Multicollinearity

San-Teng Huang,Shang-Chien Ho

National Dong Hwa University

2018/12/19

Outline

Multiple Linear Regression

- The coefficients of the estimates:

$$\hat{\beta} = \left(X^{\intercal}X\right)^{-1}X^{\intercal}y$$

- The variance of the estimates:

$$Var\left[\hat{\beta}\right] = \sigma^{2}\left(X^{\intercal}X\right)^{-1}$$

If that matrix isn't exactly singular, but is close to being non-invertible, the variances will become huge.

Collinearity

There are several equivalent conditions for any square matrix, say u, to be singular or non-invertible:

- The determinant det u or $|u|$ is 0.
- At least one eigenvalue of u is 0.
- u is rank deficient, meaning that one or more of its columns (or rows) is equal to a linear combination of the other rows.

- Dealing with Collinearity by Deleting Variables
- Diagnosing Collinearity Among Pairs of Variables
- Geometric Perspective
- Why Multicollinearity Is Harder

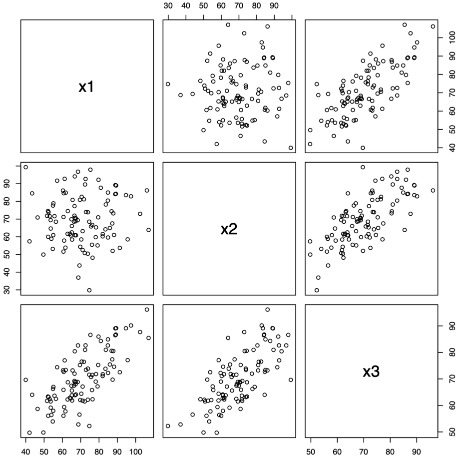

Figure: suppose X_1 and X_2 are independent Gaussians, of equal variance σ^2, and X_3 is their average, $X_3 = (X_1 + X_2)/2$

Multicollinearity

Multicollinearity means that

$$c_1X_1 + c_2X_2 + \ldots c_pX_p = \sum_{i=1}^{p} c_iX_i = c_0$$

To simplify this, let's introduce p×1 matrix a = $\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}$,so we can

write multicollinearity as $a^{\mathsf{T}}X = c_0$, for $a \neq 0$

$$Var\left[a^{\mathsf{T}}X\right] = 0, \quad a \neq 0$$

$$\begin{aligned} Var\left[a^{\mathsf{T}}X\right] &= Var\left[\sum_{i=1}^{p} c_i X_i\right] \\ &= \sum_{i=1}^{p}\sum_{j=1}^{p} c_i c_j Cov\left[X_i, X_j\right] \\ &= a^{\mathsf{T}}\, Var\left[X\right]\, a \end{aligned}$$

- The eigenvectors of Var [X], such that

$$Var\left[X\right] v_i = \lambda v_i$$

- The eigenvalues are all 0.
- Any vector can be re-written as a sum of eigenvectors:

$$a = \sum_{i=1}^{p} \left(a^\mathsf{T} v_i\right) v_i$$

- The eigenvectors can be chosen so that they all have length 1 and are orthogonal to each other.
- Var [X] can be expressed as

$$Var\left[X\right] = VDV^\mathsf{T}$$

$$
\begin{aligned}
Var[X]\,a &= Var[X]\sum_{i=1}^{p}\left(a^{\mathsf{T}}v_i\right)v_i \\
&= \sum_{i=1}^{p}\left(a^{\mathsf{T}}v_i\right)Var[X]\,v_i \\
&= \sum_{i=1}^{p}\left(a^{\mathsf{T}}v_i\right)\lambda_i v_i \\
a^{\mathsf{T}}Var[X]\,a &= \left(\sum_{i=1}^{p}\left(a^{\mathsf{T}}v_i\right)v_j\right)^{\mathsf{T}}\sum_{i=1}^{p}\left(a^{\mathsf{T}}v_i\right)\lambda_i v_i \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p}\left(a^{\mathsf{T}}v_i\right)\left(a^{\mathsf{T}}v_i\right)v_j^{\mathsf{T}}v_i \\
&= \sum_{i=1}^{p}\left(a^{\mathsf{T}}v_i\right)^2\lambda_i
\end{aligned}
$$

- The predictors are multi-collinear if and only if Var [X] has zero eigenvalues.
- Every multi-collinear combination of the predictors is either an eigenvector of Var [X] with zero eigenvalue, or a linear combination of such eigenvectors.