In-depth notes on data science related topics

Bart Frenk

April 15, 2018

1 Topics

1.1 Generalized linear models

1.1.1 Introduction

Ingredients:

- a **dependent variable** Y that follows a particular distribution in the exponential family,
- a vector of **independent variables** X,
- an invertible function $g : \mathbb{R} \to \mathbb{R}$, referred to as the **link function**,
- an **unknown vector of coefficients** β, of the same length as X

The following relation is assumed to hold:

$$g^{-1}(\mathbb{E}[Y|X]) = \beta \cdot X. \tag{1}$$

1.1.2 Examples

1.1.3 Bayesian regression

1.2 Stochastic approximation

This is important for both online learning (e.g., iteratively computing maximum likelihood estimators as the data comes in, as well as for dealing with large data sets).

1.2.1 Sequential estimators and the Robbins-Monro algorithm

The original article is [1]. It deals solely with the case of stochastically approximating solutions of equations of the form $M(x) = \alpha$, in which α is a fixed constant, and $M(x)$ is the conditional expectation of a random variable Y given $X = x$, i.e.,

$$M(x) = \mathbb{E}[Y \mid X = x] \tag{2}$$

The function M need not be completely specified, but there should be a way of sampling from Y given $X = x$.

In section 2.3.5 of [2] there is a method to derive a sequential method for computing maximum likelihood estimator using the method of Robbins and Monro.

There is a good exposition of this method in the introduction of [3].

From the article:

[1] Robbins, Monro. A stochastic approximation method (1951)

[2] Christopher M. Bishop. Pattern recognition and machine learning (2009)

[3] Toulis ea - Stable Robbins-Monro approximations through stochastic proximal updates (2018)

Consider the problem of estimating the zero θ_* of a function $h : \mathbb{R}^p - > \mathbb{R}$, where (θ) is unknown bu can be unbiasedly estimated by a random variable W_θ such that $\mathbb{E}[W_\theta] = h(\theta)$. Starting from an estimate θ_0, Robbins and Munro iteratively estimated θ_* as follows:

$$\theta_n = \theta_{n-1} - \gamma_n W_{\theta_{n-1}}, \tag{3}$$

where (θ_n) is usually a decreasing sequence of positive numbers, known as the **learning rate sequence**. Typically, we choose $\gamma_n \propto 1/n$, for $n = 1, 2, \ldots$, so that $\sum \gamma_i^2 < \infty$ and $\sum \gamma_i = \infty$.

1.2.2 Stochastic gradient descent

This should be an example of stochastic approximation. Does it fit in the Robbins-Monro framework?

1.2.3 Multi-armed bandits

1. Simplest formulation Ingredients:
   - observation Y in some space $\mathcal{Y}$.
   - reward function $r : \mathcal{Y} \to \mathbb{R}$:
   - set of actions $\mathcal{A}$.
2. Separate reward from observations
3. With context
4. Time dependence of the parameters