Machine Learning Introduction

Blanca Cano Camarero

October 2022

Contents

Introdution

What is Machine learning

As it is said in the introduction of chapter one in Bishop (2006) **machine learning** *is the field of pattern recognition that is concerned with the automatic discovery of regularities in data and the use of these regularities to take actions such as classifying the data into different categories*.

The mathematical formulation of this according to Hastie, Tibshirani, and Friedman (2001) (chapter two) is the following.

Let $X \in \mathbb{R}^p$ a real valued random input vector, and $Y \in \mathbb{R}$ a real valued random out-put variable, with joint distribution $Pr(X, Y)$. We seek a function $f(X)$ for predicting Y given values of the input X.

This theory requires a *loss function* $L(Y, f(X))$ for penalizing errors in prediction, by far the most common and convenient is (**why?**) *squared error loss*

$$L(Y, f(X)) = ((Y - f(X))^2$$

This leads us to a criterion for choosing f.

The expected (squared) prediction error, EPE.

$$EPE(f) = E[(Y - f(X))^2] = \int (y - f(x))^2 Pr(d_x, d_y) \tag{0.1}$$

By conditioning on X and using the Law of total expectation (or law of iterated expectations) we can write EPE as

$$EPE(f) = E_x E_{Y|X} \left([Y - f(X)]^2 | X\right) \tag{0.2}$$

and we see that is suffices to minimize EPE pointwise:

$$f(x) = argmin_c E_{Y|X} \left([Y - c]^2 | X = x\right) \tag{0.3}$$

.

The solution is

$$f(x) = E(Y|X = x). \tag{0.4}$$

(**Why**)

Lineal models

The simplest linear model for regression is

$$y(x, w) = w \cdot x^T \tag{0.5}$$

where $x \in \{1\} \times \mathbb{R}^d$ and $w \in \mathbb{R}^{d+1}$.

Significant models other solutions:

- Splines

- Transformation of x by *basic functions*

$$y(x, w) = w \cdot \phi(x)^T \tag{0.6}$$

Questions related with the relationship between splines and basic functions

- **Why not continuing using polynomial functions as splines?**
- **Think about relation with splines.** Differences -> create table with pro and cons
- **What about simulating analytic functions with?**

Choices for the basic functions

Gaussian basic functions

$$\phi_j(x) = exp\left\{-\frac{(x - \mu_j)^2}{2s^2}\right\} \tag{0.7}$$

Bibliography

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2001. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY, USA: Springer New York Inc.