Tutorial: Linear Regression

Cheat Sheet and Practice Problems

ES335 - Machine Learning
IIT Gandhinagar

July 23, 2025

1 Summary from Slides

1.1 Key Concepts

What is Linear Regression?

- Prediction of continuous output variables
- Models linear relationship between features and target
- Examples: $F = ma$, $v = u + at$
- Finds best fit line/hyperplane through data

Mathematical Model: For single feature: $y = \theta_0 + \theta_1 x$ For multiple features: $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + ... + \theta_M x_M$

1.2 Matrix Formulation

Normal Equation Form:

$$\hat{Y} = X\theta$$

Where:

- $\hat{Y}$: $N \times 1$ predicted output vector
- X: $N \times (M + 1)$ design matrix (includes bias column of 1s)
- θ: $(M + 1) \times 1$ parameter vector
- N: number of samples, M: number of features

Design Matrix Structure:

$$X = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,M} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,M} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{N,1} & x_{N,2} & \dots & x_{N,M} \end{bmatrix}$$

1.3 Normal Equation Solution

Objective: Minimize sum of squared errors

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

Closed-form Solution:

$$\theta = (X^T X)^{-1} X^T Y$$

1.4 Basis Expansion

Polynomial Features: Transform x to $[1, x, x^2, x^3, ...]$ to capture non-linear relationships
Still Linear: Model remains linear in parameters θ

1.5 Geometric Interpretation

Projection: Linear regression projects target vector Y onto column space of design matrix X
Residuals: Difference between actual and predicted values, orthogonal to column space

1.6 Dummy Variables and Multicollinearity

Categorical Variables: Use one-hot encoding with dummy variables
Multicollinearity: When features are highly correlated

- Problem: $X^T X$ becomes singular (non-invertible)
- Solution: Remove redundant features or use regularization

2 Practice Problems

Problem : Basic Linear Regression

Given data points: $(1, 3)$, $(2, 5)$, $(3, 7)$, $(4, 9)$
Find the linear regression line $y = \theta_0 + \theta_1 x$ using normal equation.

Problem : Matrix Setup

For a dataset with 5 samples and 2 features, write out the complete matrix equation $\hat{Y} = X\theta$ with proper dimensions.

Problem : Normal Equation Calculation

Given design matrix:

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, \quad Y = \begin{bmatrix} 5 \\ 7 \\ 9 \end{bmatrix}$$

Calculate $\theta = (X^T X)^{-1} X^T Y$ step by step.

Problem : MSE Calculation

For the predictions $\hat{y} = [2.1, 3.9, 6.2, 7.8]$ and actual values $y = [2, 4, 6, 8]$:
a) Calculate the Mean Squared Error b) Calculate the residuals c) Verify that residuals sum to zero (approximately)

Problem : Polynomial Regression

Transform the dataset $x = [1, 2, 3]$ into polynomial features up to degree 3. Write the resulting design matrix.

Problem : Multicollinearity Detection

Given correlation matrix between three features:

$$R = \begin{bmatrix} 1.0 & 0.9 & 0.1 \\ 0.9 & 1.0 & 0.2 \\ 0.1 & 0.2 & 1.0 \end{bmatrix}$$

Identify which features are highly correlated and explain the potential problems.

Problem : Dummy Variables

Convert the categorical variable "Color" with values [Red, Blue, Green, Red, Blue] into dummy variables. Show the resulting design matrix.

Problem : Geometric Interpretation

Explain why the residual vector is orthogonal to the column space of the design matrix. What does this mean geometrically?

Problem : Non-invertible Matrix

When does $(X^T X)$ become non-invertible? Give three specific scenarios and explain how to handle each case.

Problem : Feature Scaling

Dataset before scaling: $X = \begin{bmatrix} 1000 & 2 \\ 2000 & 4 \\ 3000 & 6 \end{bmatrix}$

Apply standardization (z-score normalization) to both features. Explain why scaling might be important for linear regression.

Problem : Model Selection

You have three models: - Model 1: $y = \theta_0 + \theta_1 x$ (MSE = 4.2) - Model 2: $y = \theta_0 + \theta_1 x + \theta_2 x^2$ (MSE = 3.8) - Model 3: $y = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$ (MSE = 3.9)

Which model would you choose and why? Consider overfitting.

Problem : Residual Analysis

After fitting a linear regression model, you observe that residuals show a curved pattern when plotted against fitted values. What does this indicate and how would you address it?

Problem : Computational Complexity

Compare the computational complexity of: a) Normal equation method: $(X^T X)^{-1} X^T Y$ b) Gradient descent method

When would you prefer each approach?

Problem : Real-world Application

Design a linear regression model to predict house prices with features: - Size (sq ft) - Number of bedrooms - Age of house - Distance to city center
a) Write the mathematical model b) Identify potential multicollinearity issues c) Suggest preprocessing steps d) How would you validate the model?

Problem : Advanced Challenge

Given that normal equation requires matrix inversion which is $O(n^3)$:
a) For what size datasets does this become computationally prohibitive? b) Explain why iterative methods like gradient descent might be preferred c) What are the trade-offs between exact solution (normal equation) vs approximate solution (gradient descent)?