Math 342W/642/742W

Recitation – Day #5 (2.13.25)

I. Multinomial Classification

(i) What is the output space, $\mathcal{Y}$, for the response variable in the context of *multinomial classification* problems?

$$\mathcal{Y} = \{1, 2, \ldots, L\} \quad (\textit{a nominal categorical response})$$

(ii) What is the *null model*, g_0, in the context of *multinomial classification* problem?

$$g_0 = \text{SampleMode}[\boldsymbol{y}]$$

(iii) What is the L^1 *loss*? L^2 *loss*?

For $p = 1$,

- L1 loss $= |x_i - x_\star|$
- L2 loss $= (x_i - x_\star)^2$

For $p > 1$,

- L1 loss $= \sum_{j=1}^{p} |x_{i,j} - x_{\star,j}|$ (*Manhattan/Taxicab*)
- L2 loss $= \sum_{j=1}^{p} (x_{i,j} - x_{\star,j})^2$ (*default loss*)

II. Regression

(i) What is the output space, $\mathcal{Y}$, for the response variable in the context of *regression*?

$$\mathcal{Y} = \mathbb{R}$$

(ii) What is the *null model*, g_0, in the context of *regression*?

$$g_0(\boldsymbol{x}_*) = \bar{y}$$

(iii) What is the candidate set of functions, $\mathcal{H}$, in the context of *regression*?

$$\mathcal{H} = \{\boldsymbol{w} \cdot \boldsymbol{x} \,:\, \boldsymbol{w} \in \mathbb{R}^{p+1}\}$$

(iv) Define the following errors:

- Residual Error:

$$e_i = \textit{actual output value} - \textit{predicted output value} = y_i - \hat{y}_i = y_i - g(\boldsymbol{x}_i)$$

- Sum of Absolute Errors (SAE):

$$\text{SAE} = \sum_{i=1}^{n} |e_i| \quad \text{(L1 loss)}$$

- Sum of Squared Errors (SSE):

$$\text{SSE} = \sum_{i=1}^{n} e_i^2 \quad (\text{L2 loss}, \textit{default loss})$$

III. Simple Linear Regression

(i) Suppose $p = 1$ for linear regression, define the following:

- $\mathcal{H} := \{w_0 + w_1 x \; : \; w_0, w_1 \in \mathbb{R}\}$ = set of all linear functions in $\mathbb{R}^2$
- $\hat{y} := g(\boldsymbol{x}) = b_0 + b_1 x$ = the model given by linear regression
- Objective function: $\text{SSE} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - g(\boldsymbol{x}_i))^2$
- $\mathcal{A} := b_0, b_1 = \underset{w_0, w_1 \in \mathbb{R}}{\text{argmin}} \{\text{SSE}\} = \underset{w_0, w_1 \in \mathbb{R}}{\text{argmin}} \left\{ \sum_{i=1}^{n} (y_i - w_0 - w_1 x)^2 \right\}$

(ii) Derive the bias term b_0 in simple OLS: Set $\frac{\partial}{\partial w_0}(\text{SSE}) = 0$ and solve for w_0.

$$\begin{aligned}
\frac{\partial}{\partial w_0} \left(\sum_{i=1}^{n} (y_i - w_0 - w_1 x_i)^2 \right) &= 0 \\
\sum_{i=1}^{n} \frac{\partial}{\partial w_0} (y_i - w_0 - w_1 x_i)^2 &= 0 \\
\sum_{i=1}^{n} 2(y_i - w_0 - w_1 x_i)(-1) &= 0 \\
\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} w_0 - w_1 \sum_{i=1}^{n} x_i &= 0 \\
\sum_{i=1}^{n} y_i - n w_0 - n w_1 \bar{x} = 0 \implies w_0 = \frac{\sum_{i=1}^{n} y_i - w_1 n \bar{x}}{n} &\implies w_0 = \bar{y} - w_1 \bar{x}
\end{aligned}$$

(iii) Derive the weight b_1 in simple OLS: Set $\frac{\partial}{\partial w_1}(\text{SSE}) = 0$ and solve for w_1.

$$\begin{aligned}
\frac{\partial}{\partial w_1} \left(\sum_{i=1}^{n} (y_i - w_0 - w_1 x_i)^2 \right) &= 0 \\
\sum_{i=1}^{n} \frac{\partial}{\partial w_1} (y_i - w_0 - w_1 x_i)^2 &= 0 \\
\sum_{i=1}^{n} -2 x_i (y_i - w_0 - w_1 x_i) &= 0 \\
\sum_{i=1}^{n} (x_i y_i - w_0 x_i - w_1 x_i^2) &= 0 \\
\sum_{i=1}^{n} x_i y_i - w_0 n \bar{x} - w_1 \sum_{i=1}^{n} x_i^2 &= 0
\end{aligned}$$

We substitute expression we found above for w_0:

$$\sum_{i=1}^{n} x_i y_i - (\bar{y} - w_1 \bar{x}) n \bar{x} - w_1 \sum_{i=1}^{n} x_i^2 = 0$$

III. Simple Linear Regression (cont.)

$$\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} + w_1 n\bar{x}^2 - w_i \sum_{i=1}^{n} x_i^2 = 0$$

$$w_1 \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \right) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

$$\implies w_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

Thus, the desired weights for OLS are:

$$b_0 = \bar{y} - b_1\bar{x}; \quad b_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \quad \blacksquare$$