John Smith, Ph.D.

Department of Computer Science
University Name

Email: email@university.edu
Website: www.university.edu

July 19, 2025

Introduction to Chapter 3: Supervised Learning - Regression

Overview of Supervised Learning

- Supervised Learning is a type of machine learning where an algorithm is trained on a labeled dataset.
- The algorithm maps input features (independent variables) to known outputs (dependent variable).
- The goal is to make predictions based on new, unseen data after training.

What is Regression?

Definition

Regression is a statistical method used in supervised learning to predict a continuous outcome:

- Unlike classification, which predicts discrete labels.
- Examples include estimating house prices, predicting temperatures, or forecasting sales.

Key Concepts in Regression

- **Dependent Variable (Target)**: Outcome variable to predict (e.g., house prices).
- **Independent Variables (Features)**: Factors used for prediction (e.g., size, number of bedrooms).
- **Regression Line**: Best-fitting line that represents the relationship between variables.

Common Types of Regression

- **Linear Regression**:

$$y = mx + b \tag{1}$$

- **Multiple Linear Regression**:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n \tag{2}$$

- **Polynomial Regression**:

$$y = b_0 + b_1 x + b_2 x^2 + \ldots + b_n x^n \tag{3}$$

Importance of Regression Analysis

- Understand relationships between variables.
- Valuable insights for forecasting and decision-making.

Next Steps

Overview of Key Concepts in Supervised Learning - Regression

What is Supervised Learning?

- **Definition**: A type of machine learning where the model is trained on labeled data (input-output pairs).
- **Objective**: To learn a mapping from inputs (features) to outputs (labels) based on training examples.

Key Concepts in Regression

What is Regression?

- **Definition**: A statistical method modeling the relationship between a dependent variable and independent variables.
- **Goal**: Predict continuous outcomes (e.g., predicting house prices based on size and location).

Dependent and Independent Variables

- **Dependent Variable**: The value we aim to predict (e.g., house price).
- **Independent Variables**: Features used for prediction (e.g., square footage, number of bedrooms).

Types of Regression and Metrics

Types of Regression

- **Linear Regression**: Uses a straight line (y = mx + b).
- **Polynomial Regression**: Fits a polynomial equation.
- **Logistic Regression**: For binary classification, represented with an S-shaped curve.

Loss Function

Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

where y_i is the true value and $\hat{y}_i$ is the predicted value.

Example: Simple Linear Regression

Scenario

Predicting the price of a house based on its square footage.

Dataset

- Square Footage: [1500, 2000, 2500, 3000]
- Prices: [300,000,400,000, 500,000,600,000]

Model Representation

$$\text{Price} = m \cdot \text{Square Footage} + b \tag{5}$$

Interpretation

- The slope m indicates the price increase per additional square foot.
- The intercept b represents the base price of the house when square footage is zero.

Important Considerations in Regression

Overfitting vs. Underfitting

- **Overfitting**: Model fits the training data too well; poor performance on unseen data.
- **Underfitting**: Model too simple to capture the underlying trend.

Key Points to Remember

- Supervised Learning is essential for training accurate predictive models.
- Regression is a critical technique for predicting continuous outcomes.
- Understanding types of regression and applications is crucial for effective data modeling.

Conclusion - Summary of Supervised Learning

Key Concepts Recap

- **Supervised Learning**: A type of machine learning that uses labeled data to learn mappings from inputs to outputs.
- **Regression**: A method for predicting continuous numerical outputs, modeling relationships between variables.

Conclusion - Types of Regression

Major Types of Regression

1. **Linear Regression**: Fits a straight line to data.

$$y = mx + b \tag{6}$$

Example: Predicting house prices based on square footage.

2. **Polynomial Regression**: Fits a polynomial curve to data. *Example*: Modeling plant growth over time may require a curve.

3. **Ridge and Lasso Regression**: Add penalty terms to linear regression to prevent overfitting.

$$\text{Ridge:} \quad ||y - X\beta||^2 + \alpha||\beta||^2 \tag{7}$$

Example: Useful for high-dimensional datasets.

Conclusion - Evaluation Metrics and Next Steps

Evaluation Metrics

- **Mean Absolute Error (MAE)**: Average of absolute errors.
- **Mean Squared Error (MSE)**: Average of squared errors; emphasizes larger errors.
- **R^2 Score**: Proportion of variance explained by the model (0 to 1).

Next Steps

- **Practice**: Implement regression models using online datasets.
- **Explore**: Advanced topics like ensemble methods and neural networks.
- **Reflect**: Consider ethical implications of predictive modeling.