Class 1: Linear Regression and Neural Networks

Slide 13

We are given

$$x = [16.09, 15.56, 15.85, 15.69, 15.0], \quad y = [17.62, 14.88, 16.32, 16.88, 14.96]$$

Our model is

$$\hat{y}_i = b + \omega x_i, \quad i = 1, \ldots, 5$$

Assume $b = 0$, $\omega = 0.5$. Then

$$\hat{y} = [8.05, 7.78, 7.92, 7.85, 7.51].$$

Define residuals $r_i = y_i - \hat{y}_i$. The loss function is

$$L(\omega, b) = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

For $b = 0$, $\omega = 0.5$:

$$L(\omega, b) = (17.62 - 8.05)^2 + \ldots + (14.96 - 7.51)^2 = 339.29.$$

For $b = 1$, $\omega = 1$:

$$L(\omega, b) = 4.65.$$

Slide 17

The loss function is

$$L(\omega, b) = \sum_{i=1}^{n} (y_i - b - \omega x_i)^2.$$

We want $\frac{\partial L}{\partial \omega}$ and $\frac{\partial L}{\partial b}$.

$$\frac{\partial L}{\partial \omega} = \sum_{i=1}^{n} -2(y_i - b - \omega x_i)x_i = -2\sum_{i=1}^{n} r_i x_i.$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n} 2(y_i - b - \omega x_i)(-1) = -2\sum_{i=1}^{n} r_i.$$

Check with $b = 0$, $\omega = 0.5$:

$$\frac{\partial L}{\partial \omega} = -2\sum r_i x_i = -1283.7, \quad \frac{\partial L}{\partial b} = -2\sum r_i = -81.92.$$

So increasing b and ω will reduce the loss.

Slide 29

Now we have multiple features:

$$\hat{y}_i = b + \omega_1 x_{i1} + \omega_2 x_{i2}.$$

Or in vectorised form:

$$\hat{y} = b\mathbf{1} + X\omega,$$

where

$$\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix}, \quad \mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{bmatrix}, \quad \omega = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix}.$$

The loss is still

$$L(\omega, b) = \sum (y_i - \hat{y}_i)^2 = r^T r.$$

So,

$$\frac{\partial L}{\partial \omega} = -2X^T r, \quad \frac{\partial L}{\partial b} = -2\mathbf{1}^T r = -2\sum r_i.$$

Slide 30

Now consider logistic regression. The predicted probability that $y = 1$ is

$$\hat{p}_i = \sigma(z_i) = \frac{1}{1 + e^{-z_i}}, \quad z_i = b + \omega x_i.$$

Example: if $x = 16.09$, $b = 0$, $\omega = 0.1$,

$$z = 0 + 0.1 \times 16.09 = 1.61, \quad \hat{p} = \frac{1}{1 + e^{-1.61}} = 0.833.$$

So there is an 83.3% chance that $y_i = 1$.
The new loss function is the cross-entropy:

$$L(\omega, b) = -\sum_{i=1}^{n} \Big[y_i \log \hat{p}_i + (1 - y_i) \log(1 - \hat{p}_i) \Big].$$

We compute derivatives as follows:

$$\frac{\partial L}{\partial \omega} = \sum_i (\hat{p}_i - y_i) x_i, \quad \frac{\partial L}{\partial b} = \sum_i (\hat{p}_i - y_i).$$