Regularization quiz questions

Q.1 Which one of the following is true?

(A) Ridge and Lasso regression are methods to lessen the complexity of the model and avoid over-fitting that may be brought on by simple linear regression.

(B) Ridge regression causes the coefficients to decrease and helps in reducing multi-collinearity and model complexity.

(C) Lasso regression not only helps in reducing over-fitting but it can help us in feature selection

(D) All of the above

Correct answer: (D) - All of the above

Explanation:

(A) Ridge and Lasso regression both add a regularization term to the objective function to reduce the model complexity and prevent overfitting, which is a common problem with simple linear regression.

(B) Ridge regression, in addition to reducing overfitting, shrinks the coefficients towards zero, which helps to reduce the model complexity and multi-collinearity.

(C) Lasso regression, in addition to reducing overfitting, can also perform feature selection by shrinking the coefficients of irrelevant or redundant features to zero.

Q.2 What's the penalty term for the Lasso regression?

(A) the square of the magnitude of the coefficients

(B) the square root of the magnitude of the coefficients

(C) the absolute sum of the coefficients

(D) the sum of the coefficients

Correct answer: (C) - the absolute sum of the coefficients

Explanation:

Lasso regression adds a penalty term equal to the absolute sum of the coefficients to the objective function. This penalty term is also known as L1 regularization. The purpose of this penalty term is to shrink the coefficients of irrelevant or redundant features towards zero, effectively performing feature selection. By contrast, Ridge regression adds a penalty term equal to the square of the magnitude of the coefficients (L2 regularization).

Q.3 What's the penalty term for the Ridge regression?

(A) the square of the magnitude of the coefficients

(B) the square root of the magnitude of the coefficients

(C) the absolute sum of the coefficients

(D) the sum of the coefficients

Correct Answer: (A) - the square of the magnitude of the coefficients

Explanation:

Ridge regression adds a penalty term equal to the square of the magnitude of the coefficients to the objective function. This penalty term is also known as L2 regularization. The purpose of this penalty term is to shrink the coefficients towards zero, reducing the model complexity and preventing overfitting. The amount of regularization is controlled by a hyperparameter called the regularization parameter (lambda or alpha).

Q.4 What is the purpose of the regularization parameter in Ridge and Lasso regression?

(A) To control the strength of the penalty term

(B) To control the number of iterations of the algorithm

(C) To control the learning rate of the algorithm

(D) To control the batch size of the algorithm

Correct Answer: (A) - To control the strength of the penalty term

Explanation:

The purpose of the regularization parameter in Ridge and Lasso regression is to control the strength of the penalty term that is added to the cost function. This penalty term helps to reduce the overfitting of the model by shrinking the coefficients towards zero. A higher value of the regularization parameter results in stronger penalty, leading to more coefficients being shrunk towards zero, while a lower value of the regularization parameter results in a weaker penalty, allowing more coefficients to take on non-zero values. Therefore, the regularization parameter plays a critical role in determining the balance between bias and variance in the model, and it needs to be tuned carefully to achieve the best performance.

Q.5 Which of the following is a hyperparameter of the Ridge regression algorithm?

(A) The learning rate

(B) The number of iterations

(C) The regularization parameter

(D) The batch size

Correct Answer: (C) - The regularization parameter

Explanation:

In Ridge regression, the regularization parameter controls the amount of regularization applied to the model. This hyperparameter is typically denoted by λ and determines the trade-off between fitting the training data well and keeping the model weights small.

A higher value of λ results in stronger regularization, which can lead to a simpler model with smaller weights and less overfitting. Conversely, a lower value of λ results in weaker regularization, which can lead to a more complex model with larger weights and more overfitting.

Therefore, the regularization parameter is a crucial hyperparameter that needs to be tuned carefully to achieve good performance on the test data. It is not related to the learning rate, the number of iterations, or the batch size, which are hyperparameters of other machine learning algorithms such as gradient descent.

Q.6 Which of the following is a disadvantage of using regularization?

(A) It can improve the generalization performance of a model

(B) It can reduce the variance of a model

(C) It can make the model more interpretable

(D) It can introduce bias into the model

Correct Answer: (D) - It can introduce bias into the model

Explanation:

Regularization is a technique used to prevent overfitting in machine learning models by adding a penalty term to the loss function that encourages smaller model weights. This penalty term introduces a bias into the model, which can lead to a decrease in its overall accuracy. Therefore, the disadvantage of using regularization is that it can introduce bias into the model.

However, this bias is usually small compared to the variance reduction achieved by regularization. Regularization can help to improve the generalization performance of a model by reducing overfitting and improving its ability to make accurate predictions on new, unseen data. It can also make the model more interpretable by reducing the number of features used in the model and making the feature importance more transparent.

Therefore, while regularization can introduce some bias into the model, it is typically a small price to pay for the benefits it provides in terms of reducing overfitting and improving the generalization performance of the model.

Q.7 In Ridge regression, as the regularization parameter increases, do the regression coefficients decrease?

(A) True

(B) False

Correct Answer: (A) - True

Explanation: Ridge regression is a type of linear regression that adds a penalty term to the cost function. This penalty term is proportional to the square of the magnitude of the coefficients. The regularization parameter, λ, controls the strength of the penalty term.

When λ increases, the penalty term becomes more important and the optimization algorithm tends to reduce the magnitude of the coefficients. This is because the objective function (which includes both the loss function and the penalty term) is minimized when the coefficients have smaller magnitudes. As a result, the regression coefficients are shrunk towards zero, but not exactly to zero.

Therefore, we can say that as the regularization parameter increases in Ridge regression, the regression coefficients decrease.

Q.8 Ridge regression can reduce the slope close to zero (but not exactly zero) but Lasso regression can reduce the slope to be exactly equal to zero.

(A) Both statements are True about Ridge and Lasso.

(B) Both statements are False about Ridge and Lasso.

(C) True statement about Ridge but not about Lasso.

(D) True statement about Lasso but not about Ridge.

Correct Answer: (A) - Both statements are True about Ridge and Lasso

Explanation:

Ridge and Lasso are both regularization techniques used to address the issue of overfitting in linear regression. However, they differ in the way they apply the penalty term to the coefficients.

In Ridge regression, the penalty term is proportional to the square of the magnitude of the coefficients. This means that the coefficients are reduced, but not exactly to zero. As the regularization parameter increases, the coefficients are shrunk towards zero, but none of them become exactly zero.

On the other hand, in Lasso regression, the penalty term is proportional to the absolute value of the coefficients. This results in some of the coefficients becoming exactly zero as the regularization parameter increases. Therefore, Lasso regression can be used for feature selection, where it eliminates some of the less important features by setting their coefficients to zero.

Hence, both statements are true about Ridge and Lasso. Ridge can reduce the slope close to zero (but not exactly zero), whereas Lasso can reduce the slope to be exactly equal to zero.

Q.9 What is the main difference between L1 and L2 regularization?

(A) L1 regularization adds a penalty term proportional to the absolute value of the coefficients, while L2 regularization adds a penalty term proportional to the squared magnitude of the coefficients.

(B) L1 regularization adds a penalty term proportional to the squared magnitude of the coefficients, while L2 regularization adds a penalty term proportional to the absolute value of the coefficients.

(C) L1 regularization is used for sparse models, while L2 regularization is used for models with many non-zero coefficients.

(D) L2 regularization is used for sparse models, while L1 regularization is used for models with many non-zero coefficients.

(E) There is no difference between L1 and L2 regularization.

Correct Answer: (A) - L1 regularization adds a penalty term proportional to the absolute value of the coefficients, while L2 regularization adds a penalty term proportional to the squared magnitude of the coefficients.

Explanation:

L1 regularization adds a penalty term proportional to the absolute value of the coefficients, while L2 regularization adds a penalty term proportional to the squared magnitude of the coefficients. This leads to different effects on the coefficients and the resulting models. L1 regularization tends to produce sparse models with many zero coefficients, while L2 regularization tends to produce models with smaller but non-zero coefficients.

Q.10 How does regularization affect the variance and bias? Which of the following is true?

(A) Regularization can significantly reduce the variance of the model.

(B) Regularization can be used for feature selection.

(C) Shrinkage methods are used to regularize linear regression models

(D) Regularization is used to prevent fitting below the Bayes error.

(E) Regularization increases the bias of a model.

Correct Answer: (A),(B),(C),(E)

Explanation:

Regularization can effectively reduce the variance of a model by adding a penalty term to the loss function that discourages complex models. This helps to prevent overfitting and improve the generalization performance of the model. However, if the model already has low bias, regularization may lead to a slight increase in bias. This is because regularization adds a constraint to the model, which may limit its ability to fit the training data perfectly. Nevertheless, the overall impact of regularization on bias is usually minor, and the benefits of reduced variance often outweigh the small increase in bias.

Regularization is used to prevent fitting below the Bayes error. It is not possible to fit below the irreducible error.

Shrinkage methods are used to regularize linear regression models.

Shrinkage methods can be used to reduce regression coefficients to zero or near zero allowing one to eliminate them, therefor regularization can be used for feature selection.