Questions	Category	Answer
How can we deal with categorical data?	Data preprocessing	One hot encoding
How can we process missing datas in dataset?	Data preprocessing	mean, median and mode
What should be value of learning rate and why?	ML	it is not fixed varies problem to problem and it should start from 0.0001
What are outliers and why we need to remove them?	Data preprocessing	some features have a different value from all values it can be so much high or so much low so we have to take care all this values because it effect on our model
How can we find outliers?	Data visualization	using box plot diagram
Which graph is used for categorical data ?	Data visualization	sea born has catplot() which has multiple funtionalities for categorical data
Why boxplot is mainly used for?	Data visualization	To see how the data is distributed w.r.t mean,median and to check outliers
Which graph is use to represent continuous data and discrete data?	Data visualization	histogram can show the representation for both data
How do we find global minima if we have multiple local minima?	ML	We can give large iteration and store the values were min and search for min amoust them and use the weights
If a child have to go out to play and it depends on weather it's raining or not and also depend on days of week and his health.What type of learning it is	ML	Supervised: classification as the details have to be given earlier to train when can it go out
If an officer comes at office at different time on different days of week What type of learning it is	ML	If we're predicting the day of week using the past timings, it should be supervised classification
How normalization works in back end.	Data preprocessing	It's basic task is to bring data between 0 and 1 using min-max scalar is normalization
How to plot points for multi linear regression. Along with the regression line.	Data visualization	You'll have to use dimentionality reduction, feature extraction and extract less 1, 2 or columns to plot as we can max plot in 3 dimentions
Why accuracy remain constant while implementing algorithm ?	ML	As the values of weights are updated after calculating error on complete data we might not get sufficient error to manipulate our weight very large as a result the changes are very minute and does not show much affect on accuracy.
How many hidden layers we can take in Neural network(maximum)	DL	There is no such limit to number of layers you can add however more layers will slow down the training process. As good practise the number of hidden layers should be 2/3 the size of the input layer, plus the size of the output layer. The number of hidden neurons should be less than twice the size of the input layer.
How many nodes in each hidden layer?	DL	The number of hidden neurons should be less than twice the size of the input layer.
is it only deal with categorical data when implementing neural network?	DL	All the algorithms or models work just with numeric values you'll have to convert the categorical values in numeric using one hot encoding.
How to plot points for multi linear regression. Along with the regression line.	Data visualization	Already answered above
What is perceptron in Neural network?	ML	Perceptron is a single layer neural network and a multi-layer perceptron is called Neural Networks.
what is role of activation function in neural network	DL	Activation function decides the significance of features for prediction, so if the activation fuction calculates less values it shows that a feature is not much corelated to output
Why we set weight values randomly in NN?		So that all the nodes should not predict same things are all the weights in each node is multiplied by each weight if the weights are same then the values will also be predicted same on a;ll nodes.
Why we use backword elimination in NN?		accuracy metric is considered in neural network, In categorical one should consider the f1 score.
In regression Metrics, which one prefferd to find accuracy?		accuracy metric is considered in neural network, In categorical one should consider the f1 score.
Can we set "threshold" other than 0.5 in logistic regression ?		Yes we candefinitely set the threshold depending on the problem we're solving.
OR Is it necessary to take decision boundary as 0.5 for logistic regression ?		
What is OLS?		Linear regression's other name is Ordinary Least-Squares (OLS) Regression,
Do we need to apply feature scaling on dependent variable(y) ?		No it's not needed, just apply it on the independent variables and the same scaler should be used for testing data as well and then pass test data to model.
Do we need to fit and tranform dummy variables?		Dummy varibles value has to be one in the complete column and we pass the dataframe or 2d array consisiting dummy var to our model which adjusts the weights accordingly
Wha is rbf kernal in SVR?		**rbf : radial basis function kernel**
Why SVR need Feature scaling ?		feature scalaing is needed so that all the variables are in same scale and in some algo it' s done internally and in some we need to do it.
Why Polynomial regression called 'Linear"?		As we apply just a transformation on the single row and then we apply multi linear regression on the same
While preprocessing the dataset, we remove skewness only for dependent variable or for whole dataset?		We do it for the the independent data which is having skewness.
What is confusion matrix , how it is useful for Logistic regression?		Confusion matrix is a metric for categorical data which shows the right and wrong predictions done by the machine.
What is the difference between sigmoid and percepron?		Perceptron is a single layer neural network and a multi-layer perceptron is called Neural Networks while sigmoid is an activation function which can be used in a nodes or neural network. Perceptron can have any function. and sigmoid is just 1 of th function that can be used.
How to avoid overfitting using cross validation and stepwise regression?		When we use Kfold cross validation for training it uses random samples of data each time hence it gets various points and is not over fitted.
If you increase the number of hidden layers in a Multi Layer neural network, the error of data always decreases. True/False?		It does increases however it will overfit the data which increases the error in testing.
How to deal with mixed data type in neural network ?		When we pass the data to neural netwrok we have to pass it in numerical form, if categorical just do one hot encoding and it'll be handled.
How do I measure information loss when converting categorical data to numerical?		If you use dummy variables, you won't lose information, since you can reconstruct the categorical variable - without error - from the dummies.
How to preprocess the Categorical data with large num of columns ?		Just do one hot encoding on each column separately
how to remove Skewness with large num of columns ?		
How to deal when you have too many outliers?		We can set those to the 75% quartile, or if single row has many outliers we can remove that column
How can outlier value be treated?	Data preprocessing	Answered above
Is feature scaling mandatory for data preprocessing?	Data preprocessing	
does feature scaling affect accuracy?	Data preprocessing	Yes it does
What is weights initialization in neural networks?	DL	Just assigning random weights to each of the neurons is called weights initialization. Here we need to do it randomly and not same for all. here all the inputs are passed to each neuron so if all the weights are same they'll predict the same value.
Can we use linear regression for neural network?	DL	Yes we can however it won't capture complex relation ships as linear models are just linear.
How can overfitting recover in neural network?	DL	dropout, early stopping, regularization
Feature scaling and when to use which?		Feature scaling should be used whenever we have numerical data in differnt scaling.
How can underfitting recover in neural networks?	DL	Using more columns and using more hidden layers or even by increasing number of neurons in hidden layers.
Should we remove correlated columns before or after converting to dummy vars?		We can do it before converting to dummy vars as the correlation is shown only for numerical cols. And when we do one hot encoding we're already removing 1 columns from the number of cols for each column.
Is standerdization required before training logistic regression?	ML	It can faster the calculations when the numbers are small so it depends on you.
How the bais will change on high regularization?	ML	
Is neural network differentiate time complexity in single layer and multiple layer ?	DL	
Why is ReLU better and more often used than Sigmoid in Neural Networks?		As very low value of activation fucntion can be considered as they are not active hence storing negetive values is of no use and relu will conver negetive values to 0. Hence we consider relu mostly.
Why linear regression is flawed?	ML	As it models just linear relationship. And in real life things are not just linearly related.
Whar are effect of null values while training model?	ML	If we have null values it cuases error in calculations, as when calculating the mean we need the sum of all the numbers and the total number of rows. here the null value won't be considered in addition however it'll be considered while calculating the number of rows which gives us inappropriate mean.
When would you use gradient descent (GD) over stochastic gradient descent (SGD) ?		Gradient descent is used for whole dataset(batch processing), stochastic is used for online training
What are the population,sample,training set, design set, test set and validation set?	ML	
What are different data preprocessing techniques to handle outliers?		We can set those to the 75% quartile, or if single row has many outliers we can remove that column