Linear Regression

1. What are the assumptions of linear regression?
2. What is the functional form of the loss function that is minimized in linear regression?
3. When does the ordinary least square estimate of parameters become equal to the maximum likelihood estimate of parameters in linear regression?
4. Given a set of predictions from linear regression model and the ground truth as follows, can you find the value of the loss function

n_observation	Predicted	Actual
1	10	9
2	12	12
3	16	18
4	21	19
5	27	29
6	33	28
7	45	43
8	54	56

5. What is meant by the p-value of a distribution? Should you always take those coefficients whose p-values are less than 0.05?
6. In a linear regression output, you get a table as follows.

coefficient	mean	std deviation	p-value
pressure	0.01	0.002	0.02
temperature	1.5	0.13	0.017
humidity	2.5	1.91	0.23
wind_speed	0.2	0.17	0.34

    a. Which distribution does the p-values depicted above correspond to?
    b. Does the data need to be scaled before fitting a linear regression model? Justify
    c. What checks do you need to perform after fitting the linear regression model?
7. What is meant by R-squared value of a model? How is it different from adjusted R-squared?
8. Define AIC, BIC.
    a. How are these used to select the best models?
    b. Why are these quantities needed?
    c. How can you calculate the AIC or BIC of a decision tree?

Logistic Regression

1. Which probability distribution can be used to model a random variable that has only two outcomes? How is the probability distribution parameterized?
2. What do you mean by odds of an event? What is log-odds?

3. Which assumptions of linear regression are violated if we want to model a binary classification problem with it?
4. What is the sigmoid function? Mention its parameters
5. What is the shape of the decision boundary for a logistic regression?
6. What is the loss function for logistic regression?
7. Can you use regularization in logistic regression? How does the loss function change in that case?
8. How can you solve a multiclass classification problem using logistic regression?

Decision Trees

1. What is entropy? What is information gain in context of decision trees? What is Gini impurity?
2. What are the different types of algorithms that are used to train decision trees?
3. Briefly explain any one decision tree algorithm in the context of a classification problem
4. How does the decision tree treat categorical variables?
5. What is inductive bias in the context of different algorithms for fitting decision trees?
6. Why are decision trees not affected by outliers?
7. Why is pruning necessary for decision trees?
8. Is feature scaling necessary for decision trees?
9. What are the limitations of decision trees?

Random Forest

1. Why do we need to have random forests on top of decision trees?
2. What are the different ensemble methods that random forest uses to improve upon the decision tree algorithm?
3. What do you mean by bagging? What is feature subsampling?
4. Prove that in bagging, nearly $2/3^{rd}$ of the original dataset appears in the training bagged samples
5. What are the hyperparameters that you should tune for while fitting RF? Are there any heuristics to choosing them?

Clustering

1. Describe the algorithm of KMeans. How do you determine k?
2. What is the loss function for KMeans?
3. What are the drawbacks of KMeans? What are the workarounds?
4. Describe the algorithm of Hierarchical Agglomerative Clustering. Can categorical and continuous variables both be given as input? What about feature scaling?
5. What disadvantages does DBSCAN and GMM overcome with respect to clustering as compared to KMeans?