Machine Learning

1. What is the effect of increasing or decreasing the Ridge parameter on the model's coefficients and predictions?

Answer:

Increasing the Ridge parameter leads to a reduction in the magnitude of the coefficients, as the regularization term penalizes large coefficients.

This reduction in the magnitude of the coefficients can reduce the test loss of the model but may also increase its training error.

2. **Can Ridge regression be used for feature selection, and if so, how?**

Answer:
Ridge regression is not typically used for feature selection, as it does not set coefficients to exactly zero. However, the magnitude of the coefficients is reduced, which can effectively shrink the coefficients of irrelevant features towards zero. As a result, Ridge regression can be used to identify and prioritize important features in a dataset.

3. **Can Lasso regression be used for feature selection, and if so, how?**

Yes, Lasso regression can be used for feature selection. Because the Lasso penalty term can set coefficients to exactly zero, it can effectively remove features that are not important for the model's prediction. The process of setting coefficients to zero is also called "shrinkage", which is why Lasso regression is sometimes referred to as a "shrinkage method".

The Lasso regression model is trained with different values of the regularization parameter (lambda), and the coefficients are estimated for each value of lambda. As the value of lambda increases, more coefficients are set to zero, resulting in a model with fewer features.

4. **Do we penalize intercept term in ridge and Lasso?**

Answer:

In Ridge and Lasso regression, the intercept term is not penalized, because it is not included in the regularization term. The regularization term only penalizes the sum of squared coefficients, which does not include the intercept.

The reason for this is that the intercept is usually not subject to the same issues of multicollinearity that affect the other coefficients. Penalizing the intercept in Ridge Lasso regression can lead to biased predictions and is generally not recommended.

5. **What are some limitations of Ridge and Lasso regression models?**

Answer:

- They can be sensitive to outliers: Ridge and Lasso regression are sensitive to outliers, which can have a strong influence on the model coefficients. Outliers can lead to biased estimates and poor performance.
- They require tuning of the regularization parameter: The performance of Ridge and Lasso regression models depends on the choice of the regularization parameter. Choosing an optimal value for the regularization parameter can be challenging and requires careful tuning.

6. **Can the Naive Bayes model handle missing data? If so, how?**

Answer:
Yes, the Naive Bayes model can handle missing data. In fact, the model can be quite robust to missing data, especially if the amount of missing data is small relative to the size of the dataset.

For Bernoulli Naive Bayes, which is used for binary data, the most common imputation method is to simply assume that the missing values are the same as the most frequent value of the corresponding feature in the training set.
For Gaussian Naive Bayes, which is used for continuous data, a common approach is to impute missing values with the mean or median of the corresponding feature values, or to use a regression model to predict the missing values based on the other features.

7. **How does the choice of distribution (e.g., Gaussian, Bernoulli, etc.) affect the performance of the Naive Bayes model?**

Answer:

The choice of distribution for the Naive Bayes model depends on the nature of the features and the type of problem being solved.

For example, if the features are continuous, the Gaussian distribution can be used to model the likelihood probabilities. If the features are binary (0/1), the Bernoulli distribution can be used. If the features are counts (i.e., integer-valued), the multinomial distribution can be used.

The performance of the Naive Bayes model can be affected by the choice of distribution if the distribution assumption does not hold true for the data. For example, if the features have a highly skewed distribution, a Gaussian distribution may not be appropriate and a different distribution (e.g., log-normal) may be more suitable. In such cases, the model may perform poorly due to violation of the distribution assumption. Therefore, it is important to choose an appropriate distribution based on the characteristics of the data to achieve optimal performance.

8.

Can the KNN algorithm be used for regression problems, and if so, how?

Answer:

Yes, the KNN algorithm can also be used for regression problems, and it is called the KNN regression algorithm or KNN regressor. In the KNN regression algorithm, the predicted value of the target variable is calculated as the average (or median) of the values of the k nearest neighbors.

Exercise: (For your own knowledge)

1. How feature selection is different from dimensionality reduction?
2. What is curse of dimensionility? How it affects the model?
3. How to handle missing values?
4. Does increasing the size of training data always help in improving the performance of the model?
5. Can L2 or L1 regualrization be used in decision trees? If yes, How?