

ICS1502 - Introduction to Machine Learning

Theory Assignment 1

Regression and Linear Classification Analysis - Mobile Phone
Price Prediction and Bank Note Authentication

I S Rajesh

3122237001042

M.Tech CSE Integrated 5 Years Course

Contents

1	Introduction	2
2	Regression: Mobile Phone Price Prediction	3
2.1	Dataset Description	3
2.2	Mathematical Foundation	3
2.3	Closed-Form Solution	4
2.4	Gradient Descent Approach	4
2.5	Matrix Configuration and Interpretation	4
2.6	Error Analysis and Performance Metrics	5
2.7	Results and Discussion	5
2.7.1	Unregularized performance	5
2.7.2	Closed-form with L2 regularization (no standardization)	5
2.7.3	Closed-form with L2 regularization (with standardization)	6
2.7.4	Feature Importance	6
2.8	Predicted vs Actual Plots	7
2.9	Effect of Regularization and Standardization	7
3	Linear Classification: Bank Note Authentication	9
3.1	Dataset Overview	9
3.2	Linear Classification Model	9
3.3	Training and Testing	10
3.4	Performance Comparison	10
3.5	3D Visualization of Features	11
3.6	Effect of Outliers	11
3.7	Discussion	12
4	Overall Conclusion	13

Chapter 1

Introduction

Machine learning models are widely applied for both regression and classification tasks. In this report, we study two problems:

1. Predicting mobile phone prices using multiple numerical features (regression problem).
2. Authenticating bank notes based on extracted image features (binary classification problem).

For regression, we implement **Linear Regression** using both the closed-form solution (normal equation) and Gradient Descent (GD) optimization. We further analyze the impact of **L2 regularization (Ridge Regression)** and data standardization.

For classification, we apply a **Linear Classifier (Logistic Regression)** with and without L2 regularization to the Bank Note Authentication dataset and assess the effect of outliers on the model.

Chapter 2

Regression: Mobile Phone Price Prediction

2.1 Dataset Description

The Mobile Phone Price dataset contains features such as Ratings, RAM, ROM, Mobile_Size, Primary_Cam, Selfi_Cam, Battery_Power and the continuous target variable *Price*.

2.2 Mathematical Foundation

Given a dataset with m samples and n features:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1n} \\ 1 & x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

Here, $X \in \mathbb{R}^{m \times (n+1)}$ is the design matrix with an intercept term, $\mathbf{y} \in \mathbb{R}^m$ is the label vector, and $\mathbf{w} \in \mathbb{R}^{(n+1)}$ is the parameter vector.

The linear regression model predicts:

$$\hat{\mathbf{y}} = X\mathbf{w}$$

2.3 Closed-Form Solution

The parameters that minimize the Mean Squared Error (MSE):

$$J(\mathbf{w}) = \frac{1}{m} \|X\mathbf{w} - \mathbf{y}\|^2$$

are obtained by solving:

$$\mathbf{w} = (X^T X)^{-1} X^T \mathbf{y}$$

With **L2 regularization** (Ridge Regression), the cost function becomes:

$$J(\mathbf{w}) = \frac{1}{m} \|X\mathbf{w} - \mathbf{y}\|^2 + \lambda \|\mathbf{w}\|^2$$

and the closed-form solution:

$$\mathbf{w} = (X^T X + \lambda I)^{-1} X^T \mathbf{y}$$

2.4 Gradient Descent Approach

The model can also be optimized iteratively using batch gradient descent:

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \alpha \frac{2}{m} X^T (X\mathbf{w}^{(t)} - \mathbf{y})$$

where α is the learning rate.

For Ridge Regression, we add a penalty term:

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \alpha \left[\frac{2}{m} X^T (X\mathbf{w}^{(t)} - \mathbf{y}) + 2\lambda \mathbf{w}^{(t)} \right]$$

2.5 Matrix Configuration and Interpretation

For our dataset (7 features), the matrices used are:

- $X: m \times 8$ (includes intercept)
- $\mathbf{y}: m \times 1$
- $\mathbf{w}: 8 \times 1$

Each parameter w_j corresponds to a feature coefficient. Larger magnitudes indicate greater influence on the predicted price (after appropriate standardization when comparing magnitudes).

2.6 Error Analysis and Performance Metrics

We use the following performance measures:

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2, \quad \text{RMSE} = \sqrt{\text{MSE}}, \quad \text{MAE} = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|, \quad R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

2.7 Results and Discussion

The experiments used an 80/20 train/test split. Below are the observed metrics from the runs.

2.7.1 Unregularized performance

Method	MSE	RMSE	MAE	R^2
Closed-form (no reg)	239,357,657.43	15,471.19	10,005.45	0.4332
Gradient Descent (no reg)	333,333,858.38	18,257.43	11,987.07	0.2107

Table 2.1: Unregularized regression performance on test data.

Interpretation: Closed-form (normal equation) outperformed GD — GD showed worse MSE and R^2 , likely due to suboptimal learning rate / convergence or the numerical difficulty of the specific feature scales. The closed-form solution gave a reasonable explanatory power ($R^2 \approx 0.43$), meaning about 43% of variance in price is explained by the linear model.

2.7.2 Closed-form with L2 regularization (no standardization)

λ	MSE	RMSE	MAE	R^2
0	239,357,657.43	15,471.19	10,005.45	0.4332
1	238,951,584.52	15,458.06	9,986.70	0.4342
10	237,174,595.07	15,400.47	9,861.57	0.4384
100	252,459,040.24	15,888.96	10,095.50	0.4022
1000	278,487,625.55	16,687.95	10,571.58	0.3406

Table 2.2: Closed-form regression with L2 regularization (unstandardized features).

Interpretation: Small λ values (1 and 10) slightly improved R^2 and reduced errors — indicating modest variance reduction. Very large λ values (100, 1000) caused underfitting

and worse performance.

2.7.3 Closed-form with L2 regularization (with standardization)

λ	MSE	RMSE	MAE	R^2
0	239,357,657.43	15,471.19	10,005.45	0.4332
1	239,200,533.96	15,466.11	9,998.79	0.4336
10	237,857,213.76	15,422.62	9,942.02	0.4368
100	229,594,215.31	15,152.37	9,515.14	0.4563
1000	254,307,026.19	15,947.01	9,527.68	0.3978

Table 2.3: Closed-form regression with L2 regularization (standardized features).

Interpretation: Standardization improved the behavior of regularization. Notably $\lambda = 100$ on standardized data improved R^2 to 0.4563 (best observed), showing that scaling features can change the effective bias–variance trade-off and allow larger λ to be useful.

2.7.4 Feature Importance

Using standardized L2 weights (reported for $\lambda = 1$), absolute coefficient magnitudes indicate feature importance:

Bias	14287.049612
Ratings	9692.156085
RAM	5700.134255
Primary_Cam	5286.525143
ROM	3941.592734
Battery_Power	2208.697763
Selfi_Cam	828.568473
Mobile_Size	465.636685

Interpretation: **Ratings** and **RAM** are strong predictors of price, followed by primary camera and ROM. Selfie camera and mobile size have relatively small influence.

2.8 Predicted vs Actual Plots

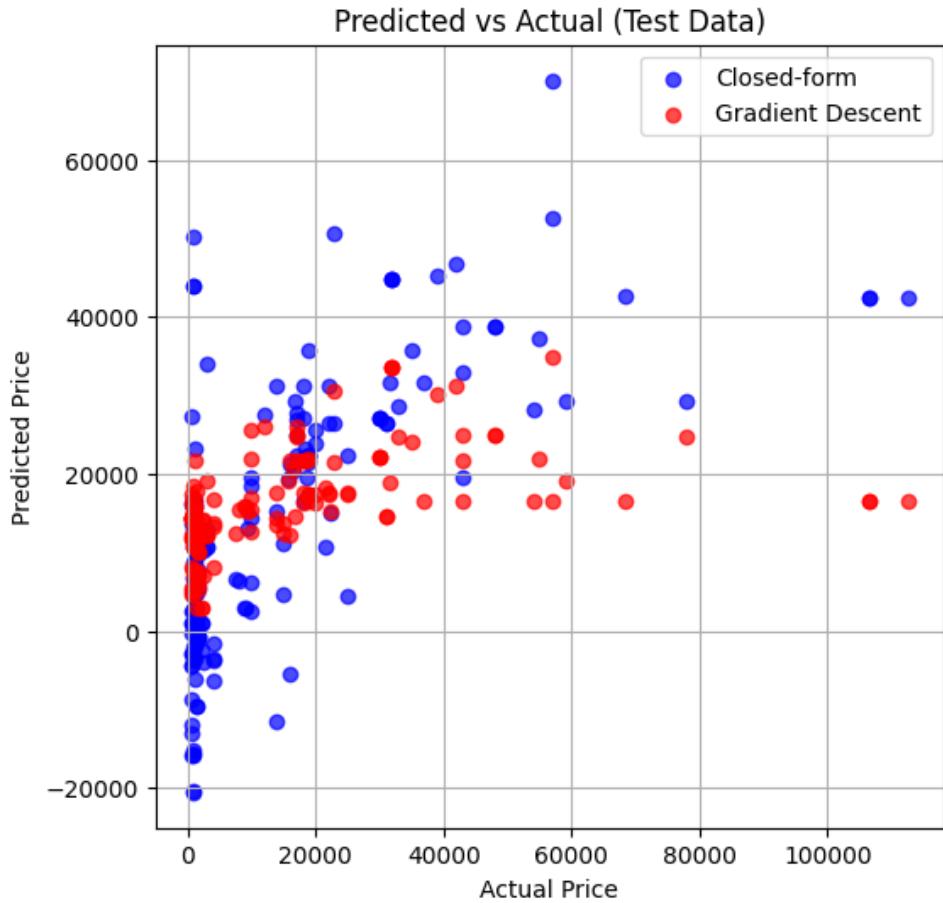


Figure 2.1: Predicted vs Actual Prices for Test Data (Closed-form and GD).

2.9 Effect of Regularization and Standardization

L2 regularization (small λ) reduces variance and slightly improves generalization; large λ causes underfitting. Standardization is important prior to regularization to ensure features contribute comparably to the penalty term; this was borne out in the experiments where standardized inputs allowed $\lambda = 100$ to improve performance.

Effect of Regularization on Model Performance

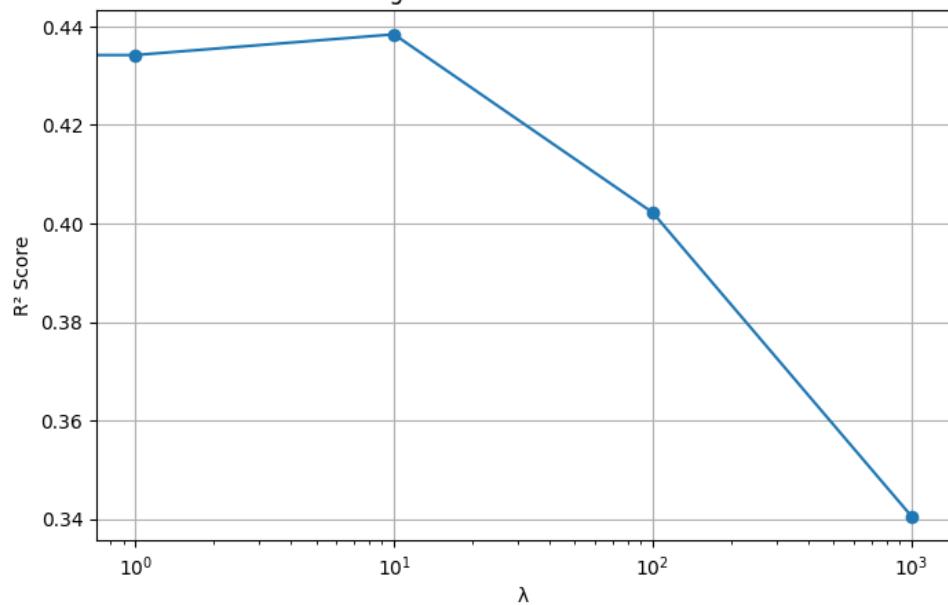


Figure 2.2: Effect of Regularization.

Chapter 3

Linear Classification: Bank Note Authentication

3.1 Dataset Overview

The Bank Note Authentication dataset contains 1372 samples with four features:

- Variance of wavelet-transformed image
- Skewness of wavelet-transformed image
- Curtosis of wavelet-transformed image
- Entropy of image

The target variable is binary: 0 (genuine) or 1 (forged).

3.2 Linear Classification Model

We employ Logistic Regression as the linear classifier.

$$\hat{y} = \sigma(\mathbf{w}^T \mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}}}$$

where $\sigma(\cdot)$ is the sigmoid function.

The cost function with L2 regularization is:

$$J(\mathbf{w}) = -\frac{1}{m} \sum_{i=1}^m [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] + \frac{\lambda}{2m} \|\mathbf{w}\|^2$$

3.3 Training and Testing

The data was divided into 80% training and 20% testing sets. Models were trained for multiple λ values.

3.4 Performance Comparison

λ	Train Accuracy	Test Accuracy
0	0.9863	0.9818
0.1	0.9863	0.9818
1	0.9872	0.9818
10	0.9891	0.9855
100	0.9809	0.9782

Table 3.1: Training and Test accuracy for logistic regression at different λ .

Interpretation: The classifier achieves very high accuracy across lambda values (approximately 98%). Slight improvements are observed around $\lambda = 10$; excessive regularization ($\lambda = 100$) slightly reduces accuracy.

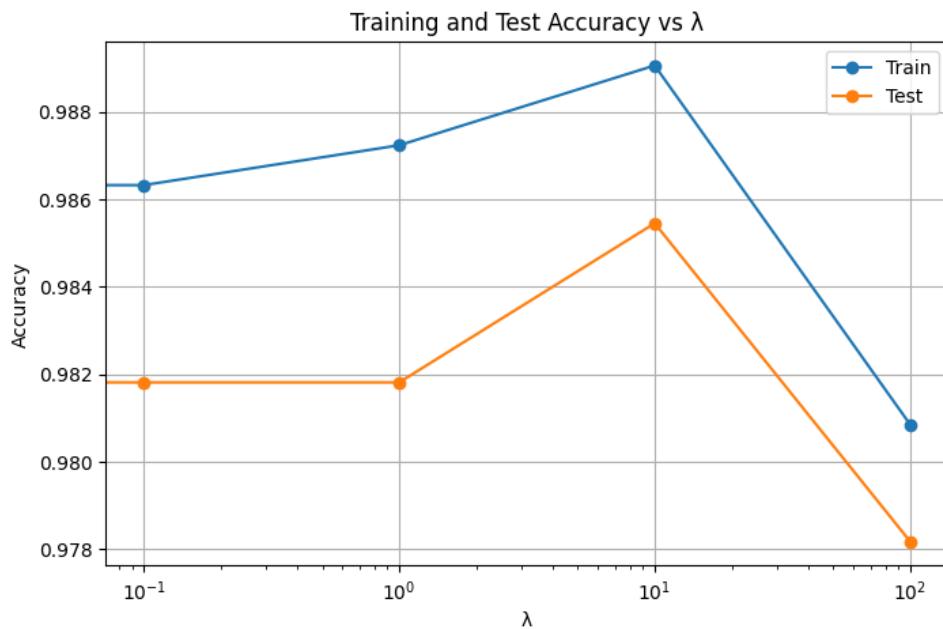


Figure 3.1: Effect of λ on Training and Test accuracy.

3.5 3D Visualization of Features

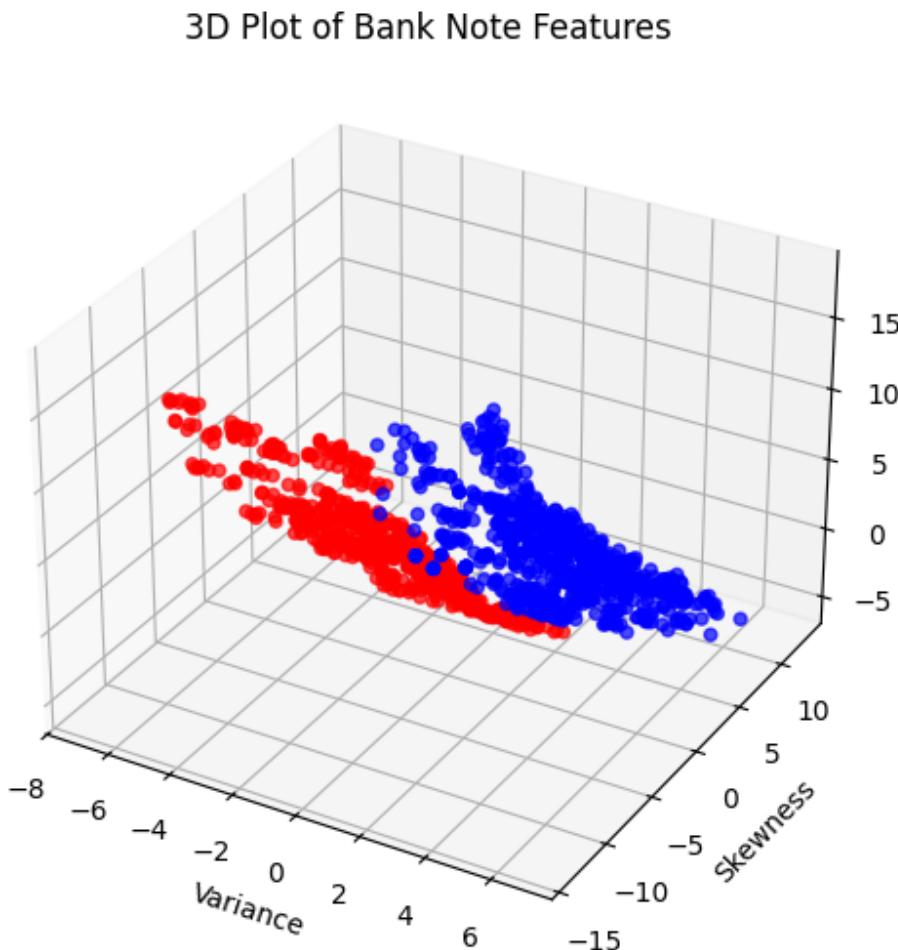


Figure 3.2: 3D visualization using top-3 features: variance, skewness, curtosis.

3.6 Effect of Outliers

Outliers were introduced by shifting 5% of the training points and re-training classifiers.

Model	Test Accuracy (after outlier injection)
Without Regularization	0.9164
With L2 Regularization ($\lambda = 1$)	0.9164

Table 3.2: Impact of outliers on classifier accuracy (test set).

Interpretation: Both regularized and unregularized models see a drop to 0.9164 accuracy after outlier injection. On this dataset and with the specific injection method, L2 at $\lambda = 1$ did not provide a large recovery effect — suggesting that the outliers shifted the distribution in a way that affected class boundaries substantially. Different outlier

strategies or robust methods (e.g., trimming, robust loss, or using stronger regularization) could be explored further.

3.7 Discussion

Linear classification is highly effective on this dataset (nearly linearly separable). L2 regularization provides modest improvements and helps control model complexity; however, outliers still negatively impact performance, indicating the need for robust preprocessing or alternative loss functions for severe contamination.

Chapter 4

Overall Conclusion

Both regression and classification models benefited from L2 regularization and standardization. The matrix-based regression (closed-form) was reliable and interpretable; gradient descent underperformed in our run (likely due to hyperparameter tuning or scaling issues). Logistic regression worked very well for the Bank Note dataset with test accuracies above 98% before outlier injection.

Key Takeaways:

- Closed-form regression is efficient for moderate feature sizes and provides interpretable weights.
- Gradient descent requires careful choice of learning rate and preprocessing (standardization) to match closed-form performance.
- L2 regularization improves generalization in most cases; proper feature scaling is important.
- Linear classifiers can be highly effective when the data is near-linearly separable, but robustness to outliers must be considered in deployment.