Linear Algebra Review

Let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$ be a $m \times n$ matrix and $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix}$ a column vector of length n.

1. Write AX in two ways: in terms of column vectors of A and in terms of row vectors of A.

2. For compatible matrices A and B, write AB in terms of A and column vectors B_i of B.

Linear Regression

Suppose some real-valued outcome y depends on p covariates $x_1, x_2, ..., x_p$ with random noise. That is, for $y \in \mathbb{R}$ and $X \in \mathbb{R}^p$:

$$y = f(X) + \epsilon \quad \text{for some } f : \mathbb{R}^p \to \mathbb{R} \tag{1}$$

In linear regression, we approximate a model of f with a linear function of the covariates $x_1, ..., x_p$:

$$y = f(X) + \epsilon = \beta_0 + \sum_{j=1}^{p} \beta_j x_j + \epsilon \tag{2}$$

Given n data points, $(y^{(1)}, X^{(1)}), (y^{(2)}, X^{(2)}), ..., (y^{(n)}, X^{(n)})$, we want to estimate the true linear function f. One natural way to do this is to find a $\hat{\beta} \in \mathbb{R}^{p+1}$ that minimizes the **R**esidual **S**um of **S**quares:

$$\hat{\beta} \quad \text{minimizes} \quad RSS(\beta) = \sum_{i=1}^{N} (y^{(i)} - \beta_0 - \sum_{j=1}^{p} \beta_j x_j^{(i)})^2 \tag{3}$$

Letting $Y = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(n)} \end{bmatrix}$ and $X = \begin{bmatrix} 1 & X^{(1)^T} \\ \vdots & \vdots \\ 1 & X^{(n)^T} \end{bmatrix}$, this can be rewritten as:

$$\hat{\beta} \quad \text{minimizes} \quad RSS(\beta) = (Y - X\beta)^T (Y - X\beta) \tag{4}$$

Since we want to minimize $RSS(\beta)$, we take the derivative [1]:

$$\frac{\partial RSS}{\partial \beta} = 2(Y - X\beta)^T(-X) \tag{5}$$

and set it to $\overrightarrow{0}$, the zero vector, to find $\hat{\beta}$ [2]:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{6}$$

With our $\hat{\beta}$, if we wanted to predict Y given X we would estimate $Y \approx \hat{Y} = X\hat{\beta}$.

3. Based on question 1, can you describe the subspace $\hat{Y}$ lives in?

4. $(Y - \hat{Y})$ is the residual, the error between Y and its approximation $\hat{Y}$. Since $\hat{\beta}$ was found by setting (5) to $\overrightarrow{0}$, what is the dot product between $(Y - \hat{Y})$ and any column of X? Question 2 might be helpful.

5. Putting together questions 3 and 4, what can you say about the geometric relationship between $(Y - \hat{Y})$ and $\hat{Y}$?

[1] some books write the transpose $-2X^T(y - X\beta)$.

[2] we assumed $X^T X$ is invertible, which is true if and only if X has full column rank. If X doesn't have full column rank, then $X^T X$ is not invertible and many solutions exist

Time Series

Let's work with time series data. Suppose we have a linear trend in time: $Y_t = \beta_0 + \beta_1 t + W_t$, where W_t is white noise with variance σ^2.

6. For times $1, .., k$, calculate the $k \times k$ covariance matrix for $Y_1, ..., Y_k$.

7. Let's try to do a linear regression on y_1,...,y_n for $n \geq 2$ with t as our only covariate. Show that $\hat{\beta}$ is an unbiased estimator of (β_0, β_1), the true linear weights.

Conceptual questions / Problems

For questions 9-11, write formulas for each of the terms.

8. Expectation
   (a) $E(aX + bY + c)$
   (b) $E(f(X)g(Y))$ if X and Y are independent
9. Variance
   (a) $\text{Var}(X)$
   (b) $\text{Var}(aX + bY + c)$
10. Covariance:
    (a) $\text{Cov}(X, Y)$
    (b) $\text{Cov}(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j)$
11. Correlation
    (a) $\text{Corr}(X, Y)$
    (b) Write down bounds for $\text{Corr}(X, Y)$

R Code

In the *astsa* library, there is a dataset on the price of chicken, called *chicken*. We will use that as our example here.

12. Load the *astsa* package and plot the *chicken* dataset.
13. What type of trend do you see?
14. Use the *lm()* function to fit this trend and add the trend line to the plot.
15. How does your trendline look?
16. Plot the residuals over time. Do they appear stable/stationary? If not, what might you do instead of or in addition to your current model?