A5: SVR Interpretation

A diagram of a graph

Description automatically generated

RMSE on the training set: 93.53167104564383

RMSE on the test set: 170.97988255842617

R-squared on the test set: 0.7445366176469796

1. Best Hyperparameter Values:

The best hyperparameter values obtained from the hyperparameter tuning process will depend on the specific data and the random seed used during the process. You should run your script to obtain the exact values. However, I can explain the significance of the parameters.

Parameter C: The value of C in an SVR model controls the trade-off between achieving a low training error and a low testing error. A smaller C makes the model more tolerant of errors on the training set (allowing some slack), which might be useful for handling outliers. Conversely, a larger C makes the model less tolerant of training errors and might result in a narrower margin.

Parameter Gamma: The gamma parameter determines how much influence a single training example has. Smaller values of gamma make the RBF kernel relatively wide (each example has a more global influence), while larger values of gamma make the kernel relatively narrow (each example has a more local influence). Narrow kernels can lead to more complex models that are sensitive to individual data points.

2. Trusting the Learned Model for Novel Predictions:

Whether you trust the learned model to make predictions for a novel material composition depends on the model's performance on your test set and your domain knowledge. Here are some considerations:

Performance Metrics: Check the RMSE and R-squared values on the test set. Lower RMSE and higher R-squared values indicate better predictive performance. If these values are acceptable, you can trust the model to some extent.

Domain Knowledge: Consider your understanding of the problem domain. If your dataset is representative of real-world scenarios and the model performs well on your test data, you can be more confident in its predictions.

Outliers: The value of C and the model's sensitivity to outliers is also a critical factor. If your dataset contains outliers and the model is highly sensitive to them (small C), you should be cautious when making predictions for novel compositions.

3. Number of Support Vectors:

The number of support vectors in the model depends on the value of epsilon and other factors. With a larger epsilon (epsilon of 100), you are allowing a larger margin for errors.

Increasing epsilon usually results in fewer support vectors because it allows the model to have a larger margin and tolerate errors. In other words, a larger epsilon makes the model more tolerant of errors and may lead to a simpler model with fewer support vectors.

Decreasing epsilon often results in more support vectors because it tightens the margin and makes the model less tolerant of errors.