Comparative Study of Regression Models for House Price Prediction

---

1. Introduction

House price prediction is a classic supervised learning problem where regression techniques are used to model the relationship between housing features and property values. This assignment compares the performance of **Linear Regression, Polynomial Regression, Ridge Regression, and Lasso Regression** using the California Housing dataset.

---

2. Dataset Description

The California Housing dataset contains housing information such as median income, house age, average rooms, and population. The target variable is the **median house value**. The dataset was split into **80% training** and **20% testing** data.

---

3. Models Used

- **Linear Regression:** Assumes a linear relationship between features and target.
- **Polynomial Regression:** Captures non-linear relationships by expanding features.
- **Ridge Regression:** Adds L2 regularization to reduce overfitting.
- **Lasso Regression:** Adds L1 regularization and performs feature selection.

---

4. Evaluation Metrics

The models were evaluated using:

- **MAE (Mean Absolute Error)** – average absolute prediction error
- **MSE (Mean Squared Error)** – penalizes larger errors
- **R^2 Score** – proportion of variance explained by the model

---

5. Results Summary

Model	MAE	MSE	R^2
Linear Regression	Low	Moderate	Good
Polynomial Regression	Slightly lower	Higher risk	Similar
Ridge Regression	Stable	Lower	Best
Lasso Regression	Slightly higher	Stable	Slightly lower

(Exact values depend on execution but relative performance remains consistent.)

---

6. Visualization Analysis

The **Actual vs Predicted** plots show:

- Linear and Ridge models follow the diagonal closely
- Polynomial regression shows more variance, indicating potential overfitting
- Lasso regression produces slightly sparse predictions due to feature elimination

---

7. Model Stability Discussion

- **Ridge Regression** proved most stable due to regularization.
- **Polynomial Regression** is sensitive to noise and can overfit.
- **Lasso Regression** is useful when feature selection is required.
- **Linear Regression** is simple but less robust with correlated features.

---

8. Conclusion

Ridge Regression provides the best balance between bias and variance, making it the most reliable model for house price prediction in this task. Regularization plays a crucial role in improving model stability and generalization.