

# **FLIP ROBO TECHNOLOGIES, INC.**

## **Worksheet Set 1**

**Fernando Rosales  
andresrosales7@gmail.com  
Internship 28**

**May 2022**

## Python- Worksheet 1

**1. Which of the following operators is used to calculate remainder in a division?**

C) % (Modulo Operator)

**2. In python 2//3 is equal to?**

B) 0. "It is for FLOOR DIVISION. It is only the integer part of the value after division excluding the fractional or decimal part.

for example:

$7/2=3.5$

whereas,

$7//2=3$ " (Kumar, 2021)

**3. In python, 6<<2 is equal to?**

C) 24.

They are bit shift operator which exists in many mainstream programming languages, << is the left shift and >> is the right shift, they can be demonstrated as the following table, assume an integer only take 1 byte in memory.

operate	bit value	octal value	description
-----	-----	-----	-----
	00000100	4	
4 << 2	00010000	16	move all bits to left 2 bits, filled with 0 at the right

(Tsingyi, 2020)

**4. In python, 6&2 will give which of the following as output?**

A) 2. & -AND -Sets each bit to 1 if both bits are 1 (W3SCHOOLSa)

**5. In python, 6|2 will give which of the following as output?**

A) 2. |-OR-Sets each bit to 1 if one of two bits is 1 (W3SCHOOLSa)

**6. What does the finally keyword denotes in python?**

C) The finally block will be executed no matter if the try block raises an error or not.

Example:

```

try:
    x > 3
except:
    print("Something went wrong")
else:
    print("Nothing went wrong")
finally:
    print("The try...except block is finished") (W3SCHOOLSb)

```

## 7. What does raise keyword is used for in python?

A) It is used to raise an exception.

Python raise Keyword is used to raise exceptions or errors. The raise keyword raises an error and stops the control flow of the program. It is used to bring up the current exception in an exception handler so that it can be handled further up the call stack.

Syntax of the raise keyword:

```
raise {name_of_the_exception_class} (GeeksforGeeks, 2021a)
```

## 8. Which of the following is a common use case of yield keyword in python?

C) in defining a generator.

Yield is a keyword in Python that is used to return from a function without destroying the states of its local variable and when the function is called, the execution starts from the last yield statement. Any function that contains a yield keyword is termed a generator. Hence, yield is what makes a generator.

```

# Python3 code to demonstrate
# yield keyword

```

```

# generator to print even numbers
def print_even(test_list) :
    for i in test_list:
        if i % 2 == 0:
            yield i

```

```

# initializing list
test_list = [1, 4, 5, 6, 7]

```

```
# printing initial list
print ("The original list is : " + str(test_list))

# printing even numbers
print ("The even numbers in list are : ", end = " ")
for j in print_even(test_list):
    print (j, end = " ") (GeeksforGeeks, 2021b)
```

9. Which of the following are the valid variable names?

D) None of the above. According to the PEP 8 (2013, as cited in Finer 2020) “use a lowercase single letter, word, or words. Separate words with underscores to improve readability.”  
Examples: x, var, my\_variable

10. Which of the following are the keywords in python?

A) and B)

**For the programming questions please check the *Worksheet Set 1 Jupyter Notebook* available at <https://github.com/fernandorn35/Internship/tree/main/Worksheet%20Set%201>**

1. *Bernoulli random variables take (only) the values 1 and 0.*

**a) True.** A Bernoulli experiment has only two outcomes, the coin flipping is actually the canonical example, the simplest random variable since it only takes two values Jarad Niemi (2020).

2. *Which of the following theorem states that the distribution of averages of iid<sup>1</sup> variables, properly normalized, becomes that of a standard normal as the sample size increases?*

**a) Central Limit Theorem.** “Is a statistical premise that, given a sufficiently large sample size from a population with a finite level of variance, the mean of all sampled variables from the same population will be approximately equal to the mean of the whole population”, Ganti (2022).

3. **Which of the following is incorrect with respect to use of Poisson distribution?**

4. **Point out the correct statement.**

5. \_\_\_\_\_ random variables are used to model rates.

c) Poisson. Following Hayes, “in statistics, a Poisson distribution is a probability distribution that is used to show how many times an event is likely to occur over a specified period. In other words, it is a count distribution.”(2022)

6. Usually replacing the standard error by its estimated value does change the CLT.

7. Which of the following testing is concerned with making decisions using data  
b)Hypothesis

8. Normalized data are centered at\_\_\_\_\_and have units equal to standard deviations of the original data.

---

<sup>1</sup>See ZACH (2021, September 21). What Are i.i.d. Random Variables? (Definition & Examples) Statology.  
<https://www.statology.org/i-i-d-random-variables/>

a) 0

9. Which of the following statement is incorrect with respect to outliers?

d) None of the mentioned

**10. What do you understand by the term Normal Distribution?**

*In ideal conditions*, the data distribution should be within the Bell Shape Curve pattern, and as the central axis of this ideal distribution all the measures of central tendency: mode, median and mean.

**11. How do you handle missing data? What imputation techniques do you recommend?**

First pulling up some summary statistics would give a general idea of what's missing overall, a subset of missing data features would be the next step if there is a considerable number of columns on the dataset. Next step would be checking the data documentation, mistyping errors are common, specially with categorical variables. Mean imputation for numerical features if possible would be a must in some cases.

**12. What is A/B testing?**

According to Elles(2019), A/B testing refers to an hypothesis test, typically used in online marketing for comparing click rate (binary response) of two comparable pages.

**13. Is mean imputation of missing data acceptable practice?**

That would depend entirely on the nature of the feature itself, if the variable at hand is numerical, and of the continuous subcategory, it would be reasonable to do it. But doing this might come at a cost, Mean imputation does not preserve the relationship among variables, and Mean Imputation leads to an underestimate of Standard Errors, according to Grace-Martin (2020).

**14. What is linear regression in statistics?**

Linear regression is a basic and commonly used type of predictive analysis.

The overall idea of regression is to examine two things: (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable? (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they—indicated by the magnitude and sign of the beta estimates—impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables.(Statistics Solutions)

***15. What are the various branches of statistics?***

Descriptive and Inferential Statistics.

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

A) Least Square Error.

2. Which of the following statement is true about outliers in linear regression?

A) Linear regression is sensitive to outliers. "Outliers refer to unusual or extreme values within a data set" (pharmacy180)

3. A line falls from left to right if a slope is \_\_\_\_\_?

B) Negative.

4. Which of the following will have symmetric relation between dependent variable and independent variable?

B) Correlation

5. Which of the following is the reason for over fitting condition?

C) Low bias and high variance. "In supervised learning, overfitting happens when the model captures the noise along with the underlying pattern in data. It happens when model is trained a lot over noisy dataset. These models have low bias and high variance.", as explained by Singh (2018)

6. If output involves label then that model is called as:

B) Prediction Model

7. Lasso and Ridge regression techniques belong to \_\_\_\_\_?

D) Regularization

8. To overcome with imbalance dataset which technique can be used?

C) SMOTE. "Is one of the most commonly used oversampling methods to solve the imbalance problem. It aims to balance class distribution by randomly increasing minority



class examples by replicating them. SMOTE synthesises new minority instances between existing minority instances” (GeeksforGeeks, 2021c)

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses \_\_\_\_\_ to make graph?

A) TPR and FPR. As argues Bhandari:

The AUC-ROC curve helps visualizing how well a machine learning classifier is performing, only for binary classification problems. The Receiver Operator Characteristic (ROC) curve is an evaluation metric for binary classification problems. It is a probability curve that plots the TPR against FPR at various threshold values and essentially separates the ‘signal’ from the ‘noise’. The Area Under the Curve (AUC) is the measure of the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve(2020)

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

B) False. According to Klosterman:

The ROC curve of an effective classifier would hug the upper left corner of the plot: high TPR, low FPR. You can imagine that a perfect classifier would get a TPR of 1 (recovers all the positive samples) and an FPR of 0 and appear as a sort of square starting at (0,0), going up to (0,1), and finishing at (1,1). While in practice this kind of performance is highly unlikely, it gives us a limiting case.

Further consider what the area under the curve (AUC) of such a classifier would be, remembering integrals from calculus if you have studied it. The AUC of a perfect classifier would be 1, because the shape of the curve would be a square on the unit interval [0, 1]. (p. 106, 2021)

11. Pick the feature extraction from below (?)

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

B) and C)

13. Explain the term regularization?

Regarding the regularization, Gupta (2020) states a few points worth considering:

One of the major aspects of training your machine learning model is avoiding overfitting. The model will have a low accuracy if it is overfitting. This happens because your model is trying too hard to capture the noise in your training dataset. By noise we mean the data points that don't really represent the true properties of your data, but random chance. Learning such data points, makes your model more flexible, at the risk of overfitting.

In other words, processing the dataset into a manageable and useful material for a model.

**14. Which particular algorithms are used for regularization?**

Ridge Regression and Lasso Regression.

**15. Explain the term error present in linear regression equation?**

According to Damon (2021), when making predictions its necessary to fit it having in consideration the prediction inaccuracy, having this measure, inaccuracy can be reduced algorithmically. He also states "this function that quantifies this error is called the Cost Function, and typically the mean squared error is used. The mean squared error is as the name suggests, the mean (average) of the square of all the errors."

## **References**

- Bhandari, A. (2020) AUC-ROC Curve in Machine Learning Clearly Explained. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2020/06/auc-roc-curve-machine-learning/>
- Damon, L. (2021). Build your own AI INVESTOR (2<sup>nd</sup> ed.).ISBN: 978-1-8381322-3-1
- Finer, J. (2020) How to Write Beautiful Python Code With PEP 8. Real Python. <https://realpython.com/python-pep8/>
- Ganti, A. (2022, March 06) Central Limit Theorem (CLT). Investopedia. [https://www.investopedia.com/terms/c/central\\_limit\\_theorem.asp](https://www.investopedia.com/terms/c/central_limit_theorem.asp)
- GeeksforGeeks. Python Yield Keyword (2021a). <https://www.geeksforgeeks.org/python-yield-keyword/#:~:text=Yield%20is%20a%20keyword%20in,is%20what%20makes%20a%20generator.>
- GeeksforGeeks. Python Raise Keyword (2021b). <https://www.geeksforgeeks.org/python-raise-keyword/#:~:text=Python%20raise%20Keyword%20is%20used,further%20up%20the%20call%20stack.>
- GeeksforGeeks. ML | Handling Imbalanced Data with SMOTE and Near Miss Algorithm in Python (2021c). [https://www.geeksforgeeks.org/ml-handling-imbalanced-data-with-smote-and-near-miss-algorithm-in-python/#:~:text=SMOTE%20\(synthetic%20minority%20oversampling%20technique\)%20is%20one%20of%20the%20most,instances%20between%20existing%20minority%20instances.](https://www.geeksforgeeks.org/ml-handling-imbalanced-data-with-smote-and-near-miss-algorithm-in-python/#:~:text=SMOTE%20(synthetic%20minority%20oversampling%20technique)%20is%20one%20of%20the%20most,instances%20between%20existing%20minority%20instances.)
- Grace-Martin, K. (2020) Missing Data: Two Big Problems with Mean Imputation. The Analysis Factor. <https://www.theanalysisfactor.com/mean-imputation/>
- Gupta, P. (2017) Regularization in Machine Learning. Towards Data Science. <https://towardsdatascience.com/regularization-in-machine-learning-76441ddcf99a>
- Hayes, A. (2022, May 19) Poisson Distribution. Investopedia. <https://www.investopedia.com/terms/p/poisson-distribution.asp>

- Jarad Niemi. (2020). P2-3 Discrete random variables - part 3/5 (Bernoulli random variables) [Video]. YouTube.  
<https://www.youtube.com/watch?v=NXUkzZhrrcA&t=0s>
- Klosterman, S. (2021). Data Science Projects with Python (2nd ed.). Packt Publishing. ISBN: 978-1-80056-448-0
- Kumar, H. (2021). It is for FLOOR DIVISION. [Comment on the post "What does // mean in python?"]. Quora. <https://www.quora.com/What-does-mean-in-python-13>
- Monroe, W. (2017). Lecture Notes #7: Bernoulli and Binomial Random Variables. Stanford University CS 109:  
<https://web.stanford.edu/class/archive/cs/cs109/cs109.1178/lectureHandouts/070-bernoulli-binomial.pdf>
- Pharmacy180. Sensitivity to Outliers, Outlier Rejection, and Robust Regression. <https://www.pharmacy180.com/>. <https://www.pharmacy180.com/article/sensitivity-to-outliers,-outlier-rejection,-and-robust-regression-2986/>
- –Tsingyi. (2020, Jan 9).What do >> and << mean in Python? (Answer). Stack Overflow. <https://stackoverflow.com/questions/22832615/what-do-and-mean-in-python>
- Singh, S. (2018, May 21) Understanding the Bias-Variance Tradeoff. <https://towardsdatascience.com/>. <https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229>
- Statistics Solutions. What is Linear Regression?. <https://www.statisticssolutions.com/>. <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/what-is-linear-regression/>
- W3SCHOOLSa. Python Operators. <https://www.w3schools.com>. [https://www.w3schools.com/python/python\\_operators.asp](https://www.w3schools.com/python/python_operators.asp)
- W3SCHOOLSb. Python finally Keywords. <https://www.w3schools.com>. [https://www.w3schools.com/python/python\\_operators.asphttps://www.w3schools.com/python/ref\\_keyword\\_finally.asp#:~:text=Definition%20and%20Usage&text=It%20defines%20a%20block%20of,objects%20and%20clean%20up%20resources.](https://www.w3schools.com/python/python_operators.asphttps://www.w3schools.com/python/ref_keyword_finally.asp#:~:text=Definition%20and%20Usage&text=It%20defines%20a%20block%20of,objects%20and%20clean%20up%20resources.)