

SNS COLLEGE OF TECHNOLOGY

Kurumbapalayam (Po), Coimbatore – 641 035

An Autonomous Institution

Accredited by NAAC – UGC with ‘A++’ Grade

Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

**COURSE NAME: 23ADO201- DATA SCIENCE FUNDAMENTALS
(OPEN ELECTIVE)**

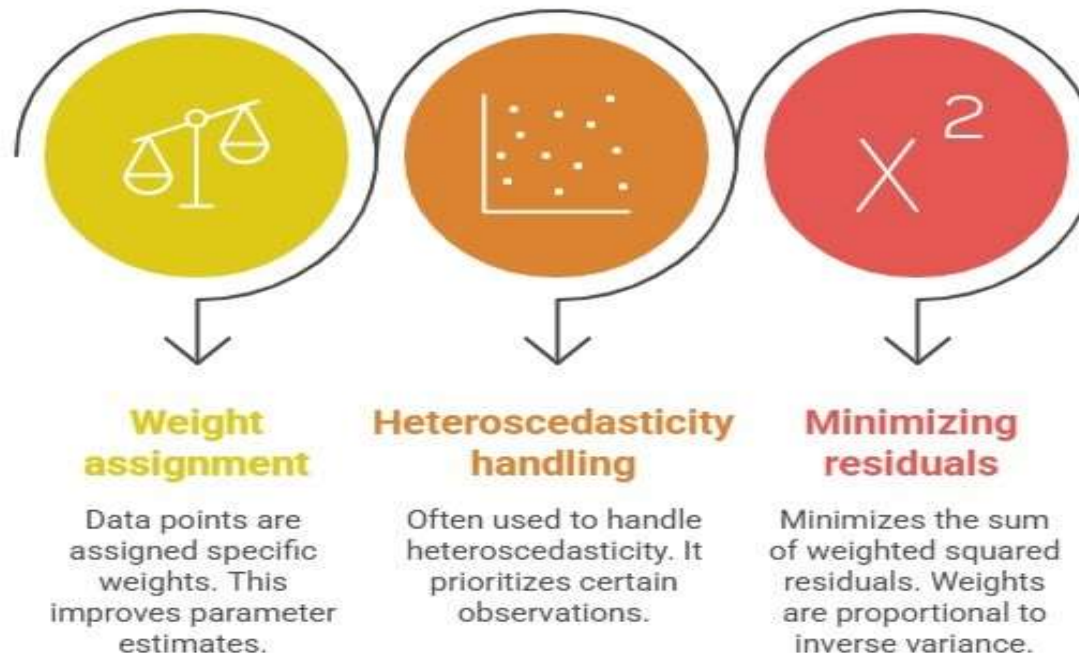
III YEAR / VI SEMESTER

Unit 5 - PREDICTIVE ANALYTICS

Topic : WEIGHTED RESAMPLING

Empathy

Weighted linear regression



Weighted Regression Models



Weighted Resampling & Ensemble Methods



```
import numpy as np
import matplotlib.pyplot as plt
```

```
# set the seed
np.random.seed(561)
```

```
# generate a single feature randomly
X0 = np.random.rand(100)
```

```
# actual interception and slope of linear regression
intercept = 2
slope = 5
```

```
# generate random observation noise (error)
# noise variance is a function of the feature (heteroscedastic noise)
noise = 3*np.abs(X0)*np.random.randn(X0.shape[0])
```

```
# generate the response variable
y = slope*X0 + intercept + noise
```

```
# generate the augmented feature matrix (bias + feature)
X = np.c_[np.ones(X0.shape[0]),X0]
```

```
# solution of linear regression
w_lr = np.linalg.inv(X.T @ X) @ X.T @ y
```

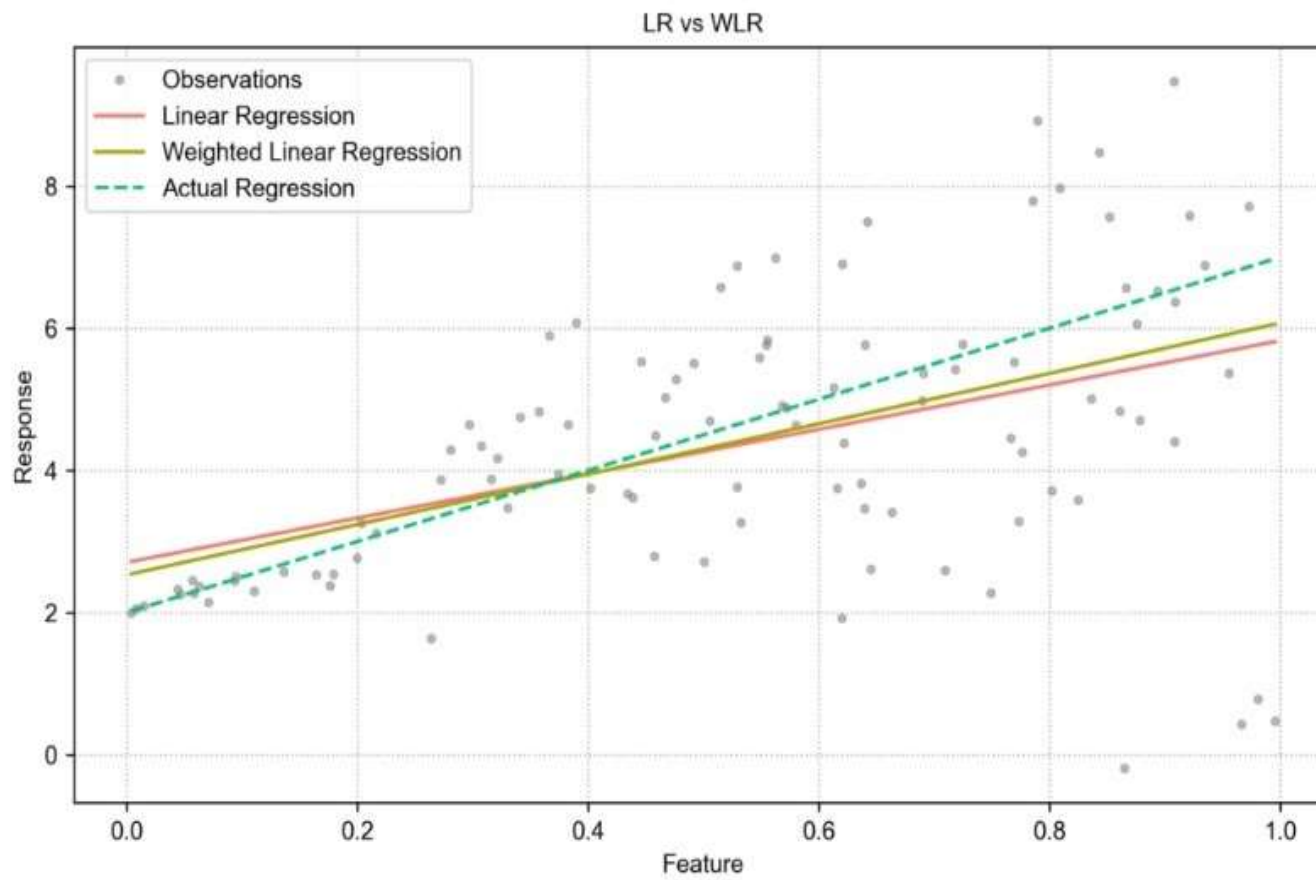
```
# calculate residuals
res = y - X @ w_lr
```

```
# estimate the covariance matrix
C = np.diag(res**2)
```

```
# solution of weighted linear regression
w_wlr = np.linalg.inv(X.T @ np.linalg.inv(C) @ X) @ (X.T @ np.linalg.inv(C) @ y)
```

```
# generate the feature set for plotting
X_p = np.c_[np.ones(2), np.linspace(X0.min(), X0.max(), 2)]
```

```
# plot the results
plt.plot(X0, y, 'b.', label='Observations')
plt.plot(X_p[:,1], X_p @ w_lr, 'r-', label='Linear Regression')
plt.plot(X_p[:,1], X_p @ w_wlr, 'g-', label='Weighted Linear Regression')
plt.plot(X_p[:,1], X_p @ [intercept, slope], 'm--', label='Actual Regression')
plt.grid(linestyle=':')
plt.ylabel('Response')
plt.xlabel('Feature')
plt.legend()
```



Activity



ACTIVITY: Weighted Resampling

Objective

To understand how **weighted resampling** works and how it helps focus on important or misclassified data points.

Given Data

Sample	Value	Weight
A	10	0.10
B	20	0.20
C	30	0.30
D	40	0.25
E	50	0.15

Student Tasks

1. Construct cumulative probability table
2. Generate 5–10 random numbers
3. Perform weighted resampling manually
4. Count frequency of each sample
5. Compare with original weights

MCQ



MCQs on Weighted Resampling (commonly used in statistics, machine learning, and bootstrapping):

1. Weighted resampling refers to:

- A) Selecting samples randomly with equal probability
- B) Selecting samples based on assigned weights
- C) Removing duplicate samples
- D) Sorting the dataset

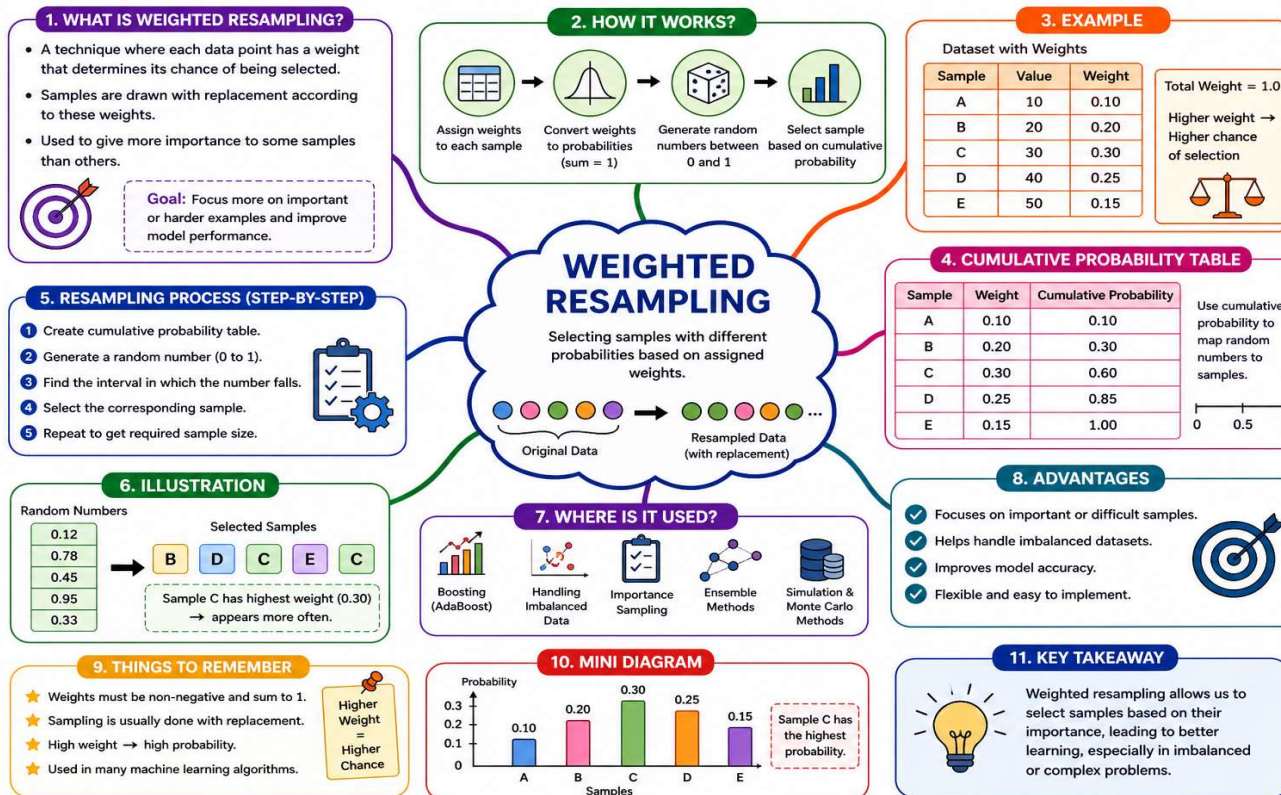
Answer: B) Selecting samples based on assigned weights

2. In weighted resampling, weights represent:

- A) Size of dataset
- B) Importance or probability of selecting each sample
- C) Number of variables
- D) Mean of data

Answer: B) Importance or probability of selecting each sample

MINDMAP



TEXT BOOKS	
1.	Srinivasan Desikan and Gopaldaswamy Ramesh, —Software Testing – Principles and Practices, Pearson Education, 2006.
REFERENCES	
1.	Ilene Burnstein, —Practical Software Testing, Springer International Edition, 2003
2.	Edward Kit, Software Testing in the Real World – Improving the Process, Pearson Education, 1995
3.	Boris Beizer, Software Testing Techniques – 2nd Edition, Van Nostrand Reinhold, New York, 1990.
4.	Aditya P. Mathur, —Foundations of Software Testing _ Fundamental Algorithms and Techniques, Dorling Kindersley (India) Pvt. Ltd., Pearson Education, 2008.

THANK YOU!

11