|  |
| --- |
| **Generative AI Consortium (Ltd)**  **AI/ML Internship: Assignment 1 (Simple Machine Learning Problem) Assignment)**  **Name: ARUNESH S** |
| **Email:** [**mailto:aruneshsakthivel@gmail.com**](mailto:aruneshsakthivel@gmail.com)   | **Size (sqft)** | **Bedrooms** | **Location** | **Price ($)** | | --- | --- | --- | --- | | 1500 | 3 | Suburb | 300,000 | | 1600 | 3 | City | 450,000 | | 1700 | 4 | Suburb | 350,000 | | 1400 | 2 | City | 400,000 | | 1800 | 4 | Suburb | 370,000 | | 10000 | 10 | Suburb | 2,000,000 | |

* **Features**: Size, Bedrooms, Location.
* **Label**: Price.
* **Prediction**: For a new house with given features, the model predicts its price.
* **Outlier**: The house with 10000 sqft and 10 bedrooms has an unusually high price, making it an outlier.
* **Training Data**: First 4 records could be used to train the model.
* **Test Data**: Last 2 records could be used to test the model.
* **Model**: A linear regression model could be used to predict house prices.
* **Validation Data**: One of the records can be used as validation data to tune hyperparameters.
* **Hyperparameter**: Learning rate of the model.
* **Epoch**: The model is trained through one complete pass over the 4 training records.
* **Loss Function**: Mean Squared Error (MSE) between predicted prices and actual prices.
* **Learning Rate**: Determines how much the model's weights are updated with each epoch.
* **Overfitting**: If the model predicts training data very well but performs poorly on test data.
* **Underfitting**: If the model cannot capture the underlying trend of the data (e.g., predicting average price for all houses).
* **Regularization**: Adding a penalty to prevent overfitting, such as Ridge or Lasso.
* **Cross-Validation**: Using k-fold cross-validation to ensure the model generalizes well.
* **Feature Engineering**: Creating a new feature like "price per bedroom" from existing features.
* **Dimensionality Reduction**: Using PCA to reduce features to "Size" and "Bedrooms".
* **Bias**: If the model consistently underestimates or overestimates house prices.
* **Variance**: If the model's predictions vary widely with small changes in the training data.