CSC311: Lecture Notes

exekis

May 2025

Week 1

Types of Machine Learning

- **Supervised Learning**: Have labeled examples of the correct output/behaviour
- **Unsupervised Learning**: no labeled examples – instead, looking for "interesting" patterns in the data
- **Reinforcement Learning**: (not covered) learning system (agent) interacts with the world and learns to maximize a scalar reward signal

Implementing Machine Learning Models and Systems

- Step 1: Understand the problem (is it prediction, learning a good representation).
- Step 2: Formulate the problem mathematically (create notation for your inputs and outcomes and model).
- Step 3: Formulate an objective function that represents success for your model.
- Step 4: Find a strategy to solve the optimization problem on pencil and paper.
- Step 5: Translate the algorithm into code.
- Step 6: Analyze, iterate, improve design choices in your model and algorithm

Nearest Neighbor Methods

Supervised Learning

Supervised Learning: Step-by-Step

Step 1: Define the Problem Build a system that uses training data to make predictions on new inputs.

Task	Inputs	Labels
object recognition	image	object category
image captioning	image	caption
document classification	text document	category
speech-to-text	audio waveform	text

Examples of supervised learning tasks

Step 2: Mathematical Notation

- Represent inputs as vectors $x^{(i)} \in \mathbb{R}^d$
- Training set $D = \{(x^{(1)}, y^{(1)}), \ldots, (x^{(N)}, y^{(N)})\}$
- For regression: $y^{(i)} \in \mathbb{R}$
- For classification: $y^{(i)} \in \{1, \ldots, C\}$

Step 3: Optimization Problem

- Given a novel input x, find optimal prediction
- First approach: Nearest Neighbor algorithm

Nearest Neighbor

For KNN, we need to define a distance metric. The most common distance metric is the Euclidean distance, which is defined as:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

where x and y are two points in n-dimensional space, and x_i and y_i are the coordinates of the points.