Chapter 2: Supervised Learning

Problem 2.1

To walk "downhill" on the loss function (equation (2.5)), we measure its gradient with respect to the parameters ϕ_0 and ϕ_1. Calculate the expressions for the slopes $\partial L/\partial\phi_0$ and $\partial L/\partial\phi_1$.

$$\begin{aligned} \mathrm{L}[\phi] &= \sum_{i=1}^{I}(\mathrm{f}[x_i, \boldsymbol{\phi}] - y_i)^2 \\ &= \sum_{i=1}^{I}(\phi_0 + \phi_1 x_i - y_i)^2 \end{aligned} \tag{2.5}$$

$$\frac{\partial L}{\partial \phi_0} = 2\sum_{i=1}^{I}(\phi_0 + \phi_1 x_i - y_i) \tag{1}$$

$$\frac{\partial L}{\partial \phi_1} = 2\sum_{i=1}^{I}x_i(\phi_0 + \phi_1 x_i - y_i) \tag{2}$$

Problem 2.2

Show that we can find the minimum of the loss function in closed form by setting the expression for the derivatives from problem 2.1 to zero and solving for ϕ_0 and ϕ_1. Note that this works for linear regression but not for more complex models; this is why we use iterative model fitting methods like gradient descent (figure 2.4)

Setting (1) and (2) to zero, we have:

$$\sum_{i=1}^{I}(\phi_0 + \phi_1 x_i - y_i) = 0 \tag{3}$$

$$\phi_0 I + \phi_1 \sum_{i=1}^{I} x_i - \sum_{i=1}^{I} y_i = 0, \tag{4}$$

and

$$\sum_{i=1}^{I} x_i(\phi_0 + \phi_1 x_i - y_i) = 0 \tag{5}$$

$$\phi_0 \sum_{i=1}^{I} x_i + \phi_1 \sum_{i=1}^{I} x_i^2 - \sum_{i=1}^{I} x_i y_i = 0. \tag{6}$$

For simplicity, use the following substitutions:

- $S_x = \sum_{i=1}^{I} x_i$
- $S_y = \sum_{i=1}^{I} y_i$
- $S_{xx} = \sum_{i=1}^{I} x_i^2$
- $S_{xy} = \sum_{i=1}^{I} x_i y_i$

Meaning that equations (4) and (6) become:

$$I\phi_0 + S_x\phi_1 - S_y = 0 \tag{7}$$

$$S_x\phi_0 + S_{xx}\phi_1 - S_{xy} = 0 \tag{8}$$

Solving for ϕ_0 from equation (7):

$$\phi_0 = \frac{S_y - S_x\phi_1}{I} \tag{9}$$

Substituting (9) into equation (8):

$$S_x\left(\frac{S_y - S_x\phi_1}{I}\right) + S_{xx}\phi_1 - S_{xy} = 0 \tag{10}$$

$$\frac{S_x S_y}{I} - \frac{S_x^2\phi_1}{I} + S_{xx}\phi_1 - S_{xy} = 0 \tag{11}$$

$$\phi_1\left(S_{xx} - \frac{S_x^2}{I}\right) + \frac{S_x S_y}{I} - S_{xy} = 0 \tag{12}$$

$$\tag{13}$$

Therefore,

$$\phi_1 = \frac{S_{xy} - S_x S_y / I}{S_{xx} - S_x^2 / I} \tag{14}$$

$$\phi_1 = \frac{I S_{xy} - S_x S_y}{I S_{xx} - S_x^2} \tag{15}$$

Substituting (15) into (9) we can get an equation for ϕ_0:

$$\phi_0 = \frac{S_y S_{xx} - SxS_{xy}}{I(IS_{xx} - S_x^2)} \tag{16}$$

Problem 2.3

Consider reformulating linear regression as a generative model, so we have $x = \mathrm{g}[y, \boldsymbol{\phi}] = \phi_0 + \phi_1 y$. What is the new loss function? Find an expression for the inverse function $y = \mathrm{g}^{-1}[x, \boldsymbol{\phi}]$., that we would use to perform inference. Will this model make the same predictions as the discriminative version for a given training set $\{x_i, y_i\}$? One way to establish this is to write code that fits a line to three data points using both methods and see if the result is the same.

The new loss function is given by:

$$\mathrm{L}[\phi] = \sum_{i=1}^{I} (x_i - (\phi_0 + \phi_1 y_i))^2 \tag{17}$$

where x_i are the observed values, and $phi_0 + \phi_1 y_i$ are the model predictions.

Given the model $x == \phi_0 + \phi_1 y$, solving for y gives:

$$y = \frac{x - \phi_0}{\phi_1} = \mathrm{g}^{-1}[x, \boldsymbol{\phi}] \tag{18}$$

The discriminative model fits parameters to minimise the loss function in equation (2.5), while the generative model fits parameters to minimise the loss function in equation (17). For the same dataset, the two models generally will not make the same predictions. This is because the generative model assumes x as a function of y, while the discriminative model assumes y as a function of x, effectively “flipping” the dependent and independent variables.