Machine learning from scratch

Lecture 1: Mathematical background

Alexis Zubiolo
alexis.zubiolo@gmail.com

Data Science Team Lead @ Adcash

January 26, 2017

Before we start

IT STEP will be organizing a Tech night on **February 16th** (Thursday) from 7pm. I will (probably) be giving a talk. The course will most likely be postponed.

Motivation, vocabulary and notations

In **supervised learning** tasks, we are given a *data set* of the form:

$$D = \left\{ (\boldsymbol{x}^{(i)} y^{(i)}), \boldsymbol{x}^{(i)} \in \mathcal{X}, y^{(i)} \in \mathcal{Y} i \in \{1, \ldots, n\} \right\}$$

Motivation, vocabulary and notations

In **supervised learning** tasks, we are given a *data set* of the form:

$$D = \left\{ (\boldsymbol{x}^{(i)} y^{(i)}), \boldsymbol{x}^{(i)} \in \mathcal{X}, y^{(i)} \in \mathcal{Y} i \in \{1, \ldots, n\} \right\}$$

- n is the size of the data set (number of *instances*/*samples*)
- In most applications:
  - $\mathcal{X} = \mathbb{R}^d$ (d is the *dimensionality*)
  - $\mathcal{Y} = \mathbb{R}$ (*regression*) or $\mathcal{Y} \subset \mathbb{N}$ (*classification*)
- $\boldsymbol{x} \in \mathcal{X}$ is the *feature vector* and $y \in \mathcal{Y}$ is the *label*

Motivation, vocabulary and notations

In **supervised learning** tasks, we are given a *data set* of the form:

$$D = \left\{ (\boldsymbol{x}^{(i)} y^{(i)}), \boldsymbol{x}^{(i)} \in \mathcal{X}, y^{(i)} \in \mathcal{Y} i \in \{1, \dots, n\} \right\}$$

- n is the size of the data set (number of *instances*/*samples*)
- In most applications:
  - $\mathcal{X} = \mathbb{R}^d$ (d is the *dimensionality*)
  - $\mathcal{Y} = \mathbb{R}$ (*regression*) or $\mathcal{Y} \subset \mathbb{N}$ (*classification*)
- $\boldsymbol{x} \in \mathcal{X}$ is the *feature vector* and $y \in \mathcal{Y}$ is the *label*

Solving a **supervised learning problem** is finding (or *learning*) a function (or *hypothesis*) $h : \mathcal{X} \mapsto \mathcal{Y}$ such that for $(x, y) \in D$, $h(\boldsymbol{x})$ is a *good* estimation (or approximation) of y.

Motivation, vocabulary and notations

In **supervised learning** tasks, we are given a *data set* of the form:

$$D = \left\{ (\boldsymbol{x}^{(i)} y^{(i)}), \boldsymbol{x}^{(i)} \in \mathcal{X}, y^{(i)} \in \mathcal{Y} i \in \{1, \dots, n\} \right\}$$

- n is the size of the data set (number of *instances/samples*)
- In most applications:
  - $\mathcal{X} = \mathbb{R}^d$ (d is the *dimensionality*)
  - $\mathcal{Y} = \mathbb{R}$ (*regression*) or $\mathcal{Y} \subset \mathbb{N}$ (*classification*)
- $\boldsymbol{x} \in \mathcal{X}$ is the *feature vector* and $y \in \mathcal{Y}$ is the *label*

Solving a **supervised learning problem** is finding (or *learning*) a function (or *hypothesis*) $h : \mathcal{X} \mapsto \mathcal{Y}$ such that for $(x, y) \in D$, $h(\boldsymbol{x})$ is a *good* estimation (or approximation) of y.

We often write $h(\boldsymbol{x}) = \hat{y}$ ($\hat{y}$ is the *prediction* of $\boldsymbol{x}$ by h).

Hypothesis parametrization

The previous slide raises 2 questions:

- How to define h
- How to assess whether $h(\boldsymbol{x}) = \hat{y}$ is a good approximation of y.