Assignment: Implementing Linear Regression using Gradient Descent

1. Dataset & Preprocessing

- Dataset: California Housing (`housing.csv`).
- Missing values in **total_bedrooms** were imputed with the column mean.
- **ocean_proximity** categorical feature encoded using One-Hot Encoding.
- Features were standardized (mean=0, variance=1).
- Dataset split: **80% training, 20% testing**.

---

2. Methodology

- Implemented a custom `LinearRegressionGD` class:
  - Gradient Descent update rule with learning rate α=0.01.
  - 1000 iterations used for convergence.
  - Ridge Regularization (λ) parameter included for extension.

---

3. Results

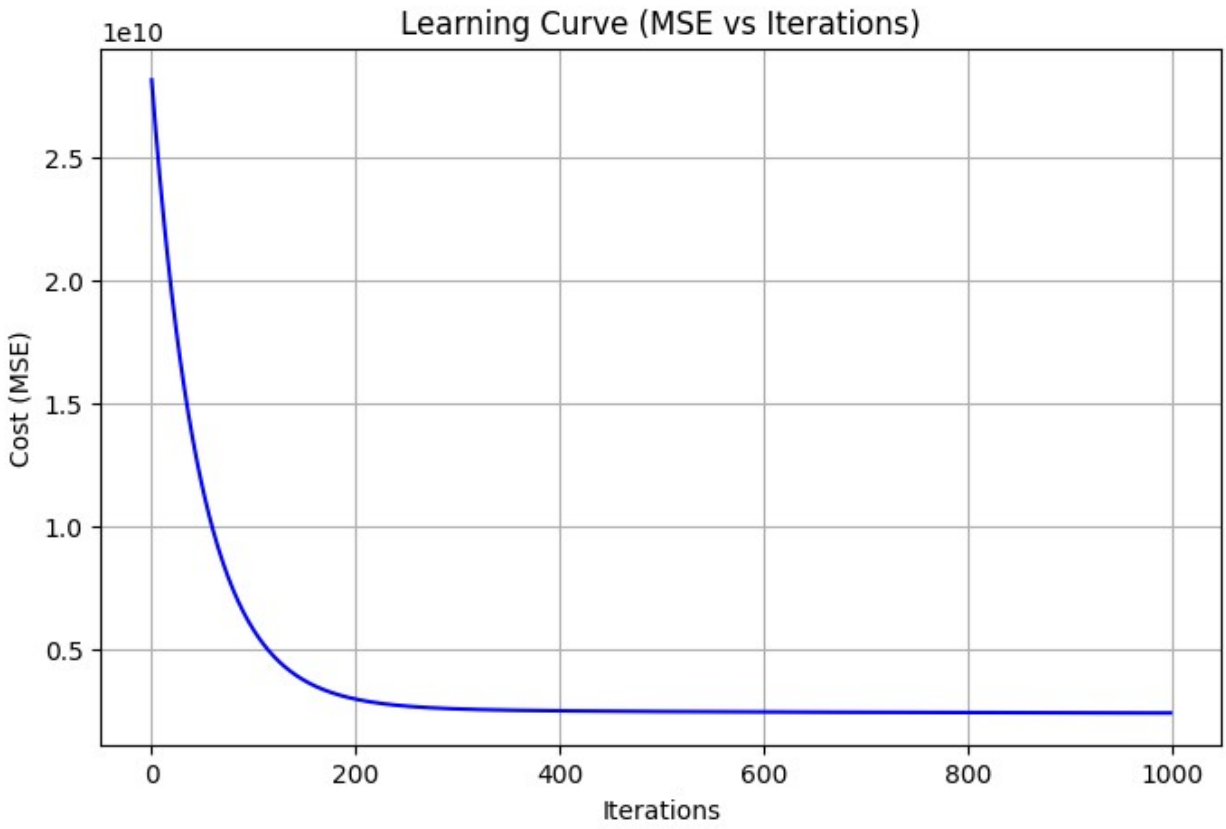
3.1 Model Evaluation

Metric	Value
Mean Squared Error (MSE)	`4970832510.2708`
R^2 Score	`0.6207`

3.2 Plots

(a) Learning Curve

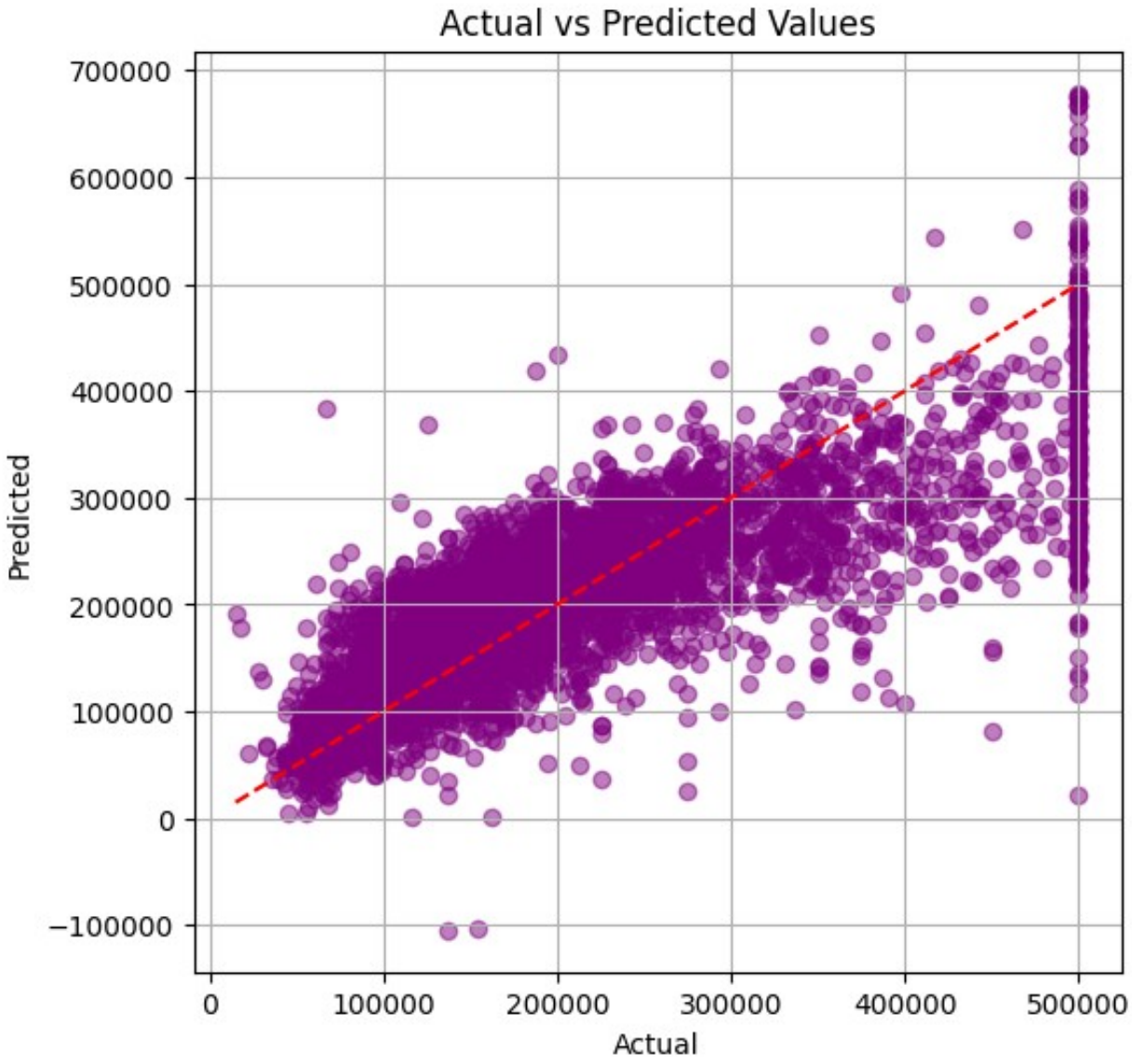
- Cost (MSE) vs Iterations shows gradual decrease.
- Model converged successfully.



(b) Actual vs Predicted

- Scatter plot shows predictions aligning with the 45° line.

- Indicates decen



here from Colab output)*

---

4. Observations

- The model converged with α=0.01 and 1000 iterations.
- Larger α caused divergence, while smaller α slowed training.
- R^2 indicates that the model explains a moderate proportion of variance.
- Predictions align reasonably with actual values, though noise exists.
- Regularization (λ > 0) reduces overfitting slightly at the cost of bias.

5. Deliverables

- **Google Colab Link:** https://colab.research.google.com/drive/1whW5Z-nK8UlOVNarcLbUP85GlF5sZ2qS?usp=sharing

- **Code:** Implemented entirely from scratch (Gradient Descent).