Pattern Recognition and Machine Learning-Exercises

Lautaro Amadei

Aix Marseille Univ, Université de Toulon, CNRS, CPT, Marseille, France

Contents

1 Chapter 1. Introduction

1.1 Exercise 1.1

1.1 ($\star$) **www** Consider the sum-of-squares error function given by (1.2) in which the function $y(x, \mathbf{w})$ is given by the polynomial (1.1). Show that the coefficients $\mathbf{w} = \{w_i\}$ that minimize this error function are given by the solution to the following set of linear equations

$$\sum_{j=0}^{M} A_{ij} w_j = T_i \tag{1.122}$$

where

$$A_{ij} = \sum_{n=1}^{N} (x_n)^{i+j}, \qquad T_i = \sum_{n=1}^{N} (x_n)^i t_n. \tag{1.123}$$

Here a suffix i or j denotes the index of a component, whereas $(x)^i$ denotes x raised to the power of i.

Figure 1: Exercise 1.1

Let us rewrite the error function using Einstein summation notation

$$E(\boldsymbol{\omega}) = \frac{1}{2}\sum_{n=1}^{N}\left\{y(x_n,\boldsymbol{\omega}) - t_n\right\}^2 = \frac{1}{2}\left(y_a - t_a\right)\left(y^a - t^a\right) \tag{1.1}$$

where $a = 1, \ldots, N$ and $y(\boldsymbol{x}, \boldsymbol{\omega}) = \omega_\alpha x^\alpha$

We look for the critical points of the function (1.1) with respect to $\boldsymbol{\omega}$

$$\frac{\partial E}{\partial \omega^\alpha} = \frac{\partial y_a}{\partial \omega^\alpha}\left(y^a - t^a\right). \tag{1.2}$$

Using that

$$\frac{\partial y_a}{\partial \omega^\alpha} = \frac{\partial \omega_\beta}{\partial \omega^\alpha} x_a^\beta = \delta_\beta^\alpha x_a^\beta = x_a^\alpha, \tag{1.3}$$

and using $y(\boldsymbol{x}, \boldsymbol{\omega}) = \omega_\alpha x^\alpha$ we obtain:

$$\frac{\partial E}{\partial \omega^\alpha} = x_a^\alpha\left(\omega_\beta x^{a\ \beta} - t^a\right) = 0 \tag{1.4}$$

From (1.4) we obtain the set of equations

$$\boxed{x_a^\alpha x^{a\ \beta}\omega_\beta = x_a^\alpha t^a}. \tag{1.5}$$

Or, using book's notation, we define

$$x_a^\alpha x^{a\ \beta} = \sum_{n=1}^{N}(x_n)^{\alpha+\beta} = A^{\alpha\beta} \tag{1.6}$$

$$x_a^\alpha t^a = \sum_{n=1}^{N}(x_n)^{\alpha} t_n = T^\alpha \tag{1.7}$$

and obtain the equations in 1

$$A^{\alpha\beta}\omega_\beta = T^\alpha \tag{1.8}$$

or, more explicitly

$$\boxed{\sum_{\beta=0}^{M} A^{\alpha\beta}\omega_\beta = T^\alpha} \tag{1.9}$$

for $\alpha = 0, \ldots, M$.

1.2 Exercise 1.2

1.2 ($\star$) Write down the set of coupled linear equations, analogous to (1.122), satisfied by the coefficients w_i which minimize the regularized sum-of-squares error function given by (1.4).

Figure 2: Exercise 1.1

The regularized sum-of-squares error function is given by

$$E_\lambda(\boldsymbol{\omega}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n,\boldsymbol{\omega}) - t_n\}^2 + \frac{\lambda}{2}|\boldsymbol{\omega}|^2 = \frac{1}{2}\left(y_a - t_a\right)\left(y^a - t^a\right) + \frac{\lambda}{2}\omega_\alpha\omega^\alpha \tag{1.10}$$

As before, we look for the critical points of (1.10)

$$\frac{\partial E_\lambda}{\partial \omega^\alpha} = 0 \implies x_a^\alpha\left(\omega_\beta x^{a\ \beta} - t^a\right) + \lambda\omega^\alpha = 0, \tag{1.11}$$

and obtain the set of equations

$$\boxed{\left(x_a^\alpha x^{a\ \beta} + \lambda\delta^{\alpha\beta}\right)\omega_\beta = x_a^\alpha t^a} \tag{1.12}$$

Defining

$$A_\lambda^{\alpha\beta} = x_a^\alpha x^{a\ \beta} + \lambda\delta^{\alpha\beta} \tag{1.13}$$

$$T^\alpha = x_a^\alpha t^a \tag{1.14}$$

we can write the set of equations in a compact form

$$\boxed{A_\lambda^{\alpha\beta}\omega_\beta = T^\alpha} \tag{1.15}$$

1.3 Exercise 1.3

1.3 $(\star\star)$ Suppose that we have three coloured boxes r (red), b (blue), and g (green). Box r contains 3 apples, 4 oranges, and 3 limes, box b contains 1 apple, 1 orange, and 0 limes, and box g contains 3 apples, 3 oranges, and 4 limes. If a box is chosen at random with probabilities $p(r) = 0.2$, $p(b) = 0.2$, $p(g) = 0.6$, and a piece of fruit is removed from the box (with equal probability of selecting any of the items in the box), then what is the probability of selecting an apple? If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box?

Figure 3: Exercise 1.2

We know the posterior probabilities for getting an apple given the color of the box

$$p(a|r) = \frac{3}{10} \tag{1.16}$$
$$p(a|b) = \frac{1}{2} \tag{1.17}$$
$$p(a|g) = \frac{3}{10} \tag{1.18}$$

Using Bayes' Theorem we can compute the marginal probability $p(a)$ of getting an apple:

$$p(a) = p(a|r)p(r) + p(a|b)p(b) + p(a|g)p(g) = \frac{3}{10} \times \frac{1}{5} + \frac{1}{2} \times \frac{1}{5} + \frac{3}{10} \times \frac{3}{5} = \frac{17}{50} \tag{1.19}$$

Now, for computing $p(g|o)$ we use Bayes' theorem

$$p(g|o) = \frac{p(o|g)p(g)}{p(o)} \tag{1.20}$$

and the marginal probability $p(o)$ of getting an orange

$$p(o) = p(o|r)p(r) + p(o|b)p(b) + p(o|g)p(g) = \frac{2}{5} \times \frac{1}{5} + \frac{1}{2} \times \frac{1}{5} + \frac{3}{10} \times \frac{3}{5} = \frac{18}{50}. \tag{1.21}$$

Putting everything together, we finally obtain the conditional probability $p(g|o)$

$$p(g|o) = \frac{3}{10}\frac{3}{5}\frac{50}{8} = \frac{1}{2}. \tag{1.22}$$

1.4 Exercise 1.5

1.5 ($\star$) Using the definition (1.38) show that $\mathrm{var}[f(x)]$ satisfies (1.39).

Figure 4: Exercise 1.5