PYTHON WORKSHEET 1

1. C) %
2. B)0
3. C)24
4. A)2
5. D)6
6. C) the finally block will be executed no matter if the try block raises an error or not.
7. A) It is used to raise an exception.
8. C) in defining a generator
9. A) _abc , C)abc2
10. A) yield , B) raise

MACHINE LEARNING WORKSHEET

1. A) Least Square Error
2. A) Linear regression is sensitive to outliers
3. B) Negative
4. C) Both of them
5. C) Low bias and high variance
6. B) Predictive modal
7. D) Regularization
8. D) SMOTE
9. A) TPR and FPR
10. B) False
11. B) Apply PCA to project high dimensional data
12. A) We don't have to choose the learning rate. B) It becomes slow when number of features is very large

13. I) Regularization is a technique used to induce some added information in machine learning model to reduce overfitting.
    ii) Overfitting is a phenomenon where machine learning models perform exceptionally well on training data but not so well on testing data i.e. variance is very high.
    iii)In regularization, we add an extra term to the models loss function, $\lambda||w^2||$
    where λ is a hyperparameter that can be tuned .
    iv) By adding an extra term, we can convince the model to fit the testing data instead of training data. As adding an extra term, reduces the loss function of the model hence leads to fitting the testing data.
    v)As a slight increase in bias, leads to reduced variance, higher λ value increases the level of regularization.
    vi) There are three types of regularization: Lasso regression, Ridge regression, elastic net regression.

14. The algorithms used for regularization are:
    i) Ridge regression: also known as L2 norm. Function of L2 is the sum of the original loss function and the squared sum of the coefficients(w) multiplied by a hyperparameter λ. If λ increases then w decreases but is never zero.

$$L=MSE + \lambda||w||^2$$

We use ridge regression in scenarios where all the input columns are important and we don't want to remove any information. We call it L2 as its multiplied by square of coefficients.

ii) Lasso regression: also known as L1 norm. It has same functionality as ridge regression except in MSE we add L1 norm (sum of modulus of coefficients). If we increase the hyperparameter λ too much, w will eventually equal 0.

$$L=MSE + \lambda||w||$$

We use lasso regression for feature selection, that is if a particular column is not that important then lasso can remove it. We will use lasso when we are sure that some input columns are important for predicting target variable and some are not important.

iii) Elastic net regression: It is a combination of ridge and lasso regression.

$$L=MSE + a||w||^2 + b||w||$$

Where : $\lambda=a+b$, $l1\_ratio=\frac{a}{a+b}$

We use elastic net regression, when we have a big dataset, where we cant guess the importance of input columns to predict the target variable . Also used in datasets where input columns have multi collinearity, i.e. when two columns are dependent on each other too much in that case elastic net regression should be used.

15. I) The term error also known as residual refers to the difference between the actual observed values and the predicted values of the linear regression model.
ii) The linear regression the goal is to model the relationship between a dependent variable y and an independent variable x by fitting a linear equation to the observed data.

$$Y=mx+b$$

Where m= slope of the line
B= y intercept
iii)For selecting the best fit line, we will choose the line with the minimum error/residual.
iv)For this, we calculate the distance wrt to all the points to calculate the residual.

$$d_1^2 + d_2^2 + ... + d_n^2 = D$$

Where D is the SS (sum of the squared residuals).

v)Now if we have three lines, to select the best line we will compare their SS and select the one with minimum SS

$$SS_1 < SS_2 < SS_3$$

In this case we will select the first line as the best fit line.