Section 05 - Regression

Al Xin

2023-03-09

Regression

Predictive regression

- **Predictive regression**: Given predictors with value $\vec{x}$, we aim to predict

$$\mu(\vec{x}) = \mathbb{E}[Y \mid \vec{X} = \vec{x}]$$

- **Regression error**: the difference between random outcome and predicted outcome

$$U(\vec{x}) = Y - \mathbb{E}[Y \mid \vec{X} = \vec{x}] = Y - \mu(\vec{x})$$

- The outcome Y is signal $\mu(\vec{x})$ plus noise $U(\vec{x})$
- Properties
  - $\mathbb{E}[U(\vec{X}) \mid \vec{X} = \vec{x}] = 0$
  - $Cov(U(\vec{X}), \vec{X}) = 0$
- **R-squared**: the proportion of variation in Y accounted for by variation in prediction

$$R^2 = \frac{Var(\mu(\vec{X}))}{Var(Y)} = 1 - \frac{Var(U(\vec{X}))}{Var(Y)}$$

Linear regression

- **Linear regression**: the regression function is a linear function of parameters $\vec{\theta}$

$$\mu(\vec{x}) = \mathbb{E}[Y \mid \vec{X} = \vec{x}.\vec{\theta}] = \theta_0 + \theta_1 x_1 + \cdots + \theta_K x_K$$

- Linear in parameters, not predictors
- **Homoskedasticity**: $Var(U_j \mid \vec{X} = \vec{x}) = \sigma^2$ for all j
- **Residual**: $\hat{U}_j = Y_j - \mu(\vec{x})$
- **Residual sum of squares (RSS)**: $RSS = \sum_{j=1}^{n} \hat{U}_j^2$

Logistic regression

- Use for binary response

$$\mu(x) = P(Y = 1|X = x) = \frac{e^{\theta x}}{1 + e^{\theta x}}$$

- This is the sigmoid function
- Likelihood function given as

$$L(\theta) = \prod_{j=1}^{n} \left(\frac{e^{\theta x_j}}{1 + e^{\theta x_j}} \right)^{y_j} \left(\frac{1}{1 + e^{\theta x_j}} \right)^{1-y_j}$$

- No closed-form solution for the MLE

Statistical models for predictive regression

- MLE for predictive regression

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \sum_{j=1}^{n} \log f(y_j \mid \vec{X}_j = \vec{x}_j, \vec{\theta})$$

- In Gaussian linear regression (noise is Gaussian) where $Y_j \mid X_j = x_j \sim \mathcal{N}(\theta x_j, \sigma^2)$

$$\hat{\theta}_{MLE} = \frac{\sum_{j=1}^{n} x_j Y_j}{\sum_{j=1}^{n} x_j^2}$$

Least squares

- Minimize RSS

$$\hat{\theta}_{LS} = \arg\min_{\theta} \left(\sum_{j=1}^{n} (Y_j - \theta x_j)^2 \right) = \frac{\sum_{j=1}^{n} x_j Y_j}{\sum_{j=1}^{n} x_j^2}$$

- Same as MLE for Gaussian

Problems

1 Which of the following is a linear regression?

1. $\mu(\vec{x} \mid \vec{\theta}) = \theta_1 \exp(x_1 + x_2) + \theta_2 x_1$
2. $\mu(\vec{x} \mid \vec{\theta}) = \theta_1 \theta_2 x_1$
3. $\mu(\vec{x} \mid \vec{\theta}) = \theta_1 x_1 x_2 + \theta_2 x_2 x_3$
4. $\mu(\vec{x} \mid \vec{\theta}) = \exp(\theta_1 x_1 + \theta_2 x_2)$
5. $\mu(\vec{x} \mid \vec{\theta}) = \theta_1 (2x_1 + 4)$

2 Linear regression model

Assume a model

$$Y_j \mid X_j = x_j, \vec{\theta} \sim \mathcal{N}(\theta_0 + \theta_1 x_j, \sigma^2)$$

Assume that σ^2 is known.

2.1 MLE Let $\hat{\theta}_0, \hat{\theta}_1$ be the MLEs for the parameters. What are the MLEs?

2.2 MLE distribution What is the distribution of $(\hat{\theta}_0, \hat{\theta}_1)$?

2.3 Independence of sample mean and MLE Show that $\bar{Y} \perp \hat{\theta}_1$. Hint: Ex. 7.5.9 in STAT 110.

2.4 Find a 95% CI for $\mu(x_0) = \theta_0 + \theta_1 x_0$ Hint: construct a pivot based on $\hat{Y}$.