Linear Algebra Primer

Matrices / Vectors A matrix is an $n \times p$ object. Matrices are often denoted by a capital letter (or Greek symbol). A few common matrices will be

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix}$$

or

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \ddots & \sigma_n^2 \end{pmatrix}$$

or

$$J_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ddots & 1 \end{pmatrix}$$

or

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 \end{pmatrix}$$

Vectors are essentially one-dimension vectors and will be denoted with an underline. We will assume vectors are $q \times 1$ dimension unless noted with a transpose.

$$\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

or

$$\underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$$

or

$$\underline{1}_n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

The transpose operator will be denoted by $\underline{y}^T = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix}$ or $\underline{y}^{'}$, both of which would result in a $1 \times n$ vector.

Matrix Multiplication The most important component in matrix multiplication is tracking dimensions.

Consider a simple case with

$$\underline{\hat{y}} = X \times \underline{\hat{\beta}},$$

where X is a 2×2 matrix, $\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}$ and $\underline{\hat{\beta}} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$.

Then

$$\underline{\hat{y}} = \begin{bmatrix} 1 \times 3 + 2 \times 2 \\ 1 \times 3 + (-1) \times 2 \end{bmatrix} = \begin{bmatrix} 7 \\ -1 \end{bmatrix}$$

In R, we use `%*%` for matrix multiplication.

```
X <- matrix(c(1,2, 1 ,-1), nrow = 2, ncol = 2, byrow = T); X
```

```
##      [,1] [,2]
## [1,]    1    2
## [2,]    1   -1
```

```
#X <- matrix(c(1, 1, 2 ,-1), nrow = 2, ncol = 2); X
beta_hat <- matrix(c(3,2),nrow =2, ncol = 1); beta_hat
```

```
##      [,1]
## [1,]    3
## [2,]    2
```

```
y_hat <- X %*% beta_hat; y_hat
```

```
##      [,1]
## [1,]    7
## [2,]    1
```

Kronecker Product Kronecker product $\otimes$, enables a different type of matrix multiplication.

If

$$A = \begin{bmatrix} 1 & 2 \end{bmatrix}$$

and $B = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$, then

$$A \otimes B = \begin{bmatrix} 3 & 6 \\ 4 & 8 \\ 5 & 10 \end{bmatrix}$$

More about matrices

Nonsingular matrix and Matrix Inverse If X is a square matrix, then X is nonsingular if there exists another matrix s.t.

$XX^{-1} = I$ and $X^{-1}X = I$

X^{-1} is the inverse of a matrix. We can calculate the inverse of a matrix for a 1×1 matrix, perhaps as 2×2, matrix and maybe even a 3×3 matrix. However, beyond that it is quite challenging and time consuming. Furthermore, it is also (relatively) time intensive for your computer.

Orthogonal matrices If a matrix X has an inverse that is also the transpose, $XX^T = I$, then X is an orthogonal matrix.

Motivating Dataset: Washington (DC) housing dataset Hopefully the connections to statistics are clear, using X and β, but let's consider a motivating dataset.

This dataset contains housing information from Washington, D.C. It was used for a STAT532 exam, so apologize in advance for any scar tissue.

```
DC <- read_csv('https://math.montana.edu/ahoegh/teaching/stat532/data/DC.csv')
```

```
## Parsed with column specification:
## cols(
##   BATHRM = col_double(),
##   HF_BATHRM = col_double(),
##   AC = col_character(),
##   BEDRM = col_double(),
##   STORIES = col_double(),
##   PRICE = col_double(),
##   CNDTN = col_character(),
##   LANDAREA = col_double(),
##   FULLADDRESS = col_character(),
##   ASSESSMENT_NBHD = col_character(),
##   WARD = col_character(),
##   QUADRANT = col_character()
## )
```

```
DC %>% group_by(WARD) %>%
  summarize(`Average Price (millions of dollars)` = mean(PRICE)/1000000, .groups = 'drop') %>%
  kable(digits = 3)
```

WARD	Average Price (millions of dollars)
Ward 1	0.879
Ward 2	1.919
Ward 3	1.294
Ward 4	0.693
Ward 5	0.592
Ward 6	0.856
Ward 7	0.321
Ward 8	0.306

```
DC %>% group_by(BEDRM) %>%
  summarize(`Average Price (millions of dollars)` = mean(PRICE)/1000000, .groups = 'drop') %>%
  kable(digits = 3)
```

BEDRM	Average Price (millions of dollars)
0	0.195
1	0.442
2	0.479
3	0.620
4	0.832
5	1.355
6	1.849
7	1.666
9	7.365

Regression Model

There are many factors in this dataset that can are useful to predict housing prices.

$$y_i = \beta_0 + \beta_1 * x_{SQFT,i} + \beta_2 x_{BEDRM,i} + \epsilon_i, \tag{1}$$

where y_i is the sales price of the i^{th} house, $x_{SQFT,i}$ is the living square footage of the i^{th} house, and $x_{BEDRM,i}$ is the number of bedrooms for the i^{th} house. Note this implies that we are treating bedrooms as continuous variables as opposed to categorical.

we usually write $\epsilon_i \sim N(0, \sigma^2)$. More on that soon.

In R we often write something like: `price ~  LANDUSE + BEDRM`.

Now let's write this model in matrix notation:

$$\underline{y} = X\underline{\beta} + \underline{\epsilon}, \tag{2}$$

where $\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_{SQFT,1} & x_{BEDRM,1} \\ 1 & x_{SQFT,2} & x_{BEDRM,2} \\ \vdots & \vdots & \vdots \\ 1 & x_{SQFT,n} & x_{BEDRM,n} \end{bmatrix}$,$\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$, and $\underline{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$

Now what are the implications of:

$\epsilon_i \sim N(0, \sigma^2)$ or $\underline{\epsilon} \sim N(\underline{0}, \Sigma)$, where

$$\Sigma = \sigma \times \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

These are equivalent statements and both imply that y_i and y_j are conditionally independent given X. In other words, after controlling for predictors (ward, square footage), then the price of house i gives us no additional information about price of house j.