Math 342W/642/742W

Recitation – Day #6 (2.20.25)

I. Regression Model Performance Metrics

Write down the performance metrics we apply to regression models:

- SSE – *sum of squared errors*
  SST = SSE_0 (SSE for g_0)
- MSE – *mean square error*
- $s_e = \text{RMSE} = \sqrt{\text{MSE}}$ – *root mean squared error*
- R^2 – coefficient of determination – proportion of variance in Y explained

II. Multivariate Linear Regression

(i) Define the set of candidate functions of interest, $\mathcal{H}$.

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

(ii) Define the following terms for *multivariate linear regression* ($p > 1$):

- $\mathbb{D} = \langle X, \boldsymbol{y} \rangle$
- X : the $n \times (p+1)$ design matrix with the vector $\mathbf{1}_n$ as its first column
- $\hat{\boldsymbol{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = X\boldsymbol{w}$ (*vector of predicted values*)
- $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ (*response vector*)
- $\boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix}$ (*error vector*)
- $\boldsymbol{w}$: vector of unknown weights

(iii) Derive the expression for *sum of squared errors* (SSE) that will be the objective function:

$$\begin{aligned}
\text{SSE} &= \sum_{i=1}^{n} e_i^2 \\
&= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\
&= \boldsymbol{e}^T\boldsymbol{e} \\
&= (\boldsymbol{y} - \hat{\boldsymbol{y}})^T(\boldsymbol{y} - \hat{\boldsymbol{y}}) \\
&= \boldsymbol{y}^T\boldsymbol{y} - 2\hat{\boldsymbol{y}}^T\boldsymbol{y} + \hat{\boldsymbol{y}}^T\hat{\boldsymbol{y}} \\
&= \boldsymbol{y}^T\boldsymbol{y} - 2(X\boldsymbol{w})^T\boldsymbol{y} + (X\boldsymbol{w})^T(X\boldsymbol{w}) \\
&= \boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{w}^T X^T\boldsymbol{y} + \boldsymbol{w}^T X^T X\boldsymbol{w}
\end{aligned}$$

(iv) Define the goal/optimization problem in multivariate linear regression:

$$\text{Find } \boldsymbol{b} = \underset{\boldsymbol{w} \in \mathbb{R}^{p+1}}{\operatorname{argmin}}\{\text{SSE}\} \quad \text{where} \quad \boldsymbol{b} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \text{and} \quad g(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{b} \quad (\textit{used for prediction}).$$

III. Linear Algebra & Calculus Interlude

Define the relevant rules from linear algebra and calculus for multivariate linear regression:

- Rule # 0: Let $a \in \mathbb{R}$. Then

$$\frac{\partial}{\partial \boldsymbol{x}}[a] = \boldsymbol{0}_n$$

- Rule # 1: Let $\boldsymbol{a} \in \mathbb{R}^n$. Then,

$$\frac{\partial}{\partial \boldsymbol{x}}[\boldsymbol{a}^T \boldsymbol{x}] = \boldsymbol{a}$$

- Rule # 2: Let $a, b \in \mathbb{R}$. Then,

$$\frac{\partial}{\partial \boldsymbol{x}}[a\, f(\boldsymbol{x}) + b\, g(\boldsymbol{x})] = a \frac{\partial}{\partial \boldsymbol{x}}[f(\boldsymbol{x})] + b \frac{\partial}{\partial \boldsymbol{x}}[\, g(\boldsymbol{x})]$$

- Rule # 3: Let $A \in \mathbb{R}^{n \times n}$ and symmetric $(A^T = A)$. Then,

$$\frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{x}^T A \boldsymbol{x} = 2A\boldsymbol{x}$$

IV. Least Squares

(i) Derive the *normal equations* that come from the **least squares** method for fitting a linear model to a given set of training data:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{w}}[\text{SSE}] &= \boldsymbol{0}_{p+1} \\
\frac{\partial}{\partial \boldsymbol{w}}\left[\boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{w}^T X^T \boldsymbol{y} + \boldsymbol{w}^T X^T X \boldsymbol{w}\right] &= \boldsymbol{0}_{p+1} && (\textit{expression for SSE}) \\
\frac{\partial}{\partial \boldsymbol{w}}\left[\boldsymbol{y}^T\boldsymbol{y}\right] - 2\frac{\partial}{\partial \boldsymbol{w}}\left[\boldsymbol{w}^T\left(X^T\boldsymbol{y}\right)\right] + \frac{\partial}{\partial \boldsymbol{w}}\left[\boldsymbol{w}^T X^T X \boldsymbol{w}\right] &= \boldsymbol{0}_{p+1} && (\textit{Rule \#2}) \\
\boldsymbol{0}_{p+1} - 2X^T\boldsymbol{y} + 2X^T X \boldsymbol{w} &= \boldsymbol{0}_{p+1} && (\textit{Rules \#1 \& \#3}) \\
2X^T X \boldsymbol{w} &= X^T \boldsymbol{y} \\
X^T X \boldsymbol{w} &= X^T \boldsymbol{y} && (\textit{Normal Equations})
\end{aligned}$$

(ii) What condition must hold for the existence of a unique solution to the *normal equations*?

the $(p+1) \times (p+1)$ matrix $X^T X$ must be invertible $\iff \text{rank}(X) = p + 1$

(iii) Express the solution to the *normal equations*:

$$\boldsymbol{b} = \left(X^T X\right)^{-1} X \boldsymbol{y}$$