Supervised Machine Learning: Theory and Advanced Modeling

References

Reference taken from *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd Edition, Part 1, **The Fundamentals of Machine Learning** (Chapter 2, Chapter 3, and Chapter 4).

Learning Objectives

By the end of this lecture, you will:

- Understand the theoretical foundations of supervised learning.
- Explore regression and classification models in-depth, including their mathematical formulations.
- Learn the significance of evaluation metrics for both regression and classification tasks.
- Familiarize themselves with advanced modeling techniques and their applications.

Supervised Learning: An Overview

Supervised learning is a machine learning paradigm where the model is trained using labeled data. Each training example consists of an input-output pair:

$$\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

- **Dependent Variable (Target):** The variable we aim to predict (e.g., y).
- **Independent Variables (Features):** The variables used as input to predict the target (e.g., x).

Supervised learning includes two primary tasks:

- **Regression:** Predicting continuous values.
- **Classification:** Predicting discrete categories.

Regression Models

What is Regression?

Regression is used to predict a continuous outcome variable based on one or more input features. The goal is to find a relationship between the dependent variable y and the independent variables $x_1, x_2, \ldots, x_p$.

Simple Linear Regression

The simplest form of regression is linear regression:

$$y = \beta_0 + \beta_1 x + \epsilon$$

- y: Dependent variable (output).
- x: Independent variable (input).
- β_0, β_1: Coefficients (intercept and slope).
- ϵ: Error term.

Multiple Linear Regression

For multiple features, the model extends to:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \epsilon$$

Objective: Minimize the residual sum of squares (RSS):

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Advanced Regression Models

- **Ridge Regression:** Adds ℓ_2 regularization to the cost function:

$$\text{RSS}_{\text{Ridge}} = \text{RSS} + \lambda \sum_{j=1}^{p} \beta_j^2$$

- **Lasso Regression:** Adds ℓ_1 regularization:

$$\text{RSS}_{\text{Lasso}} = \text{RSS} + \lambda \sum_{j=1}^{p} |\beta_j|$$

- **ElasticNet:** Combines ℓ_1 and ℓ_2 penalties.

Evaluation Metrics for Regression

- **Mean Absolute Error (MAE):** Measures average absolute differences:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- **Mean Squared Error (MSE):** Penalizes larger errors more heavily:

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

- **R-squared (R^2):** Proportion of variance explained:

$$R^2 = 1 - \frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{tot}}}$$

Classification Models

What is Classification?

Classification predicts a discrete category label. The model learns a decision boundary to separate classes.

Logistic Regression

$$P(y = 1|x) = \frac{1}{1 + e^{-z}} \quad \text{where } z = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

Advanced Classification Models

- **Decision Trees:** Recursive partitioning to separate classes.
- **Random Forest:** Ensemble of decision trees to improve accuracy.
- **Gradient Boosting:** Combines weak learners to form a strong model.

Evaluation Metrics for Classification

- **Accuracy:**

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

- **Precision:** Proportion of true positives among predicted positives.
- **Recall:** Proportion of true positives among actual positives.
- **F1 Score:** Harmonic mean of precision and recall:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Modeling Approaches

When to Use Regression?

- Predicting continuous outcomes.
- Example: Predicting house prices, temperature, or caloric value.

When to Use Classification?

- Predicting categorical outcomes.
- Example: Identifying spam emails, predicting loan approval.

Advanced Modeling Techniques

- **Support Vector Machines (SVM):** Effective for both regression and classification with high-dimensional data.
- **Neural Networks:** Suitable for complex, non-linear relationships.
- **Ensemble Methods:** Combine multiple models to improve accuracy (e.g., Bagging, Boosting).