

In [4]:

```
import pandas as pd
df = pd.read_csv('Bengaluru_House_Data.csv')
df
```

Out[4]:

	area_type	availability	location	size	society	total_sqft	bath	balcony
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0
...
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.0
13316	Super built-up Area	Ready To Move	Richards Town	4 BHK	NaN	3600	5.0	NaN
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.0
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCl	4689	4.0	1.0
13319	Super built-up Area	Ready To Move	Doddathoguru	1 BHK	NaN	550	1.0	1.0

13320 rows × 9 columns

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?

When developing an SVM regression model to predict house prices based on several characteristics, there are several regression metrics that can be used to evaluate the model's performance. Some of the commonly used regression metrics for SVM regression models include:

Mean squared error (MSE): The MSE measures the average squared difference between the predicted and actual values. It penalizes large errors more than smaller ones, which makes it a good metric for models where large errors are considered more important.

Mean absolute error (MAE): The MAE measures the average absolute difference between the predicted and actual values. It is less sensitive to outliers than MSE and can be a better metric if the distribution of errors is not normal.

R-squared (R2): The R2 measures the proportion of the variance in the target variable that is explained by the model. It ranges from 0 to 1, with higher values indicating a better fit.

In the case of predicting house prices, the Mean Squared Error (MSE) metric would be a good metric to employ. This is because predicting the correct house price is of paramount importance, and large errors in the predicted values should be penalized. Additionally, the MSE is widely used in the literature for regression tasks and is easy to interpret. However, it is always a good practice to evaluate the model using multiple metrics and choose the one that is most suitable for the specific use case.

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

MSE is a commonly used regression metric that measures the average squared difference between the predicted and actual values. It gives more weight to larger errors, which is particularly important when the goal is to minimize the difference between the predicted and actual values.

On the other hand, R-squared measures the proportion of variance in the target variable that is explained by the model. While it can be a useful metric for understanding how well the model fits the data, it may not be as directly related to the goal of predicting the actual price of a house as accurately as possible. Therefore, MSE is a better choice of evaluation metric in this case.

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, it is often more appropriate to use a regression metric that is less sensitive to outliers. One such metric is the Mean Absolute Error (MAE).

Unlike the Mean Squared Error (MSE), which gives more weight to larger errors, the MAE measures the average absolute difference between the predicted and actual values, regardless of their direction. This makes it more robust to outliers, as extreme values will have less of an impact on the overall score.

Therefore, MAE would be the most appropriate regression metric to use with an SVM model when dealing with a dataset that contains a significant number of outliers.

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 both Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are very close in value, it is generally recommended to use RMSE as the evaluation metric. The RMSE is a more interpretable metric than MSE because it is expressed in the same units as the target variable (i.e., the dependent variable). In contrast, MSE is expressed in squared units, which can make it harder to interpret the magnitude of the error.

Furthermore, since the RMSE takes the square root of the MSE, it is more sensitive to larger errors, which can be particularly important when dealing with regression models that have a polynomial kernel.

Polynomial kernels are known to be more prone to overfitting, and larger errors may indicate that the model

is not generalizing well to new data.

Therefore, in this case, RMSE should be the preferred evaluation metric to use for the SVM regression model with a polynomial kernel.

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?

If the goal is to measure how well the SVM regression models with different kernels explain the variance in the target variable, then R-squared (R^2) would be the most appropriate evaluation metric. R-squared is a commonly used metric that measures the proportion of the variance in the target variable that is explained by the model. It provides a measure of how well the model fits the data, with values ranging from 0 to 1. A higher R^2 value indicates a better fit, as it means that a larger proportion of the variance in the target variable is explained by the model.

When comparing the performance of different SVM regression models using different kernels (linear, polynomial, and RBF), R^2 would be a useful metric for selecting the model that provides the best explanation of the variance in the target variable.

In []: