Aman Mehta - 112119010

I Bearing roll no 112119010 agree and acknowledge that:

1. The assignment was answered by me as per instruction applicable to each assessment and that I haven't resorted to any unfair means to deleberately to improve my performance.

2. I have nither impersonated anyone nor I have been impersonated by any person for the purpose of assessment

Signature of the student : Aman Mehta

Full Name :- Aman Mehta

Roll NO :- 112119010

Sub Code :- CS DE 18

Mob NO :- 8201561466

Aman Mehta - 112119010

(8) Spearmen correlation coefficient =.

X value.	Y value.	X_{RA}	$X_{RA} - M_X$	Y_{RA}	$Y_{RA} - M_Y$
90	35	3	-0.5	4	0.5
151	49	6	2.5	5	1.50
145	60	5	1.5	6	2.5
55	14	2	-1.5	2	-1.5
35	31	1	-2.5	3	-0.5
140	9	4	0.5	1	-2.5

Sum diffs:

- 2.5
- 3.75
- 3.75
- 2.25
- 1.25
- -1.25

calculation

R = covariance / (X_{RA} St.Dev + Y_{RA} St.Dev)

where

X_{RA} = Rank of X value.

Y_{RA} = Rank of Y value.

$X_{RA} - M_X$ = X rank - mean of X Rank

$Y_{RA} - M_Y$ = Y rank - mean of Y rank

Sum Diff = $(X_{RA} - M_X) * (Y_{RA} - M_Y)$

Aman Mehta - 112119010

Result

$M_x = 3.5$ Standard divation = 1.87

$M_{y_1} = 3.5$ Standard divation = 1.87

Combine covariance = 9.5/5 = 1.9

$R = 1.9 / (1.87 * 1.87) = 0.543.$

$r_s = 0.54286$ p (2-tailed) = 0.2657

So the value of spearman correlation coefficient

0.54286.

As the value of this coefficient is 0.54286 which is

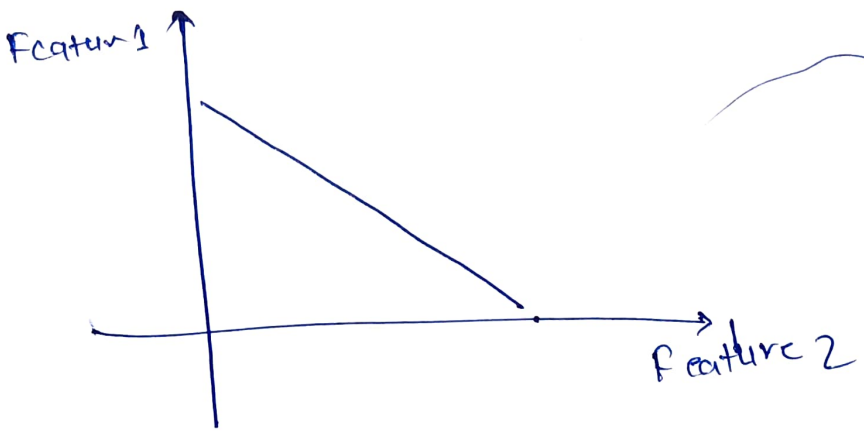
The value is very close to 0.5 hence the two are not so correlated and we can say they are corelated if we take threshold value as 0.54286 But they are no correlated as in general.

7. The formula for the false positive rate in term of TP TN fP and fN is.

$$\text{false positive rate} = \frac{fP}{fP + TN}$$

Where fP is false position numbers
TN is the number of True negative
and fP + TN being total no of
Negative.

4. If the value of the covariance between two feature is -ve and > 0.95 then it signify only the direction between the two feature that one is increasing then other is decreasing in nature.



Aman Mehta — 11211910

Q5 1) feature scaling :- ② Bring the values of a feature between certain range.

2) feature selection :- ④ Reduce the number of features.

3) PCA :- ① Represent the data in a lesser dimension space.

④ Smote — ③ Handle class inbalance problems.

Q2. (C) false positive

Q3. elbow plot.

Aman Mehta - 11211010

(1) The number of principal component axis possible from the data set comparision of 50 features are 50

If we want to diminesionality reduction using identified axes then max no of feature can be droped will be 48 feature all. we requid two feature to make 2D principal componant.

6. 

new question

We want to train our model, efficiency our training data set should have instance of both the target class as well as our test dataset should have atleast one instance of both the target class So that we could test prefomance of our model. On the 'class labels.

the data set we have has equal instance of both the target classes (0, and 1) so there is no class imbalance.

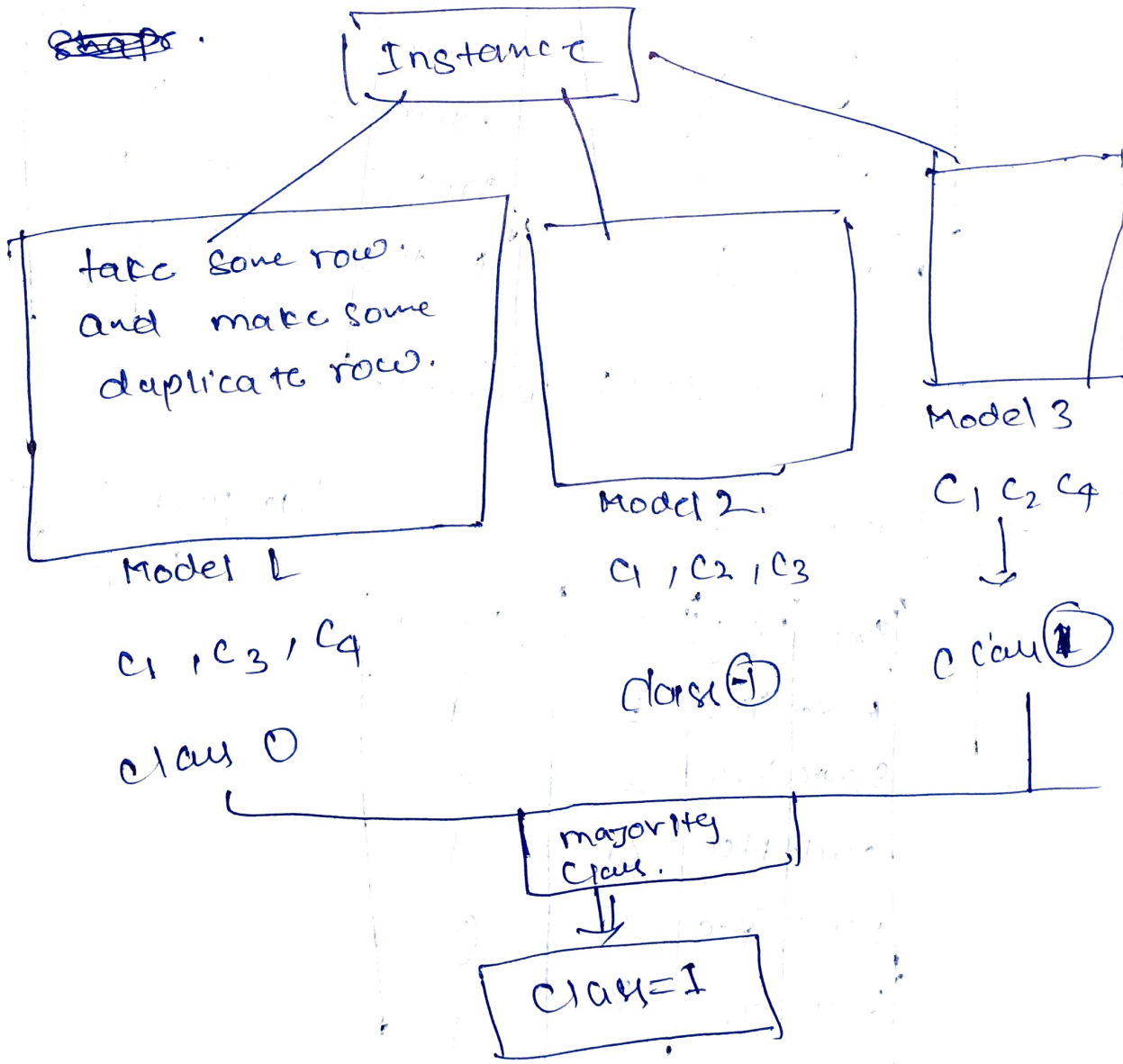
If we randomly split our data set to training and testing dataset we might miss out having both the class labeled in the respective set.

we can use leave one out cross validation that is using only one sample as testing and the rest for training but this would increase the computation time and model will be low bias.

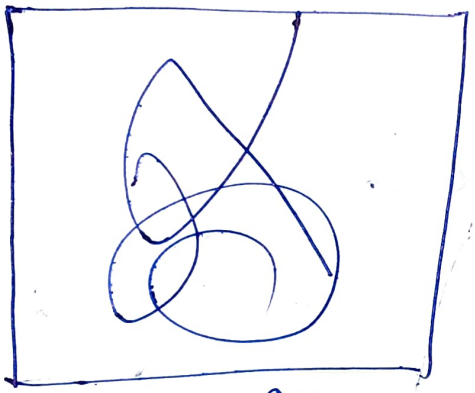
(b) The same problem happen with k fold cross validation

* We can't ensure the sample of both class labeled would make into the training and testing data set

* We must go for satisfied cross validation this makes sure that the problem instance of each class is present in both train and test dataset.



we will take few row and column and create some duplicate rows and column and also we take few column so that we can considered both row and feature Dampling with ~~[illegible]~~ replacement technique as we replace them with previous prediction



Dataset ①

Shape	# of diag	length	Hight	Target
Rect	2	100	10	1
square	2	100	100	0
square	2	100000	100000	0
Triangle	0	100	1000	0
Rectang	2	100	10	1
....				

(droped diag column)

→ Dataset 2

(droped length and hight)

Shape	Diag	# of diag	Target
Recta	Yes	2	1
Square	Yes	2	0
circle	NO	0	1
Pramo	Yes	2	0
Recta	Yes	2	1

Dataset (3)

shape.	target
Rect	1
Rect	1
Square	0
Diamond	0
Circle	1