Machine Learning

Syed Arham Naqvi

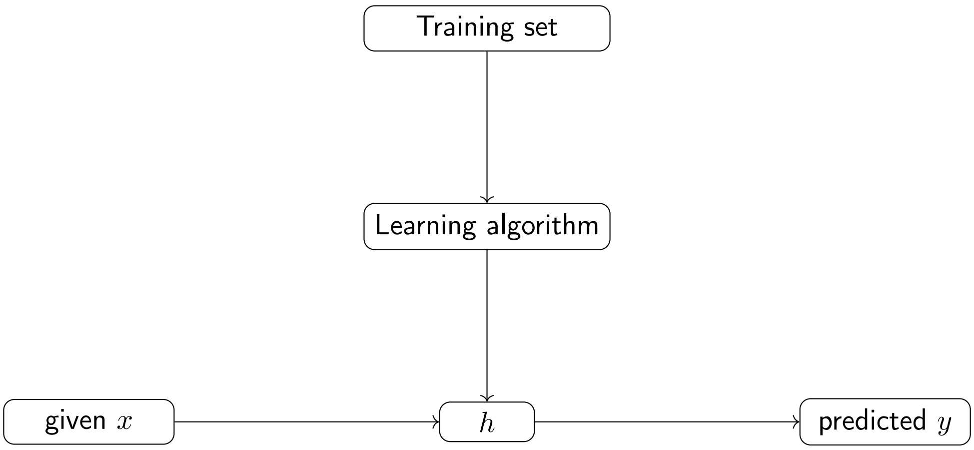
July 1, 2024

Contents

1 Introduction

Supervised learning is a technique in which the training data are pre-labeled, allowing us to learn a parameterized function $\mathbf{h}_{\tilde{\boldsymbol{\theta}}}$ called a **hypothesis**, such that $h_{\vec{\theta}} : X \mapsto Y$ where X is the space of input values and Y is the space of output values. Given input $x^{(i)}$ from the i^{th} training example, the hypothesis should ouptut a prediction $h_{\vec{\theta}}(x^{(i)})$ such that $h_{\vec{\theta}}(x^{(i)}) \approx y^{(i)}$.

A cost function $\mathbf{J}(\tilde{\theta})$ is used to measure the error between our predictions and the known target values given parameters $\vec{\theta}$. The goal is to use various optimization methods to adjust the parameter vector $\vec{\theta}$ such that $J(\vec{\theta})$ is minimized.



Regression refers to a learning problem where y is continuous while **Classification** refers to the case where y can only take on discrete values. These notes contain a breif overview of the following topics:

- Linear Regression
- Locally Weighted Regression
- Logistic Regression

2 Supervised Learning

Given a labeled training set, we can represent our parameterized hypothesis function $h_{\vec{\theta}}(x^{(i)})$ or equivalently $h(x^{(i)})$ for the i^{th} training example as a linear function of the features:

$$\begin{aligned} h_{\vec{\theta}}(x^{(i)}) &= h(x^{(i)}) \\ &= \theta_0 x_0^{(i)} + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + ... \\ &\approx y^{(i)}. \end{aligned}$$

The $x_0^{(i)}$ is not actually part of the training set, it is a dummy feature defined $x_0^{(i)} = 1 \; \forall i \in [1, ..., m]$ where m is the number of training examples. θ_0 is thus the y-intercept of the hypothesis. More succinctly, we can define the linear hypothesis function for an arbitrary training example,

$$h(x) = \sum_{i=0}^{d} \theta_i x_i = \boldsymbol{\theta}^{\mathbf{T}} \mathbf{x} \tag{1}$$

where d is the number of input variables (not counting x_0).

Next we define a cost function that takes the parameter vector $\vec{\theta}$ as input and outputs the associated error between the predicted $h(x^{(i)})$'s and the corresponding $y^{(i)}$'s across all i training examples:

$$J(\vec{\theta}) = \sum_{i=0}^{n} (h(x^{(i)}) - y^{(i)})^2. \tag{2}$$

It is the minimization function through adjustments made to $\vec{\theta}$ that will allow us to fit our hypothesis as accurately to the training data as possible and make the best predictions. Note also as the training data are fixed, the $x^{(i)}$ training examples passed to the hypothesis function are not actually variables as far as J is concerned.

2.1 Definition and Theorems

Definition 2.1 (Limit of a Sequence). *Let $\{a_n\}$ be a sequence of real numbers. We say that a_n converges to $L \in \mathbb{R}$ if for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $n > N$, $|a_n - L| < \epsilon$.*

Theorem 2.2 (Fundamental Theorem of Algebra). *Every non-constant polynomial with complex coefficients has at least one complex root.*

2.2 Examples and Applications

Example 2.3. *Consider the sequence $\{a_n\} = \frac{1}{n}$. This sequence converges to 0 as n approaches infinity.*

2.3 Notes and Remarks

Remark 2.4. *The Fundamental Theorem of Algebra implies that a polynomial of degree n has exactly n roots in the complex plane, counting multiplicities.*

3 Advanced Topics

3.1 Differential Equations

- Introduction to differential equations.
- First-order differential equations.
  - Separable equations.
  - Linear equations.
- Higher-order differential equations.

3.2 Linear Algebra

- Vector spaces.
- Linear transformations.
- Eigenvalues and eigenvectors.
  - Definition and properties.
  - Applications in solving systems of linear equations.

4 Conclusion WOT

- Summary of the main points.
- Potential areas for further study.