* Univariate, Bivariate & Multivariate Analysis
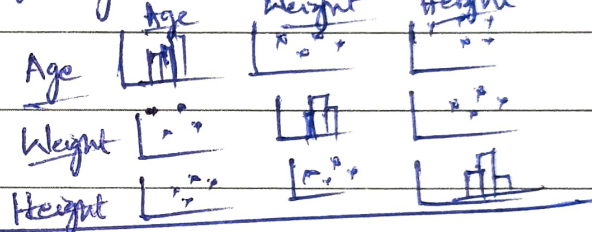
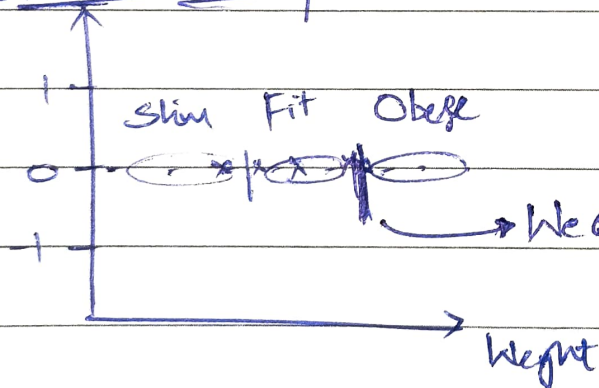| Height cm	Weight kg	o/p
180	90	Obesity
160	50	Slim
170	78	Fit
190	90	Fit
175	85	Slim

Multivariate

More than 3 features - Age, Weight, Height
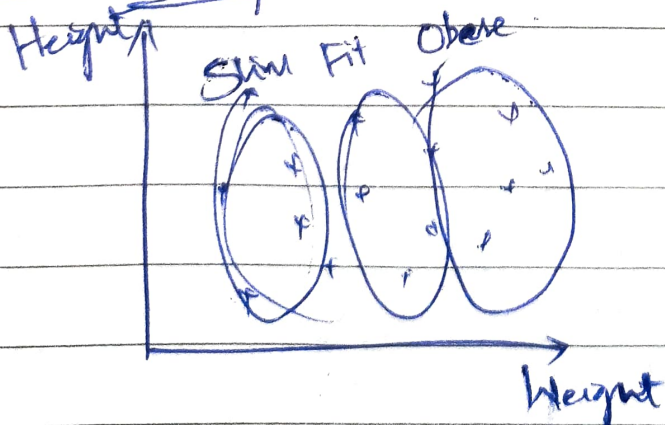
We go with Pairplot in SeaBorn

	Age	Weight	Height
Age			
Weight			
Height			

Gives concept of Correlation.

Univariate Analysis

Slim Fit Obese

→ We may be 100% Correct. Overlap point will be there

We go for Bivariate/Multivariate

Bivariate Analysis

Slim Fit Obese

Note

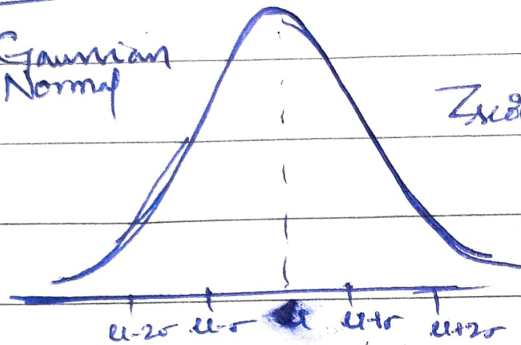If We are able to Classify point Properly, We can go for Logistic Regression

Logistic Regression :- → Linear

If Overlapping Occurs, We go with (NonLinear) Decision Tree / XG Boost / Random Forest
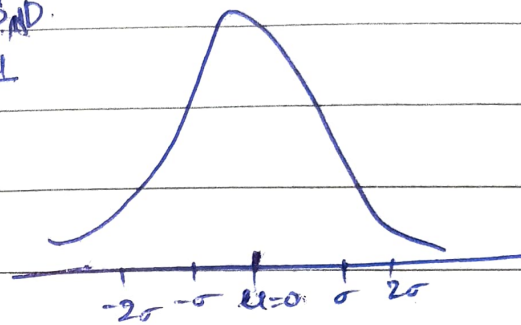SVM / Adaboost / KNN Algo.

Z-score:-

Gaussian Normal

$$Z\text{-score} = \frac{x_i - \mu}{\sigma}$$ S.N.D.

$\mu - 2\sigma \quad \mu - \sigma \quad \mu \quad \mu + \sigma \quad \mu + 2\sigma$

$-2\sigma \quad -\sigma \quad \mu = 0 \quad \sigma \quad 2\sigma$

68%

95%

$\mu + 3\sigma$

$\mu \cdot 3\sigma$

99%

$\mu = 75$
$\sigma = 10$

$P(x > 60)$

Gaussian Normal

S.N.D.

50%

44%

6%

55 60 65 75 85 95
$\mu - 2\sigma \quad \mu + \sigma \quad \mu$

-2 -1 0 +1 +2

-1.5
(z)
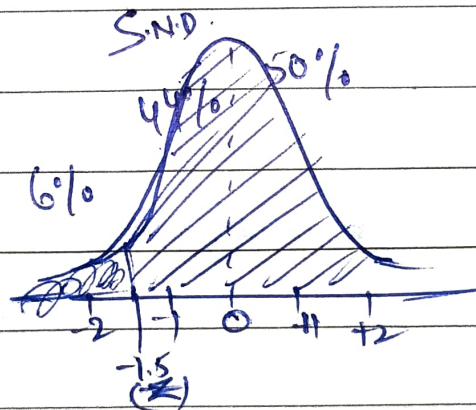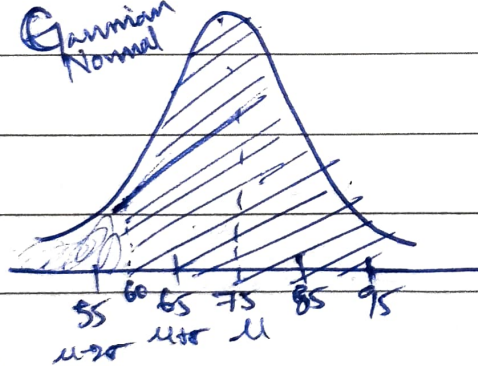
Z-score Table gives Area of left Handside (z)   7500

7.500

$= -0.06681$ ie   6.681%

$\dfrac{100 \cdot 999}{6.681}$

Remaining Area   93.319%

93.819

(Probability) $\approx 94\%$

Histograms:-
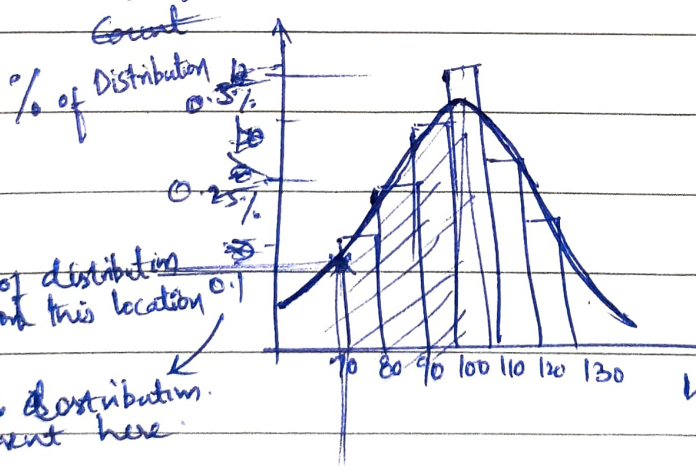
The given range

Gives Information of How many No. of points in a Given range (box)

ie ⇒ Builds Bargraph.

Count

Prob. density function
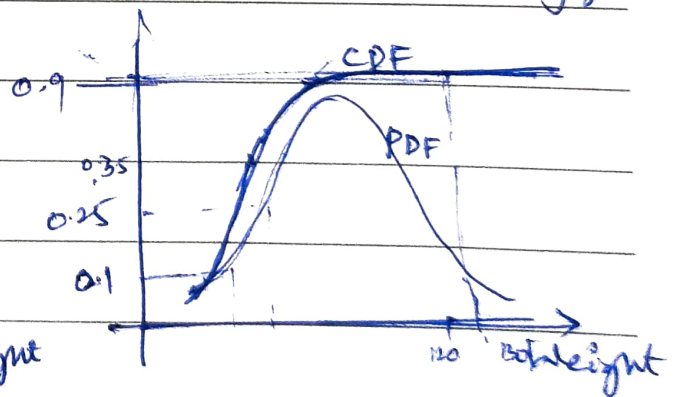BELL CURVE
NORMAL DISTRIBUTION.

20

10

70 80 90 100

Weight

Prob. density function PDF

Count

% of Distribution
0.5%

0.25%

Total % of distribution
present on this location 0.1

10% distribution
present here

70 80 90 100 110 120 130   Weight
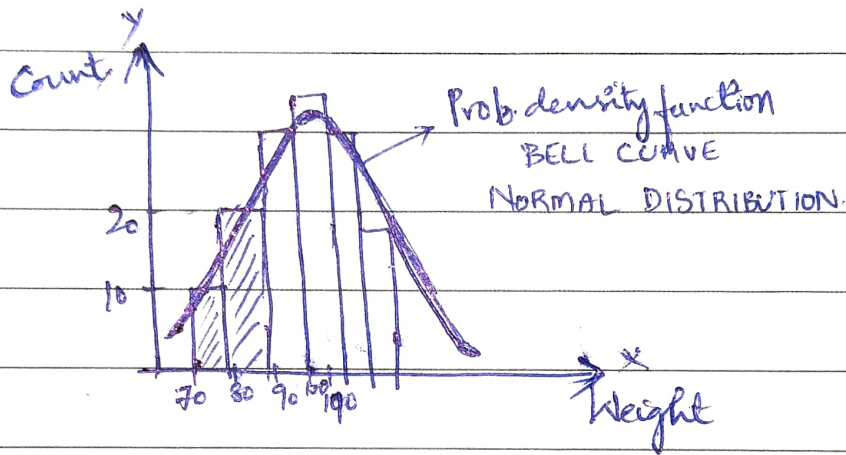
CDF

Cumulative Density function

0.9   CDF

                PDF

0.35

0.25

0.1

120   Both Weight

0.1+0.25   90% < 130 kgs