**Rohan Singh**

**428005621**

**A1: Linear Regression) Task 2**

**Question 1**: How many parameters (i.e., weights) are present in each of the three models you used? Explain how you know/determined this. Answers given without an explanation or supporting calculations receive zero credit.

The number of parameters in each model depends on the number of features considered. In our case, we have two features, "x1" and "x2." The determination of the number of parameters for the linear, quadratic, and cubic models can be calculated as follows:

* For the Linear Model, there is a weight associated with each feature, plus an intercept term, which results in 3 parameters in total.
* For the Quadratic Model, we include all quadratic combinations of the features, leading to 6 additional parameters (x1^2, x2^2, and x1 \* x2), along with the intercept, making it a total of 7 parameters.
* For the Cubic Model, we further include all cubic combinations of the features, which result in 8 additional parameters (x1^3, x2^3, x1^2 \* x2, and x1 \* x2^2), plus the previous parameters and an intercept, totaling 15 parameters.

These calculations are based on the polynomial degree and the inclusion of interaction terms in each model.

Top of Form

**Question 2:**Suppose we fit a 5th order polynomial to the same data. Based on trends in ***training*** loss you observed for the linear, quadratic and cubic models, do you expect training loss for the 5th order polynomial to be higher or lower than the cubic polynomial? Explain the rationale for your answer.

Considering the observed trends in training loss for the linear, quadratic, and cubic models, we would expect that the training loss for the 5th order polynomial will be lower than that of the cubic polynomial. This expectation arises from the fact that higher-order polynomial models tend to fit the training data more closely as their complexity increases. Therefore, the 5th order polynomial, being more complex than the cubic one, is likely to exhibit a lower training loss as it can capture intricate patterns and variations in the training data more effectively. However, it is essential to exercise caution, as excessively high-order polynomials can lead to overfitting, where the model fits the training noise and performs poorly on new, unseen data.

**Question 3**: If any of these models overfit the training data, what would you expect to see in the results?

If any of the models overfit the training data, you would expect to see the following results:

* **Training Loss Much Lower Than Test Loss**: The training loss (mean squared error) would be significantly lower than the test loss. This indicates that the model has learned to fit the training data very closely but fails to generalize well to unseen data.
* **Large Discrepancy Between Training and Test Loss**: There would be a large discrepancy between the training loss and test loss. The model may perform well on the training data but poorly on the test data.
* **Complex Model with Many Parameters**: Overfitting often occurs in complex models with many parameters, such as high-degree polynomial regression.
* **Noisy or Irrelevant Features**: Overfitting can also occur when the model tries to fit noise or irrelevant features in the training data.

**Question 4**: When considering both model accuracy and model complexity, which of the three models do you consider to be best? That is, which would you say should be the final model selected for this problem? Explain your reasoning.

The choice of the best model depends on the trade-off between model accuracy and model complexity, often referred to as the bias-variance trade-off.

* **Linear Model**: This is the simplest model among the three, with only two parameters. It has lower complexity but may underfit the data if the relationship is not linear. It's a good choice if you want a simple, interpretable model.
* **Quadratic Model**: This model has moderate complexity with six parameters and can capture more complex relationships than the linear model. It may be a good choice if you suspect a quadratic relationship in the data.
* **Cubic Model**: The cubic model has higher complexity with twenty parameters, which increases the risk of overfitting. It can capture even more complex patterns but may not generalize well to new data.

Upon evaluating both model accuracy and complexity, the cubic model emerges as the optimal choice. This conclusion is supported by the cubic model's superior performance in terms of the lowest mean squared error during testing. Furthermore, the consistency between the error values for both training and testing sets indicates that the cubic model generalizes well and provides reliable predictions. Therefore, considering the balance between model accuracy and complexity, the cubic model is the recommended final model for this problem.Top of Form