



## Linear Regression

1. State the assumptions in a linear regression model.
  2. Why do we often standardise features before applying regularised Linear Regression?
  3. Why is linear regression called “linear”?
  4. Why is the p-value important in regression coefficients?
  5. What happens if independent variables are highly correlated?
  6. If residuals show a funnel shape, which assumption is violated?
  7. Why is it important to split data into train/test sets?
  8. Suppose  $R^2 = 0.9$ . Can we say the model is good?
  9. Can Linear Regression be used for classification tasks? If yes, what's the drawback?
  10. Why do we assume errors are normally distributed in Linear Regression? What happens if this assumption is violated?
  11. Suppose you fit a Linear Regression model to predict salaries based on years of experience. The  $R^2$  is very high (0.95), but the residuals are not normally distributed. Is the model still valid?
- 

## Logistic Regression

12. What happens if classes in your dataset are highly imbalanced? How does it affect Logistic Regression?
13. What is the cost function used in Logistic Regression?
14. Logistic Regression outputs probabilities between 0 and 1. How does it achieve this transformation?
15. What is the interpretation of coefficients in Logistic Regression?



16. Logistic Regression uses Maximum Likelihood Estimation (MLE). Why not use the least squares method like in Linear Regression?
  17. Why is Logistic Regression considered more robust than Linear Regression for classification, especially with probabilities?
  18. Is Logistic Regression a generative or discriminative model? Why?
  19. Why is Logistic Regression sometimes called a one-layer neural network? What happens if we stack many such units?
- 

## **Probability, Bayes & Related**

20. Why is Bayes' theorem so fundamental in machine learning? Can you name two ML applications that directly rely on it?
21. Why does estimating a joint probability distribution become harder as the number of variables increases?
22. Why does estimating a joint probability distribution become harder as the number of variables increases? How does Naive Bayes handle this?
23. Naive Bayes assumes feature independence, which is almost never true. Why then does it still perform surprisingly well in text classification tasks?
24. Why is the Softmax function suitable for classification tasks but not regression?