MNIST and Lasso

In this problem, we'll revisit the MNIST image data where we have been trying to classify images of handwritten digits.

```
library("glmnet")
```

```
## Loading required package: Matrix
```

```
## Loading required package: foreach
```

```
## Loaded glmnet 2.0-5
```

```
load("../Data/mnist68.RData")
images_df <- mnist68
```

As you might recall from working previously with this data, many of the features (pixel) have zero of nearly zero variance (ie. always white). Use the following code to select those features whose variance is larger than the 75-th percentile of all feature variances. This will keep the computation time reasonable.

```
myv = rep(NA,784)
for (j in 1:784){myv[j] <- var(images_df[,j])}
myfeatures = (1:784)[myv > quantile(myv, .9)]
mydf = images_df[,c(myfeatures,785)]
mydf$labels = as.numeric(mydf$label==8)
```

(a)

Set up the proper model matrix from this data frame. Also, create a vector Y for the labels, so that **glmnet()** can be used. *We will be using a linear model to predict the labels, although this is really a classification problem.*

(b)

Fit a Lasso model to this model matrix using glmnet(). Produce output plot of the Lasso coefficient trajectories against λ and the L^1 norm of the coefficients. The trajectories have some unusual features. Comment on these.

(c)

Once λ is sufficiently large, there are only a few predictors (features) in the model. What are the last five features that remain in the model for sufficiently large λ? Do you think these features are independent? Explain your answer. *Use the help pages for glmnet().*