Problem Set 1 – Supervised Learning

DS542 – DL4DS

Spring, 2025

Youran Geng

Note: Refer to the equations in the *Understanding Deep Learning* textbook to solve the following problems.

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 expressions for the slopes $\frac{\partial L}{\partial \phi_0}$ and $\frac{\partial L}{\partial \phi_1}$.

Solution. Let $L_i = (\phi_0 + \phi_1 x_i - y_i)^2$, then $L = \sum_{i=1}^{I} L_i$. We have

$$\frac{\partial L_i}{\partial \phi_0} = 2\phi_0 + 2(\phi_1 x_i - y_i),$$
$$\frac{\partial L_i}{\partial \phi_1} = 2x_i^2 \phi_1 + 2x_i(\phi_0 - y_i).$$

By the additive property of derivatives, we know

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

□

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.

Solution. Setting the above two derivative to zero yields

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

In matrix form,

$$\begin{bmatrix} I & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \phi_0 \\ \phi_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

or $X^\top X\phi - X^\top y = 0$, where the sums above are all from $i = 1$ to I, and X is given by

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_I \end{bmatrix}.$$

Since $X^\top X$ is positive semidefinite, it is invertible, and we can solve for ϕ by

$$\phi = (X^\top X)^{-1} X^\top y = \frac{1}{\left(\sum\limits_{i=1}^{I} x_i\right)^2 - \sum\limits_{i=1}^{I} x_i^2} \begin{bmatrix} \left(\sum\limits_{i=1}^{I} x_i\right)\left(\sum\limits_{i=1}^{I} x_i y_i\right) - \left(\sum\limits_{i=1}^{I} x_i^2\right)\left(\sum\limits_{i=1}^{I} y_i\right) \\ \left(\sum\limits_{i=1}^{I} x_i\right)\left(\sum\limits_{i=1}^{I} y_i\right) - I \cdot \sum\limits_{i=1}^{I} x_i y_i \end{bmatrix}.$$

Therefore,

$$\phi_0 = \frac{\left(\sum\limits_{i=1}^{I} x_i\right)\left(\sum\limits_{i=1}^{I} x_i y_i\right) - \left(\sum\limits_{i=1}^{I} x_i^2\right)\left(\sum\limits_{i=1}^{I} y_i\right)}{\left(\sum\limits_{i=1}^{I} x_i\right)^2 - \sum\limits_{i=1}^{I} x_i^2},$$

$$\phi_1 = \frac{\left(\sum\limits_{i=1}^{I} x_i\right)\left(\sum\limits_{i=1}^{I} y_i\right) - I \cdot \sum\limits_{i=1}^{I} x_i y_i}{\left(\sum\limits_{i=1}^{I} x_i\right)^2 - \sum\limits_{i=1}^{I} x_i^2}.$$

$\square$