataset**
- **Explore:** How does each feature relate to whether a person survives/alives?
- Do the EDA in more detail than usual and explain the results!
  - Splitting: 80-20, stratify: y, random_state = 0
- **Preprocessing:**
  - * Drop decks
  - * Fill in the missing value using a simple imputer
  - * One hot encoding: sex, alone
  - * Ordinal encoding: class
  - * Binary encoding: embark town
- **Model selection:**
  - * Evaluation metrics used: F1_score
  - Logistic Regression

Logistic Regression

Putting z value to sigmoid function

Linear Regression Equation

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k$$

Sigmoid Function

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

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

$$p = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k}}{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k} + 1}$$

Looks like very hard to solve it, so let's try to transform it into some easy to solve equation with the help of Odds ratio.

Odds Ratio $S = \frac{p}{1-p}$

Replace p and solve

$$S = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k}$$

Take log each side and solve

$$\mathbf{Ln(S)} = \boldsymbol{\beta_0} + \underbrace{\boldsymbol{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k}}_{\text{Transformed into Linear Regression}}$$

Notes:

- The log of Odds is called Logit and transformed model is linear in β_s
- So solving the logistic regression problem essentially reduces to finding the β_s that minimizes the error.
- Now suppose with one predictor we got the Linear Regression eq. ln(s) = -20.40782+.42592*x. And now we want to classify for given x = 50 then:
- Ln(s) = -20.40782+.42592*50 = 0.89 => s = $e^{0.89}$ = 2.435
- $s = \frac{p}{1-p}$ => $p = \frac{s}{s+1} \Rightarrow p = 2.435/(1+2.435) = .709$
- So using a probability of 0.50 as a cut-off between predicting the two classes 1 or 0, this member would be classified as class 1 with a probability of 70%