Fall 2021 — STAT 2221: Advanced Applied Multivariate Analysis

Homework 2. Due by Thursday, September 23, 2021 at 9:00 a.m. EST

Topic: Multivariate Regression

Name: Giang Vu

Additional instructions. Add `\newpage` immediately before each problem so that each has its own page. Add `\begin{proof}[Solution.] ... \end{proof}` below each problem for providing your solution. You are welcome to add additional packages to the preamble, but do not modify the existing commands and formatting.

Problem 2.1. Perform the following regression analyses using the Norwegian paper quality data described in Example 6.2.2 of Izenman, available on Canvas (and retrieved via the website `http://lib.stat.cmu.edu/datasets/papir`).

1. First, write a few lines of code to carry out multivariate ridge regression (from scratch).
2. Apply your code in (1) to the data. Then compare results against running separate univariate ridge regressions on the data. Discuss your choices and findings.
3. Next, run reduced rank regression on the data. Plot the rank trace using $\mathbf{\Gamma} = \mathbf{I}_s$ as the weight matrix. Estimate the effective dimensionality of the multivariate regression. Discuss your choices and findings.
4. Compare the estimate of effective dimensionality in (3) with one obtained using cross validation.

Proof. □

1. Below, I defined a function to calculate coefficient estimates the multivariate ridge regression model based on the formula (6.41) in page 168, Izenman.
   This function requires input X_c as the centered design matrix, $vec(Y_c)$ as the vectorized centered response variables, s as the number of response Y, r as the number of independent variables, annd k as the ridge penalty term/regularization parameter.

```
#define a function to find theta hat, formula (6.41), pg. 168 Izenman
mv.ridge <- function(Xc, vecYc, s, r, k){
  I.s <- diag(s)
  I.r <- diag(r)
  K <- k*diag(s) #penalty term
  v.thetah <- solve(kronecker(Xc %*% t(Xc),I.s)+kronecker(I.r,K)) %*% kronecker(Xc, I.s) %*% vecYc
  return(matrix(v.thetah, nrow = s, ncol = r))
}
```

2. To apply the formula in part 1 on the Norway paper data, I first separated and reshaped the data to get X_c and $vec(Y_c)$ like the function requires.
   In our data, $s = 13$, $r = 9$, and $n = 29$. I chose a random value for $k = 0.5$ just to test if the formula works (Note: we could incorporate methods to find the optimal value of k here such as ridge trace, cross validation, etc.)
   The result is a matrix with dimension (13 x 9) that contains the multivariate ridge regression coefficient estimates.
   After doing that, I carried out 13 separate univariate ridge regression of each of the Y variable, and compare the ridge coefficient estimate with the multivariate method above. And the results are exactly the same, meaning that we have seen that the classical approach to multivariate (ridge) regression (stacking all the Y's on top of each other), is essentially the same as doing separate regression for each of the Y's.

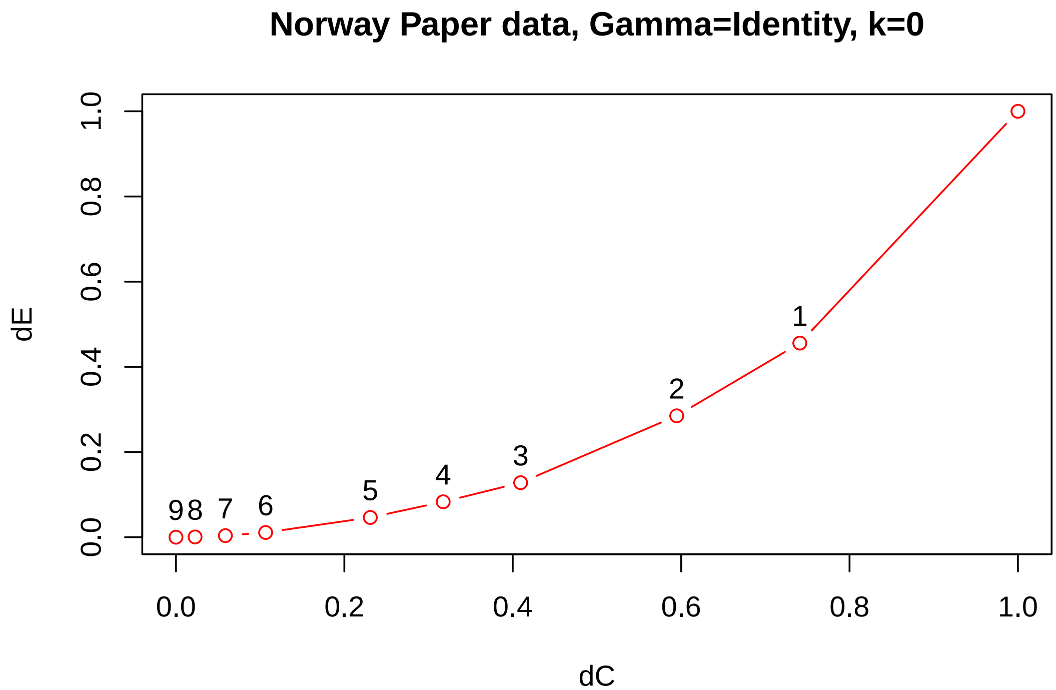
```
## [1] "This is the estimates by multivariate ridge regression model"

##             [,1]        [,2]       [,3]          [,4]        [,5]        [,6]
##  [1,]  0.7294545 -0.43425555 -0.3532010  0.09643224173 -0.27453442  0.51212440
##  [2,] -0.8183879  0.33970787  0.3583139 -0.03513833613  0.21509459 -0.53109365
##  [3,]  0.2771576 -0.64546497 -0.5504286  0.03586302661 -0.26402781  0.50457219
##  [4,]  0.4818424 -0.47305130 -0.6419119  0.12911248811 -0.36546944  0.42596586
##  [5,]  0.4993030  0.13550540 -0.5558461 -0.16817696194 -0.35246648  0.39307469
##  [6,] -0.6950667  0.03973841 -0.2068919  0.32393218809 -0.33881089  0.05290675
##  [7,]  0.8509273  0.04702584 -0.2628105 -0.30207379804 -0.01595304  0.21276460
##  [8,] -0.5468727  0.18831777 -0.3458519  0.00001472473 -0.39976843  0.04463157
##  [9,]  0.2780242  0.47994123 -0.5838285 -0.35929749931 -0.01430570  0.78352959
## [10,] -0.6341212 -0.13932858  0.1088048  0.20957751929 -0.32848384 -0.48904855
## [11,]  0.1690788 -0.68705953  0.3244435  0.13164146450  0.13956120 -0.13107410
## [12,] -0.5098788  0.26618804  0.5168062 -0.77528370281  0.39569308  0.31183426
## [13,]  0.4897273  0.04484853  0.4094243 -0.65889842020  0.52118401  0.51698401
##              [,7]       [,8]        [,9]
##  [1,]  0.10488235  0.3671176  0.20591569
##  [2,] -0.13238824 -0.5054706 -0.30337884
##  [3,]  0.06149412  0.3843059 -0.14617318
##  [4,]  0.06396471  0.1838706  0.02505337
##  [5,] -0.14911765  0.7455529  0.56128695
##  [6,]  0.13057647  0.4352353  0.12065692
```

```
##  [7,] -0.22870588  0.1186000  0.23528546
##  [8,]  0.22220000 -0.4197059  0.24211096
##  [9,]  0.13491765  0.3578588 -0.11841997
## [10,]  0.05351765 -0.6422235  0.31823275
## [11,]  0.08060000 -0.7773529 -0.67729203
## [12,]  0.20130588  0.1174235  0.14767090
## [13,] -0.21777647 -0.2303294 -0.18097639
```

```
## [1] "This is the estimates by separate univariate ridge regression models"
```

```
##             [,1]        [,2]       [,3]           [,4]        [,5]        [,6]
##  [1,]  0.7294545 -0.43425555 -0.3532010  0.09643224173 -0.27453442  0.51212440
##  [2,] -0.8183879  0.33970787  0.3583139 -0.03513833613  0.21509459 -0.53109365
##  [3,]  0.2771576 -0.64546497 -0.5504286  0.03586302661 -0.26402781  0.50457219
##  [4,]  0.4818424 -0.47305130 -0.6419119  0.12911248811 -0.36546944  0.42596586
##  [5,]  0.4993030  0.13550540 -0.5558461 -0.16817696194 -0.35246648  0.39307469
##  [6,] -0.6950667  0.03973841 -0.2068919  0.32393218809 -0.33881089  0.05290675
##  [7,]  0.8509273  0.04702584 -0.2628105 -0.30207379804 -0.01595304  0.21276460
##  [8,] -0.5468727  0.18831777 -0.3458519  0.00001472473 -0.39976843  0.04463157
##  [9,]  0.2780242  0.47994123 -0.5838285 -0.35929749931 -0.01430570  0.78352959
## [10,] -0.6341212 -0.13932858  0.1088048  0.20957751929 -0.32848384 -0.48904855
## [11,]  0.1690788 -0.68705953  0.3244435  0.13164146450  0.13956120 -0.13107410
## [12,] -0.5098788  0.26618804  0.5168062 -0.77528370281  0.39569308  0.31183426
## [13,]  0.4897273  0.04484853  0.4094243 -0.65889842020  0.52118401  0.51698401
##              [,7]       [,8]        [,9]
##  [1,]  0.10488235  0.3671176  0.20591569
##  [2,] -0.13238824 -0.5054706 -0.30337884
##  [3,]  0.06149412  0.3843059 -0.14617318
##  [4,]  0.06396471  0.1838706  0.02505337
##  [5,] -0.14911765  0.7455529  0.56128695
##  [6,]  0.13057647  0.4352353  0.12065692
##  [7,] -0.22870588  0.1186000  0.23528546
##  [8,]  0.22220000 -0.4197059  0.24211096
##  [9,]  0.13491765  0.3578588 -0.11841997
## [10,]  0.05351765 -0.6422235  0.31823275
## [11,]  0.08060000 -0.7773529 -0.67729203
## [12,]  0.20130588  0.1174235  0.14767090
## [13,] -0.21777647 -0.2303294 -0.18097639
```

3. The rank trace does provide a good idea of which value of t to choose. If we're using the "first elbow" rule, then I would say rank $t = 3$ would be a good choice, but if we want "second elbow" to include more information, we can choose $t = 6$, which does offer some more reduction in both dE and dC compared to $t = 3$. But after 6, there isn't any more room for improvement that's worth the extra complexity of higher rank models.



4. Below is the result of applying CV (5-fold, 10-fold, and LOO-CV) to assess the value of t. Average prediction error at each rank is calculated, and looking at the table, we could see that $t = 3$ is a good choice, because prediction error is minimized here. If we move forward with greater value for t, the prediction error actually increases and that's not desirable.

```
##          CV/5     CV/10       CV/n
## t=1 0.9436477 0.2604708 0.02901161
## t=2 0.9168949 0.2588325 0.02874695
## t=3 0.8619515 0.2438087 0.02703860
## t=4 0.9591024 0.2700881 0.03019063
## t=5 0.9990936 0.2785950 0.03125241
## t=6 0.9703622 0.2734133 0.03030610
## t=7 0.9923134 0.2766903 0.03073808
## t=8 0.9937230 0.2785226 0.03090053
## t=9 0.9967327 0.2786013 0.03092222
```

Problem 2.2. Establish the following.

1. Show that the eigenvalues of the matrix $\Sigma_{YY}^{-1/2}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1/2}$ necessarily lie between zero and one.

2. Consider the weighted sum-of-squares reduced rank regression objective function

$$W(t) = \mathbb{E}_{(X,Y)}\|Y - \mu - A^{(t)}B^{(t)}X\|_{\Gamma}^2.$$

Show that the above criterion is invariant under nonsingular transformations of the form $X' = \Psi + \Lambda X$ and $Y' = \Phi + \Delta Y$ for any invertible matrices Λ and Δ with appropriate dimensions.

Proof. 1. Let $R = \Sigma_{YY}^{-1/2}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1/2}$.

To show this, we need to prove that the eigenvalues of R are both non-negative and not greater than 1.

First, to show that the eigenvalues of R are non-negative, we can show R is a positive semi-definite matrix.

$$\begin{aligned} R &= \Sigma_{YY}^{-1/2}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1/2} \\ \Sigma_{YY}^{1/2}R\Sigma_{YY}^{1/2} &= \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} \\ \Sigma_{YY} - \Sigma_{YY}^{1/2}R\Sigma_{YY}^{1/2} &= \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} \\ \Sigma_{YY}^{1/2}(\mathbb{I}_s - R)\Sigma_{YY}^{1/2} &= \Sigma_{Y|X} \end{aligned}$$

The right hand side is the covariance matrix for the conditional distribution of Y given X, so this is a positive semi-definite matrix, making the left hand side positive semi-definite as well.

On the left hand side, we have Σ_{YY} as the covariance matrix for Y, so it's positive semi-definite, making $\Sigma_{YY}^{1/2}$ positive semi-definite as well.

As a result, the matrix $\mathbb{I}_s - R$ has to be positive semi-definite, and with the identity $\mathbb{I}_s$ being positive semi-definite, that means our matrix R must also be a positive semi-definite matrix.

Therefore, all of its eigenvalues must be non-negative.

Secondly, to show that the eigenvalues of R are less than or equal to 1, we can also use the fact that $\mathbb{I}_s - R$ is positive semi-definite.

$\mathbb{I}_s - R$ is positive semi-definite so all of its eigenvalues must be non-negative, but the eigenvalue of the identity $\mathbb{I}_s$ is just 1, so that means the eigenvalues of R must be no greater than 1.

As a result, the eigenvalues of $R = \Sigma_{YY}^{-1/2}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1/2}$ are between 0 and 1.

2.

$$\begin{aligned} W(t) &= \mathbb{E}_{(X,Y)}\|Y - \mu - A^{(t)}B^{(t)}X\|_{\Gamma}^2 \\ &\geq \mathbb{E}\{(Y_c - CX_c)^T\Gamma(Y_c - CX_c)\} \end{aligned}$$

where X_c and Y_c are centered versions of X and Y, and $C = AB$, based on equation (6.79), p. 178 Izenman. With $X' = \Psi + \Lambda X$ and $Y' = \Phi + \Delta Y$, we will have:

$$X'_c = X' - \mathbb{E}(X') = \Psi + \Lambda X - \Psi - \Lambda \mathbb{E}(X) = \Lambda X_c$$
$$Y'_c = Y' - \mathbb{E}(Y') = \Phi + \Delta Y - \Phi - \Delta \mathbb{E}(Y) = \Delta Y_c$$

Also,

$$\begin{aligned}
\mathbb{E}(X_c'^T X_c') &= \mathbb{E}(X_c^T \Lambda^T \Lambda X_c) \\
&= tr(\Lambda \Sigma_{XX} \Lambda^T) \\
\rightarrow \Sigma_{X'X'} &= \Lambda \Sigma_{XX} \Lambda^T
\end{aligned}$$

Similarly, we also have:

$$\begin{aligned}
\Sigma_{Y'Y'} &= \Delta \Sigma_{YY} \Delta^T \\
\Sigma_{X'Y'} &= \Delta \Sigma_{XY} \Lambda^T \\
\Sigma_{Y'X'} &= \Lambda \Sigma_{YX} \Delta^T
\end{aligned}$$

Plug X'_c and Y'_c into the objective function we will have the following criterion.

$$\begin{aligned}
W(t) &\geq \mathbb{E}\{(Y'_c - CX'_c)^T \Gamma (Y'_c - CX'_c)\} \\
&= \mathbb{E}\{(\Delta Y_c - C\Lambda X_c)^T \Gamma (\Delta Y_c - C\Lambda X_c)\} \\
&= \mathbb{E}\{Y_c^T \Delta^T \Gamma \Delta Y_c - Y_c^T \Delta^T \Gamma C \Lambda X_c - X_c^T \Lambda^T C^T \Gamma \Delta Y_c + X_c^T \Lambda^T C^T \Gamma C \Lambda X_c\} \\
&= tr\{\Gamma^{1/2} \Delta \Sigma_{YY} \Delta^T \Gamma^{1/2} - \Gamma^{1/2} C \Delta \Sigma_{XY} \Lambda^T \Gamma^{1/2} - \Gamma^{1/2} \Lambda \Sigma_{YX} \Delta^T C^T \Gamma^{1/2} + \Gamma^{1/2} C \Lambda \Sigma_{XX} \Lambda^T C^T \Gamma^{1/2}\} \\
&= tr\{\Gamma^{1/2} \Sigma_{Y'Y'} \Gamma^{1/2} - \Gamma^{1/2} C \Sigma_{X'Y'} \Gamma^{1/2} - \Gamma^{1/2} \Sigma_{Y'X'} C^T \Gamma^{1/2} + \Gamma^{1/2} C \Sigma_{X'X'} C^T \Gamma^{1/2}\} \\
&= tr\{\Sigma^*_{Y'Y'} - C^* \Sigma^*_{X'Y'} - \Sigma^*_{Y'X'} C^{*T} + C^* \Sigma^*_{X'X'} C^{*T}\} \\
&= tr\{(\Sigma^*_{Y'Y'} - \Sigma^*_{Y'X'} \Sigma^{*-1}_{X'X'} \Sigma^*_{X'Y'}) + (C^* \Sigma^{*1/2}_{X'X'} - \Sigma^*_{Y'X'} \Sigma^{*-1/2}_{X'X'})(C^* \Sigma^{*1/2}_{X'X'} - \Sigma^*_{Y'X'} \Sigma^{*-1/2}_{X'X'})^T\}
\end{aligned}$$

The new terms with asterisks are defined as follows.

$$\begin{aligned}
\Sigma^*_{X'X'} &= \Sigma_{X'X'} \\
\Sigma^*_{Y'Y'} &= \Gamma^{1/2} \Sigma_{Y'Y'} \Gamma^{1/2} \\
\Sigma^*_{X'Y'} &= \Sigma_{X'Y'} \Gamma^{1/2} \\
\Sigma^*_{Y'X'} &= \Gamma^{1/2} \Sigma_{Y'X'} \\
C^* &= \Gamma^{1/2} C
\end{aligned}$$

This criterion is minimized when $C^* \Sigma^{*1/2}_{X'X'} = \sum_j \lambda_j'^{-1/2} v'_j w_j'^T$, where v'_j are the eigenvector associated with the jth largest eigenvalue of the matrix $\Sigma^*_{Y'X'} \Sigma^{*-1}_{X'X'} \Sigma^*_{X'Y'} = \Gamma^{1/2} \Sigma_{Y'X'} \Sigma^{-1}_{X'X'} \Sigma_{X'Y'} \Gamma^{1/2}$
And $w'_j = \lambda_j'^{-1/2} \Sigma^{*-1/2}_{X'X'} \Sigma^*_{X'Y'} v'_j = \lambda_j'^{-1/2} \Sigma^{-1/2}_{X'X'} \Sigma_{X'Y'} \Gamma^{1/2} v'_j$
Therefore, the value of C that minimizes this criterion is:

$$C = \Gamma^{-1/2} (\sum_j v'_j v_j'^T) \Gamma^{1/2} \Sigma_{Y'X'} \Sigma^{-1}_{X'X'}$$

As we can see here, the solution to minimize the objective function $W(t)$ with X' and Y' is still in the same exact form with that for X and Y, but now just replacing the original data with new transformed data (equation (6.83), page 178 Izenman).

This is why we can say that the criterion is invariant under nonsingular transformations X' and Y'.

$\square$