Math 585 HW 6 Spring 2024, Due Friday, March 8, 2 pages Problems A)-E)
Quiz 6 on Wed. March 6, covers HW6 including discriminant analysis.

For the following R problems perform the perform the *source("J:/mpack.txt")* command as described in homework 3. Also copy and paste commands from (http://parker.ad.siu.edu/Olive/mrsashw.txt) for the relevant problem into R.

A), 6.10 The Venables and Ripley (2003) CPU data has variables syct = cycle time,
mmin = minimum main memory,
chmin = minimum number of channels,
chmax = maximum number of channels,
perf = published performance, and
estperf = estimated performance.

a) There are nonlinear relationships among the variables and 1 is added to each variable to make them positive. Read more about the data set and make a scatterplot matrix with the *R commands* for this part. You can make the help window small by clicking the box with the − in the upper right corner. Include the scatterplot matrix in *Word*.

b) The log rule suggests using the log transformation on all of the variables. Make the log transformations, scatterplot matrix and DD plot with the *R commands* for this part. Right click "Stop" to go from the DD plot to the R prompt. Wait until part d) until you put plots in *Word*.

c) You might be able to get a better scatterplot matrix and DD plot by doing alternative transformations on the last two variables. The commands for this part give the log transformation for the first 4 variables and possible transformations for the last variables. Clearly state which transformations you use for the 5th and 6th variable. For example if you decide logs are ok, write down the following transformations.

```
zz[,5] <- log(z[,5])
zz[,6] <- log(z[,6])
```

d) For your data set zz of transformed variables, make the scatterplot matrix and DD plot and put the two plots in *Word*.

e) Put the classical PCA output using the correlation matrix into *Word* with the command for this problem.

f) Put the robust PCA output using the correlation matrix into *Word* with the command for this problem.

g) Comment on the similarities or differences of the classical and robust PCA.

B), 7.6 Copy and paste the R commands for this problem into R. These commands make $\boldsymbol{w}$ 3×1 and $\boldsymbol{y}$ 5×1 where there are $n_1 = n_2 \equiv n_i$ cases for both $\boldsymbol{w}$ and $\boldsymbol{y}$. The eight variables $w_1, w_2, w_3, y_1, ..., y_5$ are iid $N(0, 1)$. Hence the population canonical correlations are 0. The output starts with $n_i = 500$ and gives the first sample correlation $\hat{\rho}_1$, then increases n_i by 500 and repeats. How large does n_i need to be before $\hat{\rho}_1 < 0.05$?

```
Logistic Regression Output,
Response = nodal involvement, Terms = (acid size xray)
Label      Estimate        Std. Error     Est/SE    p-value
Constant  -3.57564         1.18002        -3.030     0.0024
acid       2.06294         1.26441         1.632     0.1028
size       1.75556         0.738348        2.378     0.0174
xray       2.06178         0.777103        2.653     0.0080

Number of cases: 53, Degrees of freedom: 49, Deviance: 50.660
```

C), 8.2 Treatment for prostate cancer depends on whether the cancer has spread to the surrounding lymph nodes. Let the response variable = group y = *nodal involvement* (0 for absence, 1 for presence). Let $x_1 = acid$ (serum acid phosphatase level), $x_2 = size$ (= tumor size: 0 for small, 1 for large) and $x_3 = xray$ (xray result: 0 for negative, 1 for positive). Assume the case to be classified has $\boldsymbol{x}$ with $x_1 = acid = 0.65$, $x_2 = 0$ and $x3 = 0$.

a) Find ESP for $\boldsymbol{x}$.
b) Is $\boldsymbol{x}$ classified in group 0 or group 1?
c) Find $\hat{\rho}(\boldsymbol{x})$.

D), 8.3 Recall that X comes from a uniform(a,b) distribution, written $x \sim U(a, b)$, if the pdf of x is $f(x) = \dfrac{1}{b-a}$ for $a < x < b$ and $f(x) = 0$, otherwise. Suppose group 1 has $X \sim U(-3, 3)$, group 2 has $X \sim U(-5, 5)$, and group 3 has $X \sim U(-1, 1)$. Find the maximum likelihood discriminant rule for classifying a new observation x.

E) To get on *ARC*, send your cursor to the math progs folder, click right mouse button twice, move the cursor to `ARC` and double click. A window will appear. Move the cursor to the *ARC* icon and double click. These menu commands will be written "math progs > ARC > ARC." To quit *ARC*, move cursor to the **x** in the northeast corner and click.

Activate the banknote.lsp dataset with the menu commands
"File > Load > Data > Arcg > banknote.lsp." Scroll up the screen to read the data description. The group (status) is 0 if the banknote is genuine and 1 if the banknote is counterfeit.

a) From *Graph&Fit* select *Fit binomial response.* Select *Top* and *Diagonal* as predictors, *Status* as the response and *ones* as the number of trials. Include the output in *Word*.

b) From *Graph&Fit* select *Plot of.* Place *Status* on *V* and *B1:Eta'U* on *H.* From the *OLS* popup menu, select *Logistic* and move the slider bar to 1. From the *lowess* popup menu select *SliceSmooth* and move the slider bar until the fit is good. Include your response plot in *Word.* Are the slice means (observed proportions) tracking the logistic curve (fitted proportions) fairly well? Use *lowess* if *SliceSmooth* does not work.
c) Predict $\hat{\rho}(\boldsymbol{x})$ if $x_1 = \text{Top} = 10.7$ and $x_2 = \text{Diagonal} = 140.5$.