

Support Vector Machines Assignment - 3

March 7, 2024

Q1. In order to predict house price based on several characteristics, such as location, square footage, number of bedrooms, etc., you are developing an SVM regression model. Which regression metric in this situation would be the best to employ?

In regression problems, Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are common evaluation metrics. They measure the average squared difference between the predicted values and the actual values.

In the context of predicting house prices, RMSE is a more suitable metric to use as it provides an interpretable measure of the average error in dollars. This is particularly important because the goal is to predict house prices accurately, and RMSE provides an intuitive understanding of how well the model is performing in that respect.

Another useful metric for regression problems is R-squared, which measures the proportion of the variance in the target variable that is explained by the model. However, in the case of SVM regression, where the model is optimized for minimizing the error, RMSE is more commonly used.

Q2. You have built an SVM regression model and are trying to decide between using MSE or R-squared as your evaluation metric. Which metric would be more appropriate if your goal is to predict the actual price of a house as accurately as possible?

If your goal is to predict the actual price of a house as accurately as possible, then Mean Squared Error (MSE) would be the more appropriate evaluation metric to use.

MSE measures the average squared difference between the predicted values and the actual values. In the context of predicting house prices, MSE provides an intuitive understanding of how far the predictions are, on average, from the actual prices in terms of dollars squared. This is important because the goal is to minimize the error in predicting house prices as much as possible.

On the other hand, R-squared measures the proportion of the variance in the target variable that is explained by the model. While R-squared can be a useful metric for understanding how well the model fits the data, it is not as directly related to the accuracy of individual predictions as MSE.

Q3. You have a dataset with a significant number of outliers and are trying to select an appropriate regression metric to use with your SVM model. Which metric would be the most appropriate in this scenario?

When dealing with a dataset that contains a significant number of outliers, Mean Absolute Error (MAE) is a more appropriate metric to use with an SVM regression model than Mean Squared Error (MSE).

The reason for this is that MSE is more sensitive to outliers since it squares the errors, making the contribution of large errors to the overall error much larger than that of smaller errors. On the other hand, MAE measures the absolute difference between the predicted values and the actual values, and is therefore less sensitive to outliers.

Using MAE as the evaluation metric for the SVM regression model would ensure that the model is evaluated based on its performance in predicting the actual price of the house, regardless of any outliers that may be present in the dataset.

Q4. You have built an SVM regression model using a polynomial kernel and are trying to select the best metric to evaluate its performance. You have calculated both MSE and RMSE and found that both values are very close. Which metric should you choose to use in this case?

When evaluating the performance of an SVM regression model that uses a polynomial kernel, if the values of Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are very close, then Root Mean Squared Error (RMSE) should be used as the evaluation metric.

The reason for this is that RMSE is more interpretable since it is expressed in the same units as the target variable. In the case of predicting house prices, RMSE would be expressed in dollars, which makes it easier to understand how well the model is performing in terms of its predictive accuracy.

Additionally, RMSE is less sensitive to the scale of the target variable than MSE. This means that even if the scale of the target variable changes, the RMSE values would still be comparable across different models.

Therefore, in the scenario where MSE and RMSE are very close for an SVM regression model with a polynomial kernel, using RMSE as the evaluation metric would be the better choice.

Q5. You are comparing the performance of different SVM regression models using different kernels (linear, polynomial, and RBF) and are trying to select the best evaluation metric. Which metric would be most appropriate if your goal is to measure how well the model explains the variance in the target variable?

When comparing the performance of different SVM regression models using different kernels, if the goal is to measure how well the model explains the variance in the target variable, then R-squared (R²) would be the most appropriate evaluation metric to use.

R-squared measures the proportion of the variance in the target variable that is explained by the model. A higher R-squared value indicates that the model is explaining more of the variance in the target variable and is therefore a better fit for the data.

Since the goal is to measure how well the model explains the variance in the target variable, R-squared would provide an intuitive understanding of how well each model is performing in that respect. It would also allow for easy comparison between the different SVM regression models, regardless of the kernel used.

Therefore, R-squared would be the most appropriate evaluation metric to use when comparing the performance of different SVM regression models using different kernels, with the goal of measuring how well the model explains the variance in the target variable.