Mathematical Foundations of ML: Predicting HDB Prices with Linear Regression and Neural Networks

MIT 2026, Hwa Chong Institution

1 Introduction

This project implements **linear regression** and a **multi-layer perceptron (MLP)** from scratch using NumPy, applying them to predict HDB resale prices (229,273 transactions from data.gov.sg). We demonstrate H2 Mathematics concepts: statistics, calculus, and linear algebra. **Theme:** SDG 11 – Sustainable Cities.

2 Linear Regression: Mathematical Framework

Model: $\hat{y} = \mathbf{w}^T\mathbf{x} + b = \sum_{i=1}^{n} w_i x_i + b$
Loss (MSE): $\mathcal{L} = \frac{1}{m}\sum_{i=1}^{m}(y^{(i)} - \hat{y}^{(i)})^2$
Gradient Descent: Using partial derivatives:

$$\frac{\partial \mathcal{L}}{\partial w_j} = -\frac{2}{m}\sum_{i=1}^{m} x_j^{(i)}(y^{(i)} - \hat{y}^{(i)}) \qquad w_j := w_j - \alpha\frac{\partial \mathcal{L}}{\partial w_j} \tag{1}$$

$$\frac{\partial \mathcal{L}}{\partial b} = -\frac{2}{m}\sum_{i=1}^{m}(y^{(i)} - \hat{y}^{(i)}) \qquad b := b - \alpha\frac{\partial \mathcal{L}}{\partial b} \tag{2}$$

Normal Equation (closed-form): $\boldsymbol{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$

3 Multi-Layer Perceptron: Mathematical Framework

Forward Pass:

$$\mathbf{z}^{[1]} = \mathbf{W}^{[1]}\mathbf{x} + \mathbf{b}^{[1]} \qquad \mathbf{a}^{[1]} = \mathrm{ReLU}(\mathbf{z}^{[1]}) = \max(0, \mathbf{z}^{[1]}) \qquad \hat{y} = \mathbf{W}^{[2]}\mathbf{a}^{[1]} + b^{[2]} \tag{3}$$

Backpropagation (Chain Rule):

$$\frac{\partial \mathcal{L}}{\partial \mathbf{z}^{[2]}} = -\frac{2}{m}(\mathbf{y} - \hat{\mathbf{y}}) \qquad \frac{\partial \mathcal{L}}{\partial \mathbf{z}^{[1]}} = \left(\frac{\partial \mathcal{L}}{\partial \mathbf{z}^{[2]}} \cdot \mathbf{W}^{[2]T}\right) \odot \mathrm{ReLU}'(\mathbf{z}^{[1]}) \tag{4}$$

Weight Gradients: $\frac{\partial \mathcal{L}}{\partial \mathbf{W}^{[2]}} = \mathbf{a}^{[1]T} \cdot \frac{\partial \mathcal{L}}{\partial \mathbf{z}^{[2]}}$ and $\frac{\partial \mathcal{L}}{\partial \mathbf{W}^{[1]}} = \mathbf{x}^T \cdot \frac{\partial \mathcal{L}}{\partial \mathbf{z}^{[1]}}$

4 Results

Features: floor_area_sqm, remaining_lease_years, storey_mid, town (one-hot), flat_type (one-hot)

Model	$\mathbf{R^2}$ **(Test)**	**Key Insight**
Linear Regression	0.516	Interpretable: +$2k/sqm, +$37k/year lease
MLP (synthetic)	0.974	Captures non-linearity ($y = x^2 + \sin(x)$)
MLP (HDB data)	0.36–0.50	Comparable to LR; HDB prices are mostly linear

5 H2 Mathematics Connections

Topic	**Application in This Project**
Statistics	R^2, correlation, regression analysis
Calculus	Gradient descent, chain rule for backpropagation
Functions & Graphs	ReLU activation, non-linear mappings
Linear Algebra	Matrix operations, normal equation $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$

6 Conclusion

Machine learning builds on fundamental mathematics. Linear regression provides interpretable baselines using statistics and calculus. MLPs add flexibility through the chain rule and backpropagation. For HDB pricing, linear relationships dominate, making simple models competitive. The mathematical framework—gradients, chain rule, matrix operations—unifies both approaches.

Code: `/MIT_Project/` **Data:** data.gov.sg **AI Use:** Code structure assistance; all derivations verified independently.