Company Name - Legato
Role - Data Scientist

Q. Different activation function?

A. Binary Step Function, Linear Activation Function, Sigmoid/Logistic Activation Function, Tanh Function (Hyperbolic Tangent), ReLU Activation Function.

Q. How do you handle imbalance data?

A. Follow these techniques:

1. Use the right evaluation metrics.
2. Use K-fold Cross-Validation in the right way.
3. Ensemble different resampled datasets.
4. Resample with different ratios.
5. Cluster the abundant class.
6. Design your own models.

Q. Difference between sigmoid and softmax?

A. The sigmoid function is used for the two-class logistic regression, whereas the softmax function is used for the multiclass logistic regression (a.k.a. MaxEnt, multinomial logistic regression, softmax Regression, Maximum Entropy Classifier).

Q. Explain about optimizers?

A. Optimizers are algorithms or methods used to change the attributes of the neural network such as weights and learning rate to reduce the losses. Optimizers are used to solve optimization problems by minimizing the function.

Q. Precision-Recall Trade off ?

A. The Idea behind the precision-recall trade-off is that when a person changes the threshold for determining if a class is positive or negative it will tilt the scales. It means that it will cause precision to increase and recall to decrease, or vice versa.

Q. Decision Tree Parameters?

A. These are the parameters used for building Decision Tree: min_samples_split, min_samples_leaf, max_features and criterion.

Q. Tell us about Bagging and boosting?

A. Bagging is a way to decrease the variance in the prediction by generating additional data for training from dataset using combinations with repetitions to produce multi-sets of the original data. Boosting is an iterative technique which adjusts the weight of an observation based on the last classification.

Q. What are the different modules used in Python?

A. Modules refer to a file containing Python statements and definitions. We use modules to break down large programs into small manageable and organized files. Furthermore, modules provide reusability of code. Some of the modules used are:

1. **Numpy** - It is an amazing module for doing any kind of mathematical operations in Python. So essentially, it allows you to work with array-like objects of multiple dimensions like matrices and helps to do one, two, three dimensional math very fast. The operations performed in Numpy are fast because a lot of operations are implemented in C.
2. **Pandas** - It helps in reading and working with dataframes and just data in general. With pandas, it is easy to clean,manipulate and work with data.
3. **Regular expression(re)** - With **re** module, we can do more complex text processing using regular expression pattern matching.
4. **itertools** - This module provides various functions that work on iterators to produce complex iterators. This module works as a fast, memory-efficient tool that is used either by themselves or in combination to form **iterator algebra**.
5. **tkinter** - It comes bundled with Python, using Tk and is Python's standard GUI framework. It provides a fast and easy way to create GUI applications.

Q. What is Feature engineering?

A. Feature engineering is a machine learning technique that leverages data to create new variables that aren't in the training set. It can produce new features for both supervised and unsupervised learning, with the goal of simplifying and speeding up data transformations while also enhancing model accuracy.

Q. What is SelectK best? how does it works?

A. The SelectKBest method selects the features according to the k highest score. By changing the 'score_func' parameter we can apply the method for both classification and regression data. Selecting best features is important process when we prepare a large dataset for training. The SelectKBest class just scores the features using a function (in this case f_classif but could be others) and then "removes all but the k highest scoring features".

Q. Is outlier always need to be removed? What if. it is relevant to business?

A. Depending on the problem and dataset, we decide whether outliers are important or not. Thus, it is not necessary that Outliers need to be removed all the time because sometimes they provide important information, especially when it is relevant to business.

Q. Lambda function in python with example?

A. An anonymous function is known as a lambda function. This function can have any number of parameters but, can have just one statement.

Example: a = lambda x,y : x+y
print(a(5, 6))
Output: 11.

Q. Tricks you use to faster your model training?

A. Reduce Calculations by normalization or standardization. Use high computational memory while training (like GPU instead of local memory).

Q. Choose algorithm if your data has noise?

A. Algorithms like Probabilistic Random Forest and DBSCAN can work well when you've noise in data.

Q. Explain Deal overfitting?

A. Techniques to reduce overfitting:

1. Increase training data.
2. Reduce model complexity.
3. Early stopping during the training phase.
4. Ridge Regularization and Lasso Regularization.

NOTE:
If you want to learn Data Science from scratch and become interview ready, check out our popular Data science course.

https://ai.cloudyml.com/Learn-Data-Science-from-Scratch