**Machine Learning subjective question**

**11. In which situation One-hot encoding must be avoided? Which encoding technique can be used in such a case?**

When the categorical feature is ordinal ,one hot encoding must be avoided & label encoding should be used.

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

**Under Sampling →** Random Under sampling aims to balance class distribution by randomly eliminating majority class examples. This is done until majority and minority class instances are balanced out.

**Over Sampling →** Over-Sampling increases the number of instances in the minority class by randomly replicating them in order to present a higher representation of the minority class in the sample.

SMOTE (Synthetic Minority Over-sampling Technique)

This technique is followed to avoid overfitting which occurs when exact replicas of minority instances are added to the main dataset.

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.

**13. What is the difference between SMOTE and ADASYN sampling techniques?**

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

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

GridSearchCV is used to find the best parameters out of given parameters on the best. So it is a method of tuning the parameters of given estimator. Since GridSearchCV, takes into account each and every parameter combination to check for the best result therefore in case of large datasets it will be very time taking task to get best parameters using Grid search instead Random Search will reduce the time complexity since it takes random combinations of the parameters. It is also used for Over fitting and under fitting models to give a best output with accuracy.

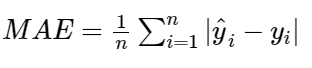
**15. List down some of the evaluation metric used to evaluate a regression model. Explain each of them in brief**

* R Square/Adjusted R Square.: This metric represents the part of the variance of the dependent variable explained by the independent variables of the model. It measures the strength of the relationship between your model and the dependent variable.
* Mean Square Error(MSE): It is the average of the squared difference between the predicted and actual value.

https://miro.medium.com/max/394/1*aLt5bWtuBr_V7TrBILe3qQ.png

MSE penalizes large errors.

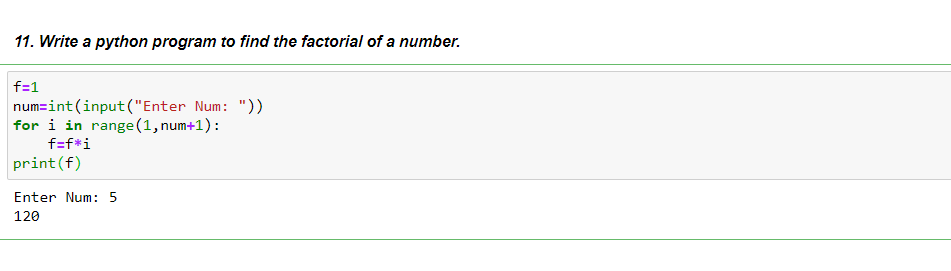
* Root Mean Square Error(RMSE): This is the square root of the average of the squared difference of the predicted and actual value.
* Mean Absolute Error(MAE): This is simply the average of the absolute difference between the target value and the value predicted by the model



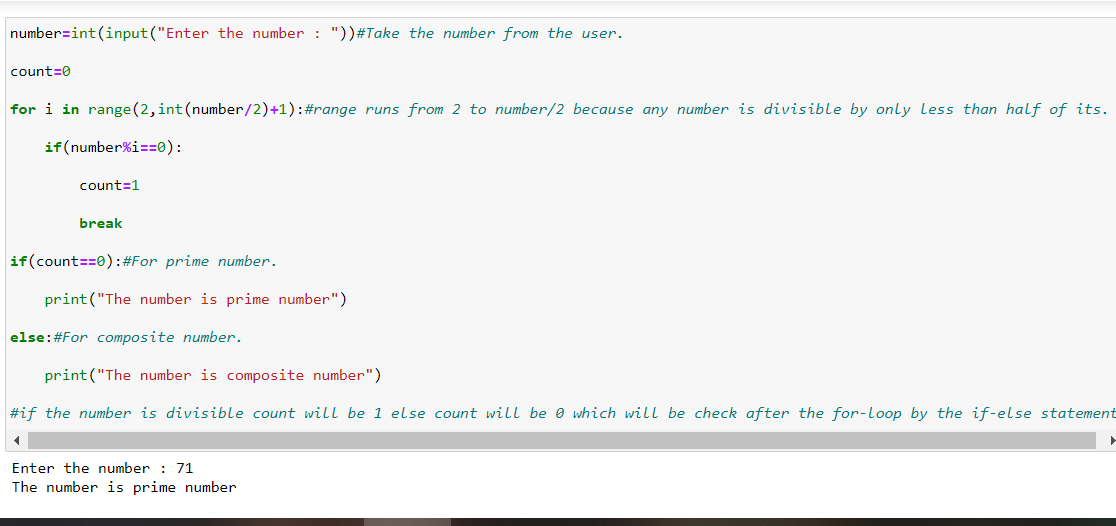
MAE does not penalize large errors.

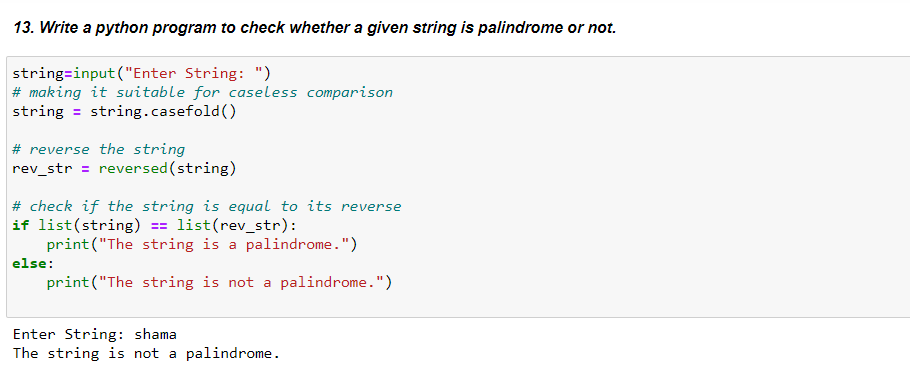
**Python subjective questions**

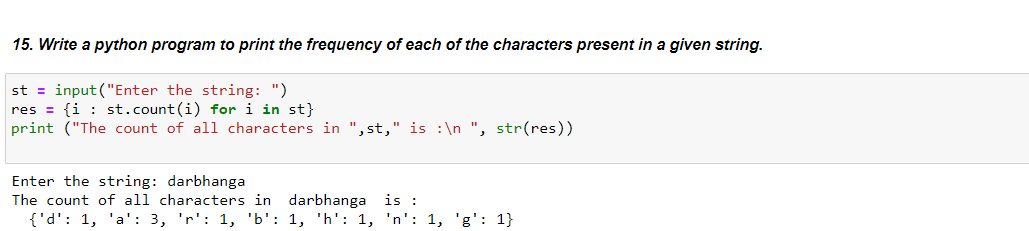
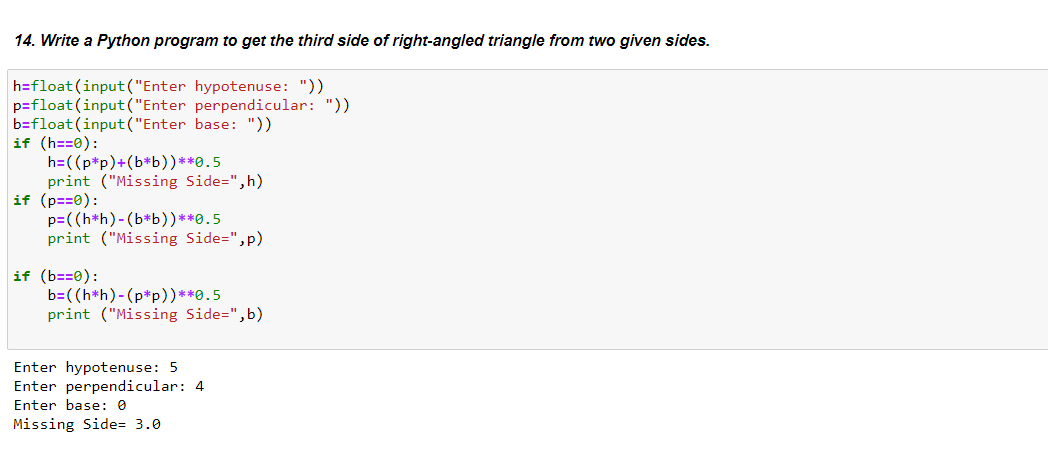
11. Write a python program to find the factorial of a number.

****

12. Write a python program to find whether a number is prime or composite.







**Statistics subjective questions**

**13. What is Anova in SPSS?**

Analysis of Variance, i.e. ANOVA in SPSS, is used for examining the differences in the mean values of the dependent variable associated with the effect of the controlled independent variables, after taking into account the influence of the uncontrolled independent variables.

**14. What are the assumptions of Anova?**

**Assumptions for ANOVA**

* Each group sample is drawn from a normally distributed population.
* All populations have a common variance.
* All samples are drawn independently of each other.
* Within each sample, the observations are sampled randomly and independently of each other.
* Factor effects are additive.

**15. What is the difference between one way Anova and two way Anova?**

A one-way ANOVA only involves one factor or independent variable, whereas there are two independent variables in a two-way ANOVA.

In a one-way ANOVA, the one factor or independent variable analyzed has three or more categorical groups. A two-way ANOVA instead compares multiple groups of two factors