1. Gradient Descent cannot get stuck in a local minimum when training a linear regression model. The cost function in linear regression is a convex function. Convex functions have the property that any local minimum is also a global minimum. This means that if gradient descent finds a minimum, it is the best possible solution.

2. It indicates that the model is likely overfitting. There are ways to prevent this: (i) Reduce the complexity of your polynomial model by using a lower degree polynomial. A simpler model is less likely to overfit the training data, allowing it to generalize better to the validation set. (ii) Increasing the size of the training dataset can help the model learn more general patterns and reduce overfitting. (iii) Apply regularization techniques such as Lasso (L1) or Ridge (L2) regression. Regularization adds a penalty for larger coefficients, which helps to prevent overfitting by discouraging the model from fitting the noise in the training data.

3. (i) With too many predictors, the model can easily fit the noise in the training data, leading to poor generalization on unseen data. It learns the specifics of the training set rather than the underlying relationships (Overfitting). (ii) High correlation between predictors can inflate the variance of coefficient estimates, leading to unreliable results. (iii) The model may have a high variance meaning that it is sensitive to

fluctuations in the training data. Small changes in the training dataset can lead to significantly different models.

4. (i) Outliers can cause the estimated coefficients to shift, leading to a line that doesn't represent the overall trend well.(ii) Outliers increase the variance of the model, making it less generalizable to unseen data. (iii) The model may overfit to outliers, attempting to minimize their large errors, which can worsen performance on typical data points. Techniques to make the model more robust to outliers: (i) Least Absolute Deviations (LAD) Regression: Minimizes the sum of absolute residuals rather than squared residuals. This approach is less sensitive to large residuals, as it does not square them. (ii) Huber Regression: Combines least squares and LAD by treating small residuals as squared (like in least squares) and large residuals as absolute, reducing the impact of outliers. (iii) RANSAC (Random Sample Consensus): Iteratively fits models using random subsets of data, ignoring points that don't fit well. This approach estimates a model that is less influenced by outliers.

5. MAE.ipynb

6. (i) T-test: Compares the means of two groups to determine if they are statistically different from each other, better for Accuracy, F1 score, or any continuous metric derived from model predictions. (ii) Wilcoxon Signed-Rank Test: A

nonparametric alternative to the paired t-test that assesses whether the median of differences between paired observations is significantly different from zero, better for Any metric; particularly useful for ordinal data or when the assumptions of the t-test are not met. (iii) ANOVA: Tests for differences between the means of three or more groups, better for Accuracy, F1 score, or error rate when evaluating multiple models.