WORKSHEET 3 PYTHON

Q1 to Q8 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following will raise a value error in python? A) int(32) B) int(3.2) C) int(-3.2) D) int('32')
2. What will be the output of round(3.567)? A) 3.5 B) 3.0 C) 4 D) 3
3. How is the function pow(a,b,c) evaluated in python? A) a**b**c B) (a**b)%c C) (a**b)*c D) (a**b)**c
4. What will be the output of print(type(type(int))) in python 3? A) <class 'type'> B) <type 'type'> C) <class 'int'> D) <type 'int'>
5. What will be the output of ord(chr(65))? A) 'A' B) 'a' C) 65 D) TypeError
6. What is called when a function is defined inside a class? A) Module B) Function C) *init* function D) Method
7. What will be the output of all([1, 0, 5 ,7])? A) 0 B) False C) True D) error
8. Is the output of the function abs() the same as that of the function math.fabs()? A) Always B) Sometimes C) Never D) None of these Q9 and Q10 have multiple correct answers. Choose all the correct options to answer your question.
9. Select all correct float numbers in python? A) -68.7e100 B) 42e3 C) 4.2038 D) 3.0
10. Which of the following is(are) correct statement(s) in python? A) You can pass positional arguments in any order. B) You can pass keyword arguments in any order. C) You can call a function with positional and keyword arguments. D) Positional arguments must be before keyword arguments in a function call Q11 to Q15 are programming questions. Answer them in Jupyter Notebook.
11. Write a python function print pyramid of stars. Level of the pyramid should be taken as an input from the user. E.g. Input = 5 Output: ASSIGNMENT
12. Write a python function print Hourglass pattern. E.g. Input = 5 Output:
13. Write a python function to print Pascal's Triangle. The number of levels in the triangle must be taken as input by the user. E.g. Input = 5 Output: 1 1 1 1 2 1 1 3 3 1 1 4 6 4 1
14. Write a python function to print Diamond Shaped Pattern shown below. Function must take integer input which represents the number of stars in the middle most line. E.g.: Input = 5 Output:
15. Write a python function to print Diamond Shaped Character Pattern shown below. Function must take integer input within range 1 to 26, which represents the rank of the alphabet. E.g.: Input = 5

Q1) Which of the following will raise a value error in python?

A) int(32) B) int(3.2) C) int(-3.2) D) int('32')

ANSWER: D) int('32')

Q2) What will be the output of round(3.567)?

A) 3.5 B) 3.0 C) 4 D) 3

ANSWER: C) 4

Q3) How is the function pow(a,b,c) evaluated in python?

A) a**b**c B) (a**b)%c C) (a**b)*c D) (a**b)**c

ANSWER: B) (a**b)%c

Q4) What will be the output of print(type(type(int))) in python 3?

A) <class 'type'> B) <type 'type'> C) <class 'int'> D) <type 'int'>

ANSWER: B) <type 'type'>

Q5) What will be the output of ord(chr(65))?

A) 'A' B) 'a' C) 65 D) TypeError

ANSWER: C) 65

Q6) What is called when a function is defined inside a class?

A) Module B) Function C) *init* function D) Method

ANSWER: D) Method

Q7) What will be the output of all([1, 0, 5 ,7])?

A) 0 B) False C) True D) error

ANSWER: B) False

Q8) Is the output of the function abs() the same as that of the function math.fabs()?

A) Always B) Sometimes C) Never D) None of these

ANSWER: B) Sometimes.

Q9) Select all correct float numbers in python?

A) -68.7e100 B) 42e3 C) 4.2038 D) 3.0

ANSWER: A) -68.7e100 C) 4.2038 D) 3.0

Q10) Which of the following is(are) correct statement(s) in python?

A) You can pass positional arguments in any order. B) You can pass keyword arguments in any order. C) You can call a function with positional and keyword arguments. D) Positional arguments must be before keyword arguments in a function call

ANSWER: B) You can pass keyword arguments in any order. C) You can call a function with positional and keyword arguments. D) Positional arguments must be before keyword arguments in a function call

Q11) Write a python function print pyramid of stars. Level of the pyramid should be taken as an input from the user. E.g. Input = 5

In [1]:
```
rows = int(input("Enter number of rows: "))

k = 0

for i in range(1, rows+1):
    for space in range(1, (rows-i)+1):
        print(end="  ")

    while k!=(2*i-1):
        print("* ", end="")
        k += 1

    k = 0
    print()
```

```
Enter number of rows: 5
        *
      * * *
    * * * * *
  * * * * * * *
* * * * * * * * *
```

Q12) Write a python function print Hourglass pattern. E.g. Input = 5

In [2]:
```
rows = int(input("Enter number of rows: "))

for i in range(rows, 1, -1):
    for space in range(0, rows-i):
        print("  ", end="")
    for j in range(i, 2*i-1):
        print("* ", end="")
```

```
    for j in range(1, i-1):
        print("* ", end="")
    print()

k = 0

for i in range(1, rows+1):
    for g in range(1, (rows-i)+1):
        print(end="  ")

    while k!=(2*i-1):
        print("* ", end="")
        k += 1

    k = 0
    print()
```

```
Enter number of rows: 5
* * * * * * *
  * * * * *
    * * *
      *
      *
    * * *
  * * * * *
* * * * * * * *
```

Q13) Write a python function to print Pascal's Triangle. The number of levels in the triangle must be taken as input by the user. E.g. Input = 5

In [3]:
```
rows = int(input("Enter number of rows: "))
coef = 1

for i in range(1, rows+1):
    for space in range(1, rows-i+1):
        print(" ",end="")
    for j in range(0, i):
        if j==0 or i==0:
            coef = 1
        else:
            coef = coef * (i - j)//j
        print(coef, end = " ")
    print()
```

```
Enter number of rows: 5
    1
   1 1
  1 2 1
 1 3 3 1
1 4 6 4 1
```

Q14) Write a python function to print Diamond Shaped Pattern shown below. Function must take integer input which represents the number of stars in the middle most line. E.g.: Input = 5

In [4]:
```
h = int(input("please enter diamond's height:"))

for i in range(h):
    print(" "*(h-i),'*' *(i*2+1))
for i in range(h-2, -1, -1):
    print(" "*(h-i), '*'*(i*2+1))
```

```
please enter diamond's height:5
      *
     ***
    *****
   *******
  *********
   *******
    *****
     ***
      *
```

Q15) Write a python function to print Diamond Shaped Character Pattern shown below. Function must take integer input within

range 1 to 26, which represents the rank of the alphabet. E.g.: Input = 5

In [5]:
```
h = int(input("please enter diamond's height:"))
ascii_value=65

for i in range(h):
    alphabet=chr(ascii_value)
    print(" "*(h-i),alphabet *(i*2+1))
for i in range(h-2, -1, -1):
    print(" "*(h-i), alphabet*(i*2+1))
```

```
please enter diamond's height:5
      A
     AAA
    AAAAA
   AAAAAAA
  AAAAAAAAA
   AAAAAAA
    AAAAA
     AAA
      A
```

STATISTICS WORKSHEET-10

Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.

1. Rejection of the null hypothesis is a conclusive proof that the alternative hypothesis is a. True b. False c. Neither
2. Parametric test, unlike the non-parametric tests, make certain assumptions about a. The population size b. The underlying distribution c. The sample size
3. The level of significance can be viewed as the amount of risk that an analyst will accept when making a decision a. True b. False
4. By taking a level of significance of 5% it is the same as saying a. We are 5% confident the results have not occurred by chance b. We are 95% confident that the results have not occurred by chance c. We are 95% confident that the results have occurred by chance
5. One or two tail test will determine a. If the two extreme values (min or max) of the sample need to be rejected b. If the hypothesis has one or possible two conclusions c. If the region of rejection is located in one or two tails of the distribution
6. Two types of errors associated with hypothesis testing are Type I and Type II. Type II error is committed when a. We reject the null hypothesis whilst the alternative hypothesis is true b. We reject a null hypothesis when it is true c. We accept a null hypothesis when it is not true
7. A randomly selected sample of 1,000 college students was asked whether they had ever used the drug Ecstasy. Sixteen percent (16% or 0.16) of the 1,000 students surveyed said they had. Which one of the following statements about the number 0.16 is correct? a. It is a sample proportion. b. It is a population proportion. c. It is a margin of error. d. It is a randomly chosen number.
8. In a random sample of 1000 students, pˆ = 0.80 (or 80%) were in favour of longer hours at the school library. The standard error of pˆ (the sample proportion) is a. .013 b. .160 c. .640 d. .800 WORKSHEET
9. For a random sample of 9 women, the average resting pulse rate is x = 76 beats per minute, and the sample standard deviation is s = 5. The standard error of the sample mean is a. 0.557 b. 0.745 c. 1.667 d. 2.778
10. Assume the cholesterol levels in a certain population have mean μ= 200 and standard deviation σ = 24. The cholesterol levels for a random sample of n = 9 individuals are measured and the sample mean x is determined. What is the z-score for a sample mean x = 180? a. –3.75 c. –2.50 c. −0.83 d. 2.50
11. In a past General Social Survey, a random sample of men and women answered the question “Are you a member of any sports clubs?” Based on the sample data, 95% confidence intervals for the population proportion who would answer “yes” are .13 to .19 for women and .247 to .33 for men. Based on these results, you can reasonably conclude that a. At least 25% of American men and American women belong to sports clubs. b. At least 16% of American women belong to sports clubs. c. There is a difference between the proportions of American men and American women who belong to sports clubs. d. There is no conclusive evidence of a gender difference in the proportion belonging to sports clubs.
12. Suppose a 95% confidence interval for the proportion of Americans who exercise regularly is 0.29 to 0.37. Which one of the following statements is FALSE? a. It is reasonable to say that more than 25% of Americans exercise regularly. b. It is reasonable to say that more than 40% of Americans exercise regularly. c. The hypothesis that 33% of Americans exercise regularly cannot be rejected. d. It is reasonable to say that fewer than 40% of Americans exercise regularly. Q13 to Q15 are subjective answers type questions. Answers them in their own words briefly.
13. How do you find the test statistic for two samples?
14. How do you find the sample mean difference?
15. What is a two sample t test example?

Q1) Rejection of the null hypothesis is a conclusive proof that the alternative hypothesis is

a. True b. False c. Neither

ANSWER: a) True

Q2) Parametric test, unlike the non-parametric tests, make certain assumptions about

a. The population size b. The underlying distribution c. The sample size

ANSWER: b) The underlying distribution.

Q3) The level of significance can be viewed as the amount of risk that an analyst will accept when making a decision

a. True b. False

ANSWER: a) True

Q4) By taking a level of significance of 5% it is the same as saying

a. We are 5% confident the results have not occurred by chance b. We are 95% confident that the results have not occurred by chance c. We are 95% confident that the results have occurred by chance

ANSWER: C) We are 95% confident that the results have occurred by chance.

Q5) One or two tail test will determine

a. If the two extreme values (min or max) of the sample need to be rejected b. If the hypothesis has one or possible two conclusions c. If the region of rejection is located in one or two tails of the distribution

ANSWER: a) If the two extreme values (min or max) of the sample need to be rejected .

Q6) Two types of errors associated with hypothesis testing are Type I and Type II. Type II error is committed when

a. We reject the null hypothesis whilst the alternative hypothesis is true b. We reject a null hypothesis when it is true c. We accept a null hypothesis when it is not true

ANSWER: c) We accept a null hypothesis when it is not true

Q7) A randomly selected sample of 1,000 college students was asked whether they had ever used the drug Ecstasy. Sixteen percent (16% or 0.16) of the 1,000 students surveyed said they had. Which one of the following statements about the number 0.16 is correct?

a. It is a sample proportion. b. It is a population proportion. c. It is a margin of error. d. It is a randomly chosen number.

ANSWER: a) It is a sample proportion.

Q8) In a random sample of 1000 students, $\hat{p} = 0.80$ (or 80%) were in favour of longer hours at the school library. The standard error of $\hat{p}$ (the sample proportion) is

a. .013 b. .160 c. .640 d. .800

ANSWER: b) 0.160

Q9) For a random sample of 9 women, the average resting pulse rate is x = 76 beats per minute, and the sample standard deviation is s = 5. The standard error of the sample mean is

a. 0.557 b. 0.745 c. 1.667 d. 2.778

ANSWER: c) 1.667

Q10) Assume the cholesterol levels in a certain population have mean μ= 200 and standard deviation σ = 24. The cholesterol levels for a random sample of n = 9 individuals are measured and the sample mean x is determined. What is the z-score for a sample mean x = 180?

a. –3.75 c. –2.50 c. −0.83 d. 2.50

ANSWER: c) -0.83

Q11) In a past General Social Survey, a random sample of men and women answered the question “Are you a member of any sports clubs?” Based on the sample data, 95% confidence intervals for the population proportion who would answer “yes” are .13 to .19 for women and .247 to .33 for men. Based on these results, you can reasonably conclude that

a. At least 25% of American men and American women belong to sports clubs. b. At least 16% of American women belong to sports clubs. c. There is a difference between the proportions of American men and American women who belong to sports clubs. d. There is no conclusive evidence of a gender difference in the proportion belonging to sports clubs.

ANSWER: c) There is a difference between the proportions of American men and American women who belong to sports clubs.

Q12) Suppose a 95% confidence interval for the proportion of Americans who exercise regularly is 0.29 to 0.37. Which one of the following statements is FALSE?

a. It is reasonable to say that more than 25% of Americans exercise regularly. b. It is reasonable to say that more than 40% of Americans exercise regularly. c. The hypothesis that 33% of Americans exercise regularly cannot be rejected. d. It is reasonable to say that fewer than 40% of Americans exercise regularly.

ANSWER: b) It is reasonable to say that more than 40% of Americans exercise regularly.

Q13) How do you find the test statistic for two samples?

ANSWER:

To calculate a test statistic:

1. Find the raw scores of the populations
2. Assume you want to perform a z-test to determine whether the means of two populations are equal.
3. Calculate the standard deviation of the population
4. Find the standard deviation of the population you're evaluating.
5. Calculate the population mean ...
6. Evaluate the z-value ...
7. Apply the t-test formula ...

Q14) How do you find the sample mean difference?

Q14) How do you find the sample mean difference?

ANSWER:

The formula for the mean of the sampling distribution of the difference between means is: μm1–m2 = μ1 – μ2

Q15) What is a two sample t test example?

ANSWER:

The two-sample t-test (also known as the independent samples t-test) is a method used to test whether the unknown population means of two groups are equal or not.

MACHINE LEARNING

ASSIGNMENT - 8 In Q1 to Q7, only one option is correct, Choose the correct option:

1. What is the advantage of hierarchical clustering over K-means clustering? A) Hierarchical clustering is computationally less expensive B) In hierarchical clustering you don't need to assign number of clusters in beginning C) Both are equally proficient D) None of these
2. Which of the following hyper parameter(s), when increased may cause random forest to over fit the data? A) max_depth B) n_estimators C) min_samples_leaf D) min_samples_splits
3. Which of the following is the least preferable resampling method in handling imbalance datasets? A) SMOTE B) RandomOverSampler C) RandomUnderSampler D) ADASYN
4. Which of the following statements is/are true about "Type-1" and "Type-2" errors?
5. Type1 is known as false positive and Type2 is known as false negative.
6. Type1 is known as false negative and Type2 is known as false positive.
7. Type1 error occurs when we reject a null hypothesis when it is actually true. A) 1 and 2 B) 1 only C) 1 and 3 D) 2 and 3
8. Arrange the steps of k-means algorithm in the order in which they occur:
9. Randomly selecting the cluster centroids
10. Updating the cluster centroids iteratively
11. Assigning the cluster points to their nearest center A) 3-1-2 B) 2-1-3 C) 3-2-1 D) 1-3-2
12. Which of the following algorithms is not advisable to use when you have limited CPU resources and time, and when the data set is relatively large? A) Decision Trees B) Support Vector Machines C) K-Nearest Neighbors D) Logistic Regression
13. What is the main difference between CART (Classification and Regression Trees) and CHAID (Chi Square Automatic Interaction Detection) Trees? A) CART is used for classification, and CHAID is used for regression. B) CART can create multiway trees (more than two children for a node), and CHAID can only create binary trees (a maximum of two children for a node). C) CART can only create binary trees (a maximum of two children for a node), and CHAID can create multiway trees (more than two children for a node) D) None of the above In Q8 to Q10, more than one options are correct, Choose all the correct options:
14. In Ridge and Lasso regularization if you take a large value of regularization constant(lambda), which of the following things may occur? A) Ridge will lead to some of the coefficients to be very close to 0 B) Lasso will lead to some of the coefficients to be very close to 0 C) Ridge will cause some of the coefficients to become 0 D) Lasso will cause some of the coefficients to become 0. MACHINE LEARNING ASSIGNMENT - 8
15. Which of the following methods can be used to treat two multi-collinear features? A) remove both features from the dataset B) remove only one of the features C) Use ridge regularization D) use Lasso regularization
16. After using linear regression, we find that the bias is very low, while the variance is very high. What are the possible reasons for this? A) Overfitting B) Multicollinearity C) Underfitting D) Outliers Q10 to Q15 are subjective answer type questions, Answer them briefly.
17. In which situation One-hot encoding must be avoided? Which encoding technique can be used in such a case?
18. In case of data imbalance problem in classification, what techniques can be used to balance the dataset? Explain them briefly.
19. What is the difference between SMOTE and ADASYN sampling techniques?
20. What is the purpose of using GridSearchCV? Is it preferable to use in case of large datasets? Why or why not?
21. List down some of the evaluation metric used to evaluate a regression model. Explain each of them in brief.

Q1) What is the advantage of hierarchical clustering over K-means clustering?

A) Hierarchical clustering is computationally less expensive B) In hierarchical clustering you don't need to assign number of clusters in beginning C) Both are equally proficient D) None of these.

ANSWER: D) None of these.

Q2) Which of the following hyper parameter(s),when increased

may cause randomforest to over fit the data?

A) max_depth B) n_estimators C) min_samples_leaf D) min_samples_splits

ANSWER: A) max_depth

Q3) Which of the following is the least preferable resampling method in handling imbalance datasets?

A) SMOTE B) RandomOverSampler C) RandomUnderSampler D) ADASYN

ANSWER: A) SMOTE

Q4) Which of the following statements is/are true about “Type-1” and “Type-2” errors?

1. Type1 is known as false positive and Type2 is known as false negative.
2. Type1 is known as false negative and Type2 is known as false positive.
3. Type1 error occurs when we reject a null hypothesis when it is actuallytrue. A) 1 and 2 B) 1 only C) 1 and 3 D) 2 and 3

ANSWER: D) 1 only.

Q5)Arrange the steps of k-means algorithm in the order in which they occur:

1. Randomly selecting the cluster centroids
2. Updating the cluster centroids iteratively
3. Assigning the cluster points to their nearest center A) 3-1-2 B) 2-1-3 C) 3-2-1 D) 1-3-2

ANSWER: A) 3-1-2

Q6) Which of the following algorithms is not advisable to use when you have limited CPU resources and time, and when the data set is relatively large?

A) Decision Trees B) Support Vector Machines C) K-Nearest Neighbors D) Logistic Regression

ANSWER: D) LOGISTIC REGRESSION

Q7) What is the main difference between CART (Classification and Regression Trees) and CHAID (Chi Square Automatic Interaction Detection) Trees?

A) CART is used for classification, and CHAID is used for regression. B) CART can create multiway trees (more than two children for a node), and CHAID can only create binary trees (a maximum of two children for a node). C) CART can only create binary trees (a maximum of two children for a node), and CHAID can create multiway trees (more than two children for a node) D) None of the above

ANSWER: C) CART can only create binary trees(a maximum of two children for a node), and CHAID can create multiway trees(more than two children for a node)

Q8) In Ridge and Lasso regularization if you take a large value of regularization constant(lambda), which of the following things may occur?

A) Ridge will lead to some of the coefficients to be very close to 0 B) Lasso will lead to some of the coefficients to be very close to 0 C) Ridge will cause some of the coefficients to become 0 D) Lasso will cause some of the coefficients to become 0.

ANSWER:

A) Ridge will lead to some of the coefficients to be very close to 0 D) Lasso will cause some of the coefficients to become 0.

Q9) Which of the following methods can be used to treat two multi-collinear features?

A) remove both features from the dataset B) remove only one of the features C) Use ridge regularization D) use Lasso regularization

ANSWER: C) Use ridge regularization D) use Lasso regularization

Q10) After using linear regression, we find that the bias is very low, while the variance is very high. What are the possible reasons for this?

A) Overfitting B) Multicollinearity C) Underfitting D) Outliers

ANSWER: A) Overfitting
C) Underfitting

Q11) In which situation One-hot encoding must be avoided? Which encoding technique can be used in

such a case?

ANSWER:

One-Hot-Encoding has the advantage that the result is binary rather than ordinal and that everything sits in an orthogonal vector space. The disadvantage is that for high cardinality, the feature space can really blow up quickly and you start fighting with the curse of dimensionality. Also for categorical variables where ordinal relationship exists, the one hot encoding is not enough. We have to use Label Encoder for ordinal data

Q12) In case of data imbalance problem in classification, what techniques can be used to balance the dataset? Explain them briefly.

ANSWER: An imbalanced classification problem is an example of a classification problem where the distribution of examples across the known classes is biased or skewed. Imbalanced classifications pose a challenge for predictive modeling as most of the machine learning algorithms used for classification were designed around the assumption of an equal number of examples for each class. This results in models that have poor predictive performance, specifically for the minority class. Two approaches to make a balanced dataset out of an imbalanced one are under-sampling and over-sampling.

1) Under-sampling Under-sampling balances the dataset by reducing the size of the abundant class. This method is used when quantity of data is sufficient. By keeping all samples in the rare class and randomly selecting an equal number of samples in the abundant class, a balanced new dataset can be retrieved for further modelling.

2) Over-sampling On the contrary, oversampling is used when the quantity of data is insufficient. It tries to balance dataset by increasing the size of rare samples. Rather than getting rid of abundant samples, new rare samples are generated by using e.g. repetition, bootstrapping or SMOTE (Synthetic Minority Over-Sampling Technique).

3) Cluster-Based Over Sampling In this case, the K-means clustering algorithm is independently applied to minority and majority class instances. This is to identify clusters in the dataset. Subsequently, each cluster is oversampled such that all clusters of the same class have an equal number of instances and all classes have the same size.

4) Modified synthetic minority oversampling technique (MSMOTE) for imbalanced data It is a modified version of SMOTE. SMOTE does not consider the underlying distribution of the minority class and latent noises in the dataset. To improve the performance of SMOTE a modified method MSMOTE is used.

Q13) What is the difference between SMOTE and ADASYN sampling techniques?

ANSWER: SMOTE: Synthetic Minority Over sampling Technique (SMOTE) algorithm applies KNN approach where it selects K nearest

neighbors, joins them and creates the synthetic samples in the space. The algorithm takes the feature vectors and its nearest neighbors, computes the distance between these vectors. The difference is multiplied by random number between (0, 1) and it is added back to feature. SMOTE algorithm is a pioneer algorithm and many other algorithms are derived from SMOTE.

ADASYN: Adaptive Synthetic (ADASYN) is based on the idea of adaptively generating minority data samples according to their distributions using K nearest neighbor. The algorithm adaptively updates the distribution and there are no assumptions made for the underlying distribution of the data.
The algorithm uses Euclidean distance for KNN Algorithm. The key difference between ADASYN and SMOTE is that the former uses a density distribution, as a criterion to automatically decide the number of synthetic samples that must be generated for each minority sample by adaptively changing the weights of the different minority samples to compensate for the skewed distributions. The latter generates the same number of synthetic samples for each original minority sample.

Q14) What is the purpose of using GridSearchCV? Is it preferable to use in case of large datasets? Why or why not?

ANSWER: Grid search is the process of performing hyper parameter tuning in order to determine the optimal values for a given model. There are libraries that have been implemented, such as GridSearchCV of the sklearn library, in order to automate this process. Grid Search can be thought of as an exhaustive search for selecting a model. In Grid Search, the data scientist sets up a grid of hyperparameter values and for each combination, trains a model and scores on the testing data. In this approach, every combination of hyperparameter values is tried and when running it on larger dataset can be very inefficient.

For example, searching 20 different parameter values for each of 4 parameters will require 160,000 trials of cross-validation. This equates to 1,600,000 model fits and 1,600,000 predictions if 10-fold cross validation is used. While Scikit Learn offers the GridSearchCV function to simplify the process, it would be an extremely costly execution both in computing power and time.

Q15) List down some of the evaluation metric used to evaluate a regression model. Explain each of them

in brief.

ANSWER: There are three main errors (metrics) used to evaluate models, Mean absolute error, Mean Squared error and R2 score.

Mean Absolute Error (MAE): Lets take an example where we have some points. We have a line that fits those points.When we do a summation of the absolute value distance from the points to the line, we get Mean absolute error.The problem with this metric is that it is not differentiable.

Mean Squared Error (MSE): Mean Squared Error solves differentiability problem of the MAE. Consider the same diagram above. We have a line that fits those points. When we do a summation of the square of distances from the points to the line, we get Mean squared error.

R2 Score: R2 score answers the question that if this simple model has a larger error than the linear regression model. However, it terms of metrics the answer we need is how much larger. The R2 score answers this question. R2 score is 1 — (Error from Linear Regression Model/Simple average model).

Best possible score is 1.0 and it can be negative (because the model can be arbitrarily worse). A constant model that always predicts the expected value of y, disregarding the input features, would get a R^2 score of 0.0.

In [ ]: