

Tutorial Notes 9 of MATH3424

1 Summary of course material

1. Collinearity: the existence of strong linear relationships among the predictor variables

Typical indications of collinearity

- (a) High R^2 , but low p-value for the coefficient estimates.
- (b) Significant estimates F statistic, but low p-value for the coefficient estimates.
- (c) If the estimated coefficient is contrary to prior expectation (which people are confident to be correct).

2. Detection of Collinearity:

Variance Inflation Factor (VIF): Let R_j^2 be the square of the multiple correlation coefficient that results when the predictor variable X_j is regressed against all the other predictor variables. The variance inflation for X_j is

$$\text{VIF}_j = \frac{1}{1 - R_j^2}, \quad j = 1, \dots, p$$

- (a) $\text{VIF}_j \rightarrow 1$, X_j is linearly independent of other predictor variables.
- (b) VIF_j large, X_j has a strong linear relationship with the other predictor variables
- (c) VIF exceeding 10 would imply that collinearity may be causing problems.
- (d) $\overline{\text{VIF}} = \frac{1}{p} \sum_{j=1}^p \text{VIF}_j$: another index of collinearity

3. Ridge Regression:

- (a) Assume the standardized form of the regression model is given by

$$\tilde{Y} = \theta_1 \tilde{X}_1 + \theta_2 \tilde{X}_2 + \dots + \theta_p \tilde{X}_p + \varepsilon'$$

The estimating equations for ridge regression coefficients are

$$\begin{array}{ccccccc} (1+k)\theta_1 & + & r_{12} \theta_2 & + & \dots & + & r_{1p} \theta_p & = & r_{1y}, \\ r_{21} \theta_1 & + & (1+k)\theta_2 & + & \dots & + & r_{2p} \theta_p & = & r_{2y}, \\ & & \vdots & & \vdots & & \vdots & & \vdots \\ r_{p1} \theta_1 & + & r_{p2} \theta_2 & + & \dots & + & (1+k)\theta_p & = & r_{py}, \end{array}$$

where k is the bias parameter, r_{ij} is the correlation between i th and j th predictor variables and r_{ij} is the correlation between i th predictor variable and response variable.

The solution is given by

$$\hat{\theta} = \left(\frac{\tilde{X}'\tilde{X}}{n-1} + kI_p \right)^{-1} \frac{\tilde{X}'\tilde{Y}}{n-1}$$

It is a biased estimator of θ . (Why?)

(b)

$$\text{Total Variance}(k) = \sum_{j=1}^p \text{Var}(\hat{\theta}_j(k)) = \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2}$$

(c) Ridge trace to detect collinearity

(d) Selection of bias parameter k : Fixed Point/Iterative Method/Ridge Trace

Remarks on Ridge Regression:

Ridge regression estimates can be obtained by the method of penalized least squares (Why?). The penalized least squares criterion combines the usual sum of squared errors with a penalty for large regression coefficients:

$$S(\theta_1, \dots, \theta_p) = \sum_{i=1}^n \left(\tilde{y}_i - \sum_{j=1}^p \theta_j \tilde{x}_{ij} \right)^2 + (n-1)k \sum_{j=1}^p \theta_j^2$$

2 Questions

2.1

Chemical shipment. The data to follow, taken on 20 incoming shipments of chemicals in drums arriving at a warehouse, show number of drums in shipment (X_1), total weight of shipment (X_2 , in hundred pounds), and number of minutes required to handle shipment (Y).

i :	1	2	3	...	18	19	20
X_{i1} :	7	18	5	...	21	6	11
X_{i2} :	5.11	16.72	3.20	...	15.21	3.64	9.57
Y_i :	58	152	41	...	155	39	90

- (a) Fit the original data by OLS and find the fitted values.

Call:

```
lm(formula = Y ~ ., data = df)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-8.8353	-3.5591	-0.0533	2.4018	15.1515

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.3243	3.1108	1.069	0.3
X1	3.7681	0.6142	6.135	1.10e-05 ***
X2	5.0796	0.6655	7.632	6.89e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.618 on 17 degrees of freedom

Multiple R-squared: 0.9869, Adjusted R-squared: 0.9854

F-statistic: 641.6 on 2 and 17 DF, p-value: < 2.2e-16

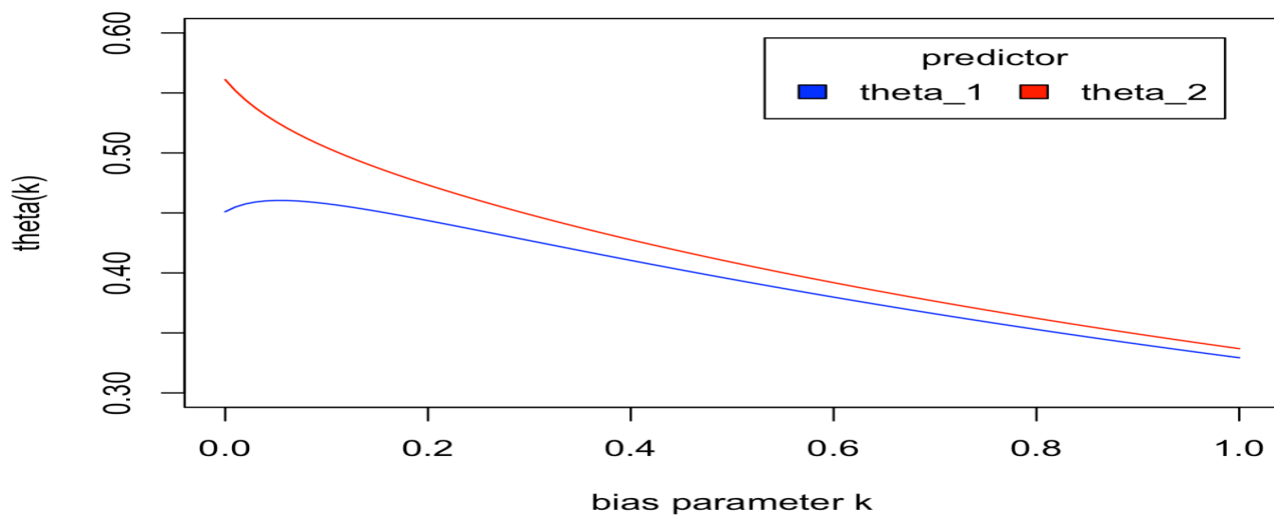
Figure 1: summary of linear regression by OLS

1	2	3	4	5	6	7	8	9	10	11	12	13
55.65775	156.08097	38.41952	91.78733	100.54737	42.68637	202.09731	72.94678	117.55927	46.34368	123.35921	96.84849	45.18459
14	15	16	17	18	19	20						
81.30495	56.83527	130.24902	133.56918	159.71512	44.42265	93.38515						

Figure 2: fitted value by OLS

(b) Make a ridge trace plot for k values ranging from 0 to 1 with standardized data.

```
df <- read.table("T9.txt", header = FALSE)
colnames(df) <- c("Y", "X1", "X2")
df_scale <- as.data.frame(scale(df))
X <- as.matrix(df_scale[,-1])
Y <- as.matrix(df_scale[,1])
n <- length(Y); p <- 2
k_seq <- seq(0,1,length.out = 101)
theta_res <- matrix(0, length(k_seq), p)
for (i in 1:101){
  theta_res[i,c(1,2)] <- solve(t(X)%*%X/(n-1)+k_seq[i]*diag(p))%*%(t(X)%*%Y/(n-1))
}
theta_res <- cbind(k_seq, theta_res)
plot(theta_res[,1], theta_res[,2], type = "l", col = "blue",
      ylim = c(0.3, 0.6), xlab = "bias parameter k", ylab = "theta(k)")
lines(theta_res[,1], theta_res[,3], type = "l", col = "red")
legend("topright", inset = .05, title = "predictor",
      c("theta_1", "theta_2"), fill = c("blue", "red"), horiz = TRUE)
```



Now given below are the estimated ridge standardized regression coefficients, the variance inflation factors, and R^2 for selected bias parameters k .

k	.000	.005	.01	.05	.07	.09	.10	.20
$\hat{\theta}_1$.451	.453	.455	.460	.460	.459	.458	.444
$\hat{\theta}_2$.561	.556	.552	.526	.517	.508	.504	.473
$VIF_1 = VIF_2$	7.03	6.20	5.51	2.65	2.03	1.61	1.46	.71

(c) Why are the VIF_1 values the same as the VIF_2 values here?

(d) Suggest a reasonable value for the bias parameter k based on the information provided above. What is the estimation of k given by fixed point method?

- (e) Transform the estimated standardized regression coefficients using the bias parameter you suggest in part (d) back to the original variables and obtain the fitted values for the 20 cases. How similar are these fitted values to those obtained with the ordinary least squares fit in part (a)?

```
m_simple <- lm(Y~.,df)
summary(m_simple)
std <- diag(sqrt(var(df)))
mean_XY <- colMeans(df)
ratio <- std[1]/std[-1]
theta_est <- theta_res[theta_res[,1]==0.05,-1]
beta_est <- ratio*theta_est
beta_0_est <- mean_XY[1]-sum(beta_est*mean_XY[-1])
X_o <- as.matrix(df[, -1])
Y_o <- as.matrix(df[, 1])
fitted_ridge <- as.numeric(X_o%%beta_est+beta_0_est)
names(fitted_ridge) <- c(1:length(fitted_ridge))
fitted_ridge
```

Figure 3: fitted value by ridge

1	2	3	4	5	6	7	8	9	10	11	12	13
55.65775	156.08097	38.41952	91.78733	100.54737	42.68637	202.09731	72.94678	117.55927	46.34368	123.35921	96.84849	45.18459
14	15	16	17	18	19	20						
81.30495	56.83527	130.24902	133.56918	159.71512	44.42265	93.38515						

Figure 4: fitted value by OLS

1	2	3	4	5	6	7	8	9	10	11	12	13
56.54120	154.14860	39.75123	92.61263	99.88431	43.75176	198.86197	73.37879	117.40364	47.18078	123.15542	96.73016	46.40789
14	15	16	17	18	19	20						
82.15683	57.95908	128.98746	132.72810	158.49749	45.69351	93.16915						

Figure 5: fitted value by ridge

2.2

We know from Chapter 3 that in standard multiple linear regression $\text{Cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}$. Show that $\text{Var}(\hat{\beta}_j) \propto \sigma^2 \cdot \text{VIF}_j$ for $j = 1, \dots, p$.

Hint: You might need to use the following matrix inverse formula for block matrix:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}$$