Assignment

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?
https://drive.google.com/file/d/1Z9oLpmt6IDRNw7IeNcHYTGeJRYypRSC0/view?

usp=share_link

Ans:To predict house prices based on several characteristics using an SVM regression model, the choice of regression metric depends on the specific requirements and characteristics of the problem. Here are some common regression metrics that can be employed in this situation:

Mean Absolute Error (MAE):

- The MAE represents the average absolute difference between the predicted and actual values.
- MAE=1�∑�=1�|��−�^�|
- MAE=
  - n
  - 1
  -
- $\sum$
- $i=1$
- n
-
- $|y$
- i
-
  - $-$
  - y
  - ^
  -
- i
-
- $|$
- MAE is easy to interpret, providing a straightforward measure of prediction accuracy.

Mean Squared Error (MSE):

- The MSE represents the average squared difference between the predicted and actual values.
- MSE=1�∑�=1�(��−�^�)2

- MSE=
  - n
  - 1
  -
- $\sum$
- $i=1$
- n
-
- $(y$
- i
-
  - $-$
  - y
  - ^
  -
- i
-
- $)$
- 2
-
- MSE penalizes larger errors more than smaller ones and is widely used in regression problems.

Root Mean Squared Error (RMSE):

- The RMSE is the square root of the MSE and has the same unit as the target variable.
- RMSE=MSE
- RMSE=
- MSE
-
-
- RMSE provides a more interpretable measure of prediction error.

R-squared (R^2) Score:

- R^2 represents the proportion of the variance in the target variable that is predictable from the independent variables.
- $\text{R2}=1-\sum\text{=1}(\text{}-\text{^})2\sum\text{=1}(\text{}-\text{¯})2$
- R
- 2
- =1−
- ∑

- $i=1$
- n
- $(y$
- i
- $-$
- y
- $\bar{}$
- $)$
- 2
- $\sum$
- $i=1$
- n
- $(y$
- i
- $-$
- y
- $\hat{}$
- i
- $)$
- 2
- R² ranges from 0 to 1, where 1 indicates a perfect fit.

The choice of the best regression metric depends on the specific goals and requirements of your prediction task. In the context of house price prediction, Mean Squared Error (MSE) or Root Mean Squared Error (RMSE) is commonly used as they penalize larger errors, which may be more relevant in a real estate context where accurate pricing is crucial.

Here's a Python example using scikit-learn to develop an SVM regression model and evaluate its performance using RMSE:

python

Copy code

```
import        as
from                          import
from                        import
from             import
from                 import
import       as


      "https://drive.google.com/uc?id=1Z9oLpmt6IDRNw7IeNcHYTGeJRYypRSC0"


                  'price'
       'price'


                                                                    0.2
            42


                         'linear'




print f"Root Mean Squared Error (RMSE): {rmse}"
```

This example uses the scikit-learn library to develop an SVM regression model and calculate the Root Mean Squared Error (RMSE) as the evaluation metric. You can modify the code to use other regression metrics based on your specific requirements.

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?
Ans:Mean Squared Error (MSE) would be more appropriate as an evaluation metric. Here's why:

MSE (Mean Squared Error):

- MSE measures the average squared difference between the predicted and actual values.
- It penalizes larger errors more than smaller ones, making it sensitive to outliers.
- Minimizing MSE corresponds to finding the parameters that result in the smallest overall squared differences, which aligns with the goal of accurate prediction.

MSE=1$\sum$=1(−^)2

$$MSE=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

R-squared (R^2):

- R^2 represents the proportion of the variance in the target variable that is predictable from the independent variables.

- While R^2 is a useful metric for understanding the proportion of variance explained, it might not be as directly interpretable in terms of predicting the absolute values of house prices.
- R^2 is a relative measure that compares the performance of the model to a simple baseline model (mean prediction).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

For house price prediction, the focus is typically on minimizing the prediction errors in absolute terms, and MSE directly reflects the average squared difference between predicted and actual prices. It provides a clear indication of how well the model is performing in terms of accuracy.

In summary, MSE is more directly aligned with the goal of predicting actual house prices accurately, making it a more suitable metric for this specific regression task.

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?
Ans:In a scenario where you have a dataset with a significant number of outliers, the Mean Absolute Error (MAE) would be a more appropriate regression metric to use with your SVM model. Here's why:

Mean Absolute Error (MAE):

- MAE measures the average absolute difference between the predicted and actual values.
- It is less sensitive to outliers compared to Mean Squared Error (MSE) because it does not square the differences.
- Outliers have a linear impact on MAE, and extreme values do not disproportionately influence the metric.

MAE=1�∑�=1�|��−�^�|

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}$$

i

$|$

Mean Squared Error (MSE):

- MSE squares the differences between predicted and actual values, making it more sensitive to outliers.
- Outliers can have a disproportionate impact on MSE because the squared differences amplify the effect of large errors.

MSE=1$\hat{}$2

MSE=

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Huber Loss:

- The Huber loss is another option that combines the characteristics of both MAE and MSE. It is less sensitive to outliers than MSE and provides a compromise between MAE and MSE.

$$\begin{cases} \frac{1}{2}(y - f(x))^2 & \text{for } |y - f(x)| \leq \delta \\ \delta(|y - f(x)| - \frac{1}{2}\delta) & \text{otherwise} \end{cases}$$

In the presence of outliers, using MAE or Huber Loss can provide a more robust evaluation of your SVM regression model. These metrics are less influenced by extreme values, making them suitable for scenarios where outliers might have a significant impact on model performance.

Consideration of the specific characteristics of your dataset and the nature of the outliers can guide the choice between MAE, MSE, or other robust regression metrics.

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?
Ans:When you have built an SVM regression model using a polynomial kernel and find that both Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are very close, either metric can be considered for evaluating the performance of your model. The choice between MSE and RMSE depends on the specific interpretation and requirements of your regression task:

MSE (Mean Squared Error):

- MSE measures the average squared difference between the predicted and actual values.
- It provides a straightforward interpretation, and the values are in the squared units of the target variable.
- If the units of the target variable are meaningful and squaring makes sense in the context of your problem, MSE is a suitable choice.

MSE=1$\sum$=1(−^)2

$$MSE=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y})$$

i

)

2

RMSE (Root Mean Squared Error):

- RMSE is the square root of MSE and has the same unit as the target variable.
- It provides a more interpretable measure of prediction error in the original units of the target variable.
- RMSE penalizes larger errors more than smaller ones, and it is suitable when you want a metric that reflects the magnitude of errors in a more intuitive way.

RMSE=MSE

RMSE=

MSE

Considerations:

- If the units of the target variable are important in your interpretation, RMSE may be preferred because it provides a measure of average error in the original units.
- If you are not concerned about the units and are interested in the overall accuracy without a specific emphasis on the magnitude of errors, MSE can be used.
- In practice, the choice between MSE and RMSE often depends on the preferences of the stakeholders, the nature of the problem, and the specific context in which the regression model is being used.

Ultimately, both MSE and RMSE are valid metrics for evaluating the performance of your SVM regression model. You may choose the one that aligns better with your goals and the way you want to communicate the accuracy of your predictions.

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?

Ans:If your goal is to measure how well the model explains the variance in the target variable, the most appropriate evaluation metric is the coefficient of determination, commonly known as R-squared (

$\hat{}2$

R

2

). R-squared provides a measure of the proportion of the variance in the target variable that is explained by the independent variables in your model.

Here's how R-squared is calculated:

$$\text{R2}=1-\sum\text{i=1n(yi}-\text{yi^)2}\sum\text{i=1n(yi}-\text{y¯)2}$$

R

2

$=1-$

$\sum$

$i=1$

n

$(y$

i

$-$

y

$-$

$)$

2

$\sum$

$i=1$

n

$(y$

i

$-$

y

$\hat{}$

i

$)$

2

Where:

- ��
- y
- i
-
- is the actual value of the target variable.
- ��^�
  - y
  - $\hat{}$
  -
- i
-
- is the predicted value of the target variable.
- ��
  - y
  - $\bar{}$

-
- is the mean of the target variable.

R-squared ranges from 0 to 1, where:

- $\hat{2}=1$
- R
- 2
- $=1$ indicates a perfect fit, where the model explains all the variance in the target variable.
- $\hat{2}=0$
- R
- 2
- $=0$ indicates that the model does not explain any variance, and its predictions are equivalent to the mean of the target variable.
- $0<\hat{2}<1$
- $0<R$
- 2
- <1 indicates the proportion of variance explained by the model.

In the context of comparing SVM regression models with different kernels (linear, polynomial, and RBF), R-squared is suitable because it provides a standardized measure of how well each model captures the variance in the target variable. Higher

$\hat{2}$

R

2

values indicate a better fit, and you can use this metric to compare the explanatory power of the different models.

python

Copy code

```
from                    import
```

```
print f"R-squared (R2) Score: {r2_value}"
```

Selecting the model with the highest

$$R^2$$

value would indicate the one that explains the variance in the target variable most effectively among the compared SVM regression models.