CS 412 — Introduction to Machine Learning (UIC)	January 23, 2025
Lecture 1	
Instructor: Aadirupa Saha	*Scribe(s): Uwadia Edomwonyi*

[This draft is not fully proofread. Please email any typos/errors to the instructor or directly edit the latex file.]

Overview

This lecture introduces the foundational concepts in supervised machine learning, with an emphasis on understanding functions, prediction tasks, and the role of loss functions in model training.

1. What is a function? How are functions used in ML?
2. Examples of prediction tasks
3. Learning a prediction function from data
4. Loss functions for regression and classification

1 Functions and Prediction

A function $f : X \to Y$ maps each input $x \in X$ to a unique output $y \in Y$. In machine learning:

- X: Input space (features)
- Y: Output space (labels)

Examples of Prediction Tasks

- **Rain Prediction:** $X \subset \mathbb{R}^d$, $Y = \{0, 1\}$
- **Digit Recognition:** $X \subset \mathbb{R}^{307200}$, $Y = \{0, 1, \ldots, 9\}$
- **House Price Prediction:** $X \subset \mathbb{R}^d$, $Y \subset \mathbb{R}_{>0}$

Categorical features can be encoded using one-hot encoding.

2 Learning from Data

We are given a dataset:

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

Our goal is to find a function f such that $f(x_i) \approx y_i$ for all i.
Assume there exists a ground truth function f^* such that:

$$\forall i, \quad f^*(x_i) = y_i$$

Since the exact f^* is usually unknown, we aim to approximate it using learning algorithms.

3 Loss Functions

A *loss function* quantifies the error between predicted and actual outputs. Formally:

$$l : Y \times \hat{Y} \to \mathbb{R}_+ \quad \text{(non-negative real number)}$$

Loss Function Goals

- Encourage predictions $\hat{y}$ that are close to y
- Minimize total error across training examples

3.1 Regression Losses ($Y \subset \mathbb{R}$)

- **Absolute Loss:** $l(y, \hat{y}) = |y - \hat{y}|$
- **Squared Loss:** $l(y, \hat{y}) = (y - \hat{y})^2$

3.2 Classification Losses (Binary)

- **Logistic Loss:**
$$l(y, \hat{y}) = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y})$$

3.3 Classification Losses (Multiclass)

Let $y \in \{0, 1\}^m$, $\hat{y} \in [0, 1]^m$, where m is the number of classes.

- **Multiclass Logistic Loss:**
$$l(y, \hat{y}) = -\sum_{i=1}^{m} y_i \log(\hat{y}_i)$$

Next Lecture

The next lecture will cover:

- Hypothesis Classes and ERM
- Linear Regression in depth
- Bias-Variance Trade-off

References

1. *Pattern Recognition and Machine Learning*, Christopher M. Bishop
2. *The Elements of Statistical Learning*, Hastie, Tibshirani, and Friedman