Statistics 5401
19. Biplots

Gary W. Oehlert
School of Statistics
313B Ford Hall
612-625-1557
gary@stat.umn.edu

We have seen that principal components are a compact way of "approximating" data.
We have seen that plotting principal components can be a useful way of visualizing data.
But we have also seen that the two-dimensional plot of principal components can be a bit confusing to interpret, because it is difficult to relate back to the original variables.
That is where the biplot comes in.
The biplot is an enhanced plot of principal components that helps us to see connections between components and the original variables.
Let $\mathbf{X}$ be the data matrix with variable means subtracted out. We may, or may not, wish to rescale to make all the columns have unit variance; that depends on the problem.
Make the svd

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}'$$

Begin with the first two columns of $\mathbf{U}$. Let

$$\ell_i = \begin{bmatrix} d_1\mathbf{U}_{i1} \\ d_2\mathbf{U}_{i2} \end{bmatrix}$$

The bivariate points ℓ_i are the points we plot in the usual principal components plot.
Now the first two columns of $\mathbf{V}$. Let

$$r_j = \begin{bmatrix} \mathbf{V}_{j1} \\ \mathbf{V}_{j2} \end{bmatrix}$$

The bivariate point r_j tell us how the jth variable enters into the first two principal components.
The first two columns of $\mathbf{U}$ and $\mathbf{V}$ along with d_1 and d_2 form a rank two approximation to $\mathbf{X}$

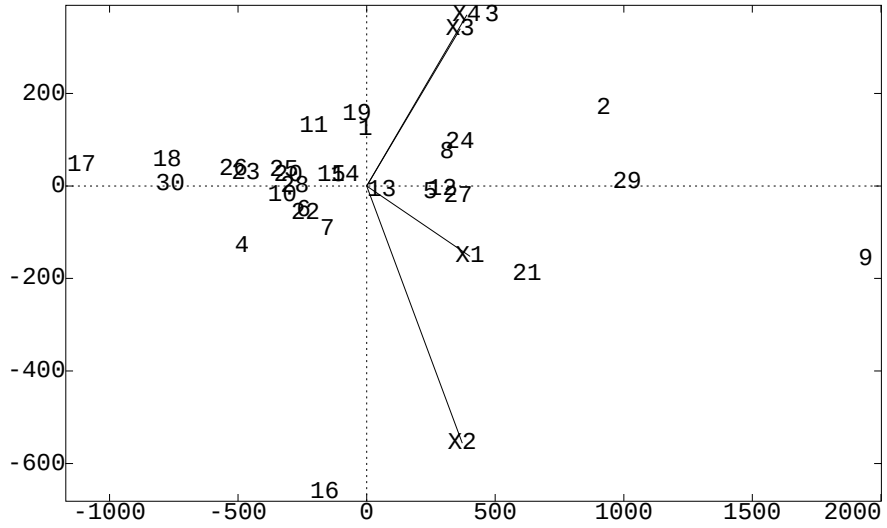
$$\begin{aligned} \mathbf{X}_{ij} &\approx \mathbf{U}_{i1}d_1\mathbf{V}_{j1} + \mathbf{U}_{i2}d_2\mathbf{V}_{j2} \\ &= < \ell_i, r_j > = \cos(\theta_{ij})||\ell_i||\,||r_j|| \end{aligned}$$

where θ_{ij} is the angle between the two-vectors ℓ_i and r_j.
Thus large values of $\mathbf{X}_{ij}$ will occur when there is a small angle between ℓ_i and r_j, and small values of $\mathbf{X}_{ij}$ will occur when there is a large angle between ℓ_i and r_j. Or put another way, positively correlated ℓ_i and r_j mean than $\mathbf{X}_{ij}$ is positive, and negatively correlated ℓ_i and r_j mean than $\mathbf{X}_{ij}$ is negative.
OK, so here is the biplot.
Plot the points ℓ_i with labels to indicate which case is which.
Plot the points cr_j with labels to indicate which variable is which, and with a line from the origin to cr_j. Choose the constant c so that the rescaled r_js are about the same size as the ℓ_is.

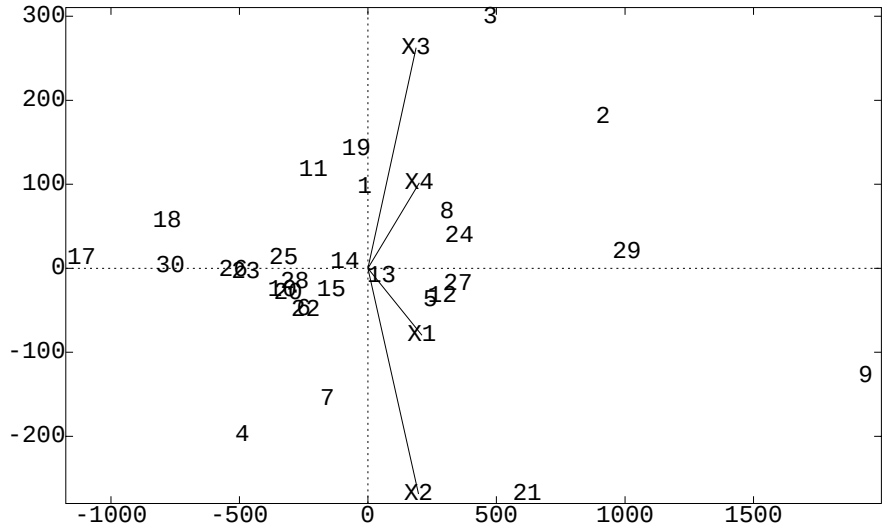
```
Cmd> X <- readdata("",x1,x2,x3,x4,d)
Read from file "~/5401/JW5data/T4-3.DAT"
Column 1 saved as REAL vector x1
```

```
Column 2 saved as REAL vector x2
Column 3 saved as REAL vector x3
Column 4 saved as REAL vector x4
Column 5 saved as REAL vector d

Cmd> X <- hconcat(x1,x2,x3,x4)

Cmd> rlab <- paste(run(30),multiline:T)

Cmd> rlab
 (1) "1"
 (2) "2"
 (3) "3"
...
(29) "29"
(30) "30"

Cmd> clab <- vector("X1","X2","X3","X4")

Cmd> biplot <- macro("
@X <- $1
@rlabs <- $2
@clabs <- $3
@X <- @X - tabs(@X,mean:T)'
@p <- ncols(@X)
@s <- svd(@X,all:T)
@p1 <- @s$leftvectors[,1]*@s$values[1]
@p2 <- @s$leftvectors[,2]*@s$values[2]
@v1 <- @s$rightvectors[,1]
@v2 <- @s$rightvectors[,2]
@f <- vector(max(@p1)/max(@v1),\
max(@p2)/max(@v2),min(@p1)/min(@v1),\
min(@p2)/min(@v2))
@f <- @f[@f > 0]
@f <- min(@f)
@v1 <- @v1*@f
@v2 <- @v2*@f
@blank <- \" \"
chplot(@p1,@p2,@rlabs,xlab:@blank,\
ylab:@blank,show:F)
for(@i,run(@p))
addpoints(@v1[@i],@v2[@i],\
          symbol:@clabs[@i],show:F)
addlines(vector(0,@v1[@i]),\
          vector(0,@v2[@i]),show:F)
```

```
;;

showplot(xmin:?,ymin:?,xmax:?,\
  ymax:?,$K)",dollars:T)

Cmd> biplot(X,rlab,clab)
```




```
Cmd> biplot(X[-16,],rlab[-16],clab)
```




```
Cmd> readdata("",name,x1,x2,x3,x4,x5,x6)
Read from file "~/5401/JW5data/T12-9.DAT"
Column 1 saved as factor name
```

```
Column 2 saved as REAL vector x1
Column 3 saved as REAL vector x2
Column 4 saved as REAL vector x3
Column 5 saved as REAL vector x4
Column 6 saved as REAL vector x5
Column 7 saved as REAL vector x6

Cmd> X <- hconcat(x1,x2,x3,x4,x5,x6)

Cmd> X <- X/tabs(X,stddev:T)’

Cmd> rlabs <- getlabels(name)

Cmd> clabs <- vector("sat","top","acc",\
"sfr","exp","grd")

Cmd> biplot(X,rlabs,clabs)
```