Q: What are the assumptions associated with a linear regression model?

A1: There are four assumptions associated with a linear regression model:

Linearity: The relationship between X and the mean of Y is linear.  
Homoscedasticity: The variance of residual is the same for any value of X.  
Independence: Observations are independent of each other.  
Normality: For any fixed value of X, Y is normally distributed.

Q: What is Ordinary Least Squares in regression?

A1: Ordinary Least Squares regression (OLS) is a common technique for estimating coefficients of linear regression equations which describe the relationship between one or more independent variables and a dependent variable. In other words, you can consider the OLS as a strategy to obtain, from your model, a ‘straight line’ which is as close as possible to your data points. Even though OLS is not the only optimization strategy, it is the most popular for regression tasks.

A2: Ordinary least squares (OLS), estimates the parameters in a regression model by minimizing the sum of the squared residuals. This method draws a line through the data points that minimizes the sum of the squared differences between the observed values and the corresponding fitted values.

A3: Ordinary Least Squares regression (OLS) often called “Linear Regression”. It’s a common technique for estimating coefficients of linear regression equations which describe the relationship between one or more independent variables and a dependent variable.

A4: Bir tahmin yaparken eldeki verilerin doğrusal bir çizgi etrafında olmasını isteriz. Eldeki verilerin çizilecek olan bu doğruya olan uzaklıklarının karelerinin toplamının minimum olması daha sonra yapılacak tahminlerin daha doğru sonuç vermesini sağlayacaktır.

A5: Error is the difference between prediction and reality: the vertical distance between a real data point and the regression line. OLS is concerned with the squares of the errors. It tries to find the line going through the sample data that minimizes the sum of the squared errors.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Q: Günün Sorusu: What is bias in machine learning and why is it important?

A1: Bias can be best described as the difference between the actual prediction of our model to the correct value trying to predict. A model with high bias might pay very little attention to the training data and hence it rather oversimplifies the model. Such a model will always lead to high errors in training and test data. Bias is a phenomenon that completely skews the result of an algorithm in favor of or against an idea. Bias is considered to be a systematic error that occurs in the machine learning model automatically due to incorrect assumptions in the ML process.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

***If a human expert can easily create a pattern in his or her own mind, it is generally not worth the time and effort of using data science to “discover” it. In general, data science becomes useful when we have a large number of data examples and when the patterns are too complex for humans to discover and extract manually.*** As a lower bound, we can take a large number of data examples to be defined as more than a human expert can check easily. With regard to the complexity of the patterns, again, we can define it relative to human abilities. We humans are reasonably good at defining rules that check one, two, or even three attributes (also commonly referred to as features or variables), but when we go higher than three attributes, we can start to struggle to handle the interactions between them. By contrast, data science is often applied in contexts where we want to look for patterns among tens, hundreds, thousands, and, in extreme cases, millions of attributes.

*-- Kelleher, John D., and Brendan Tierney. Data science. MIT Press, 2018.*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Q: What is bias variance trade-off in machine learning?

A1: If our model is too simple and has very few parameters then it may have high bias and low variance. On the other hand if our model has large number of parameters then it’s going to have high variance and low bias. So we need to find the right/good balance without overfitting and underfitting the data.

Q: What are the metrics used to evaluate a Regression Model?

A1: Mainly, there are five metrics that are commonly used to evaluate the regressionmodels:Mean Absolute Error (MAE)Mean Squared Error (MSE) Root Mean Squared Error (RMSE) R-Squared (Coefficient of Determination)Adjusted R-Squared

Q: What are the flaws in R-squared?

A1: There are two major flaws of R-squared:

Problem- 1: As we are adding more and more predictors, R² always increases irrespective of the impact of the predictor on the model. As R² always increases and never decreases, it can always appear to be a better fit with the more independent variables(predictors) we add to the model. This can be completely misleading.

Problem- 2: Similarly, if our model has too many independent variables and too many high-order polynomials, we can also face the problem of over-fitting the data. Whenever the data is over-fitted, it can lead to a misleadingly high R² value which eventually can lead to misleading predictions.

Q: How is your DS journey going on? What do you think about preclass materials, classes, labs and daily questions?

Q: What is ROC curve?

A1: It stands for Receiver Operating Characteristic. It is basically a plot between a true positive rate and a false positive rate, and it helps us to find out the right tradeoff between the true positive rate and the false positive rate for different probability thresholds of the predicted values. So, the closer the curve to the upper left corner, the better the model is. In other words, whichever curve has greater area under it that would be the better model.

Q: What is recall (in statistics)?

A1 :

Q: How to calculate the accuracy of a binary classification algorithm using confusion matrix?

A1 : In a binary classification algorithm, we have only two labels, which are True and False. Before we can calculate the accuracy, we need to understand a few key terms:

• True positives: Number of observations correctly classified as True

• True negatives: Number of observations correctly classified as False  
• False positives: Number of observations incorrectly classified as True  
• False negatives: Number of observations incorrectly classified as False  
To calculate the accuracy, we need to divide the sum of the correctly classified observations by the number of total observations.

Q: Question of the day: How can we deal with outliers?

A: Below are some of the methods of treating the outliers:

* Trimming/removing the outlier
* Quantile based flooring and capping
* Mean/Median imputation

A : Outliers can be dealt with in several ways. One way is to drop them. We can only drop the outliers if they have values that are incorrect or extreme. In case the outliers are not that extreme, then we can try:  
A different kind of model. For example, if we were using a linear model, then we can choose a non-linear model Normalizing the data, which will shift the extreme values closer to other data points  
Using algorithms that are not so affected by outliers, such as random forest, etc.

 Q: Please explain how KNN works?

A : The value of k in the KNN algorithm is related to the error rate of the model. A small value of k could lead to overfitting as well as a big value of k can lead to underfitting.

Q:What are the main parameters of SVM algorithm and how do they affect model performance?

A: There are mainly 3 parameters in SVM. C- It is a hypermeter in SVM to control error. Low C means low error and large C means large error. Gamma is used when we use the Gaussian RBF kernel. Gamma high means more curvature. Gamma low means less curvature. And kernel enable us to change dimentionality.

Q:What is entropy in a decision tree algorithm?

A:In a decision tree algorithm, entropy is the measure of impurity or randomness. The entropy of a given dataset tells us how pure or impure the values of the dataset are. In simple terms, it tells us about the variance in the dataset.

Q:Question of the day: What is information gain in a decision tree algorithm?

A:When building a decision tree, at each step, we have to create a node that decides which feature we should use to split data, i.e., which feature would best separate our data so that we can make predictions. This decision is made using information gain, which is a measure of how much entropy is reduced when a particular feature is used to split the data. The feature that gives the highest information gain is the one that is chosen to split the data.

Q : What is bagging?

A: Bagging is an ensemble learning method. It stands for bootstrap aggregating. In this technique, we generate some data using the bootstrap method, in which we use an already existing dataset and generate multiple samples of the N size. This bootstrapped data is then used to train multiple models in parallel, which makes the bagging model more robust than a simple model. Once all the models are trained, when we have to make a prediction, we make predictions using all the trained models and then average the result in the case of regression, and for classification, we choose the result, generated by models, that has the highest frequency.

Q:What is ensemble learning?

A:When we are building models using Data Science and Machine Learning, our goal is to get a model that can understand the underlying trends in the training data and can make predictions or classifications with a high level of accuracy. However, sometimes some datasets are very complex, and it is difficult for one model to be able to grasp the underlying trends in these datasets. In such situations, we combine several individual models together to improve performance. This is what is called ensemble learning.

Q:What is boosting in ML?

A:Boosting is one of the ensemble learning methods. Unlike bagging, it is not a technique used to parallelly train our models. In boosting, we create multiple models and sequentially train them by combining weak models iteratively in a way that training a new model depends on the models trained before it. In doing so, we take the patterns learned by a previous model and test them on a dataset when training the new model. In each iteration, we give more importance to observations in the dataset that are incorrectly handled or predicted by previous models. Boosting is useful in reducing bias in models as well.

Q: How does KMeans Clustering work?

A:k-Means is just a clustering algorithm. It uses the distance between points as a measure of similarity. k-Means takes data points as input and groups them into k clusters. This process of grouping is the training phase of the learning algorithm. The result would be a model that takes a data sample as input and returns the cluster that the new data point belongs to, according the training that the model went through.

Q: Explain the difference between L1 and L2 regularization.

A: Where L1 regularization attempts to estimate the median of data, L2 regularization makes estimation for the mean of the data in order to evade overfitting.

L1 regularization can add the penalty term in cost function. On the other hand, L2 regularization appends the squared value of weights in the cost function.

Q: Question of the day: How does PCA work?

A: Principal component analysis (PCA) is a technique for reducing the dimensionality of such datasets, increasing interpretability but at the same time minimizing information loss. It does so by creating new uncorrelated variables that successively maximize variance.