Fashion MNIST

Group Member 1 Name: _____ **Group Member 1 SID:** _____

Group Member 2 Name: _____ **Group Member 2 SID:** _____

The dataset contains $n = 18,000$ different 28×28 grayscale images of clothing, each with a label of either *shoes*, *shirt*, or *pants* (6000 of each). If we stack the features into a single vector, we can transform each of these observations into a single $28 * 28 = 784$ dimensional vector. The data can thus be stored in a $n \times p = 18000 \times 784$ data matrix $\mathbf{X}$, and the labels stored in a $n \times 1$ vector $\mathbf{y}$.

Once downloaded, the data can be read as follows.

```
library(readr)
FMNIST <- read_csv("FashionMNIST.csv")
```

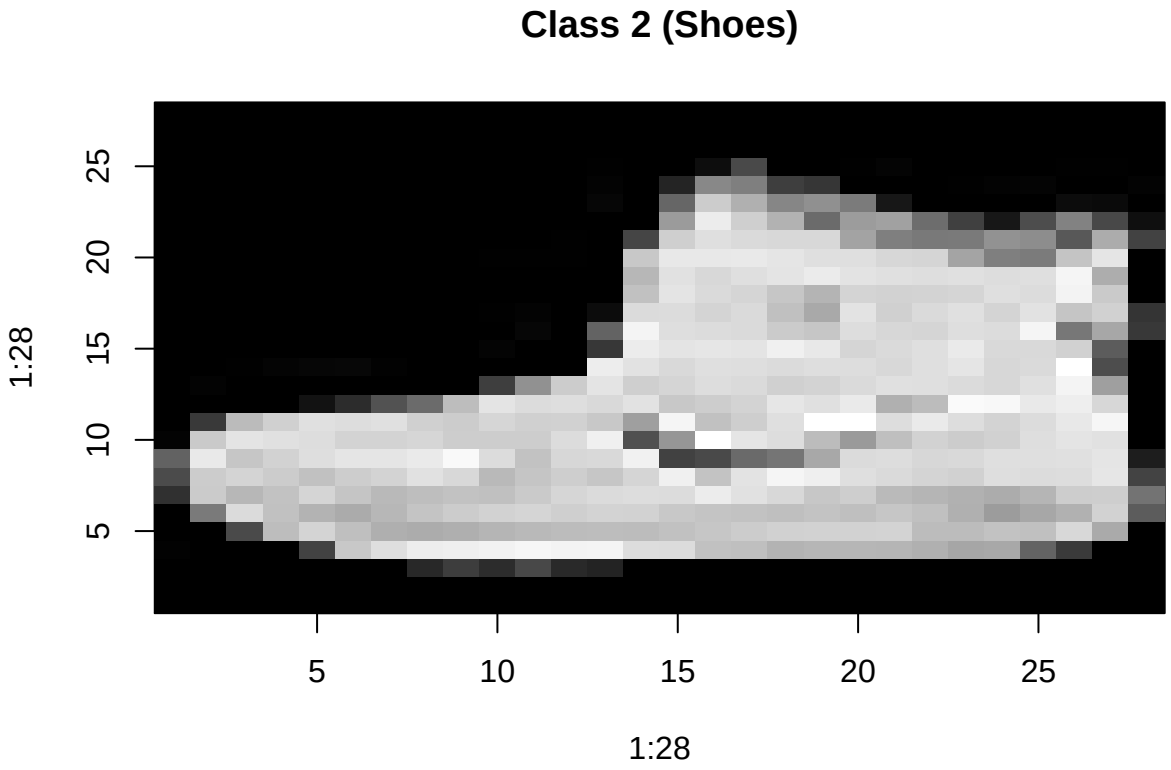
```
## Parsed with column specification:
## cols(
##   .default = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

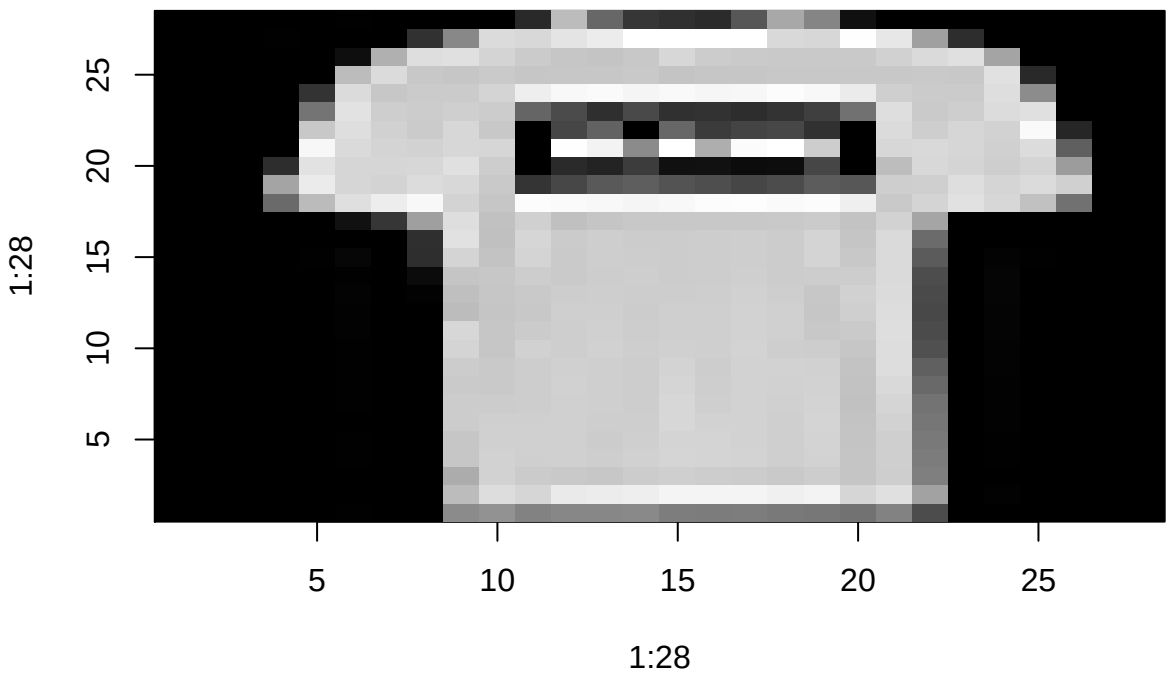
```
y <- FMNIST$label
X <- subset(FMNIST, select = -c(label))
rm('FMNIST') #remove from memory -- it's a relatively large file
#print(dim(X))
```

We can look at a few of the images:

```
X2 <- matrix(as.numeric(X[1,]), ncol=28, nrow=28, byrow = TRUE)
X2 <- apply(X2, 2, rev)
image(1:28, 1:28, t(X2), col=gray((0:255)/255), main='Class 2 (Shoes)')
```

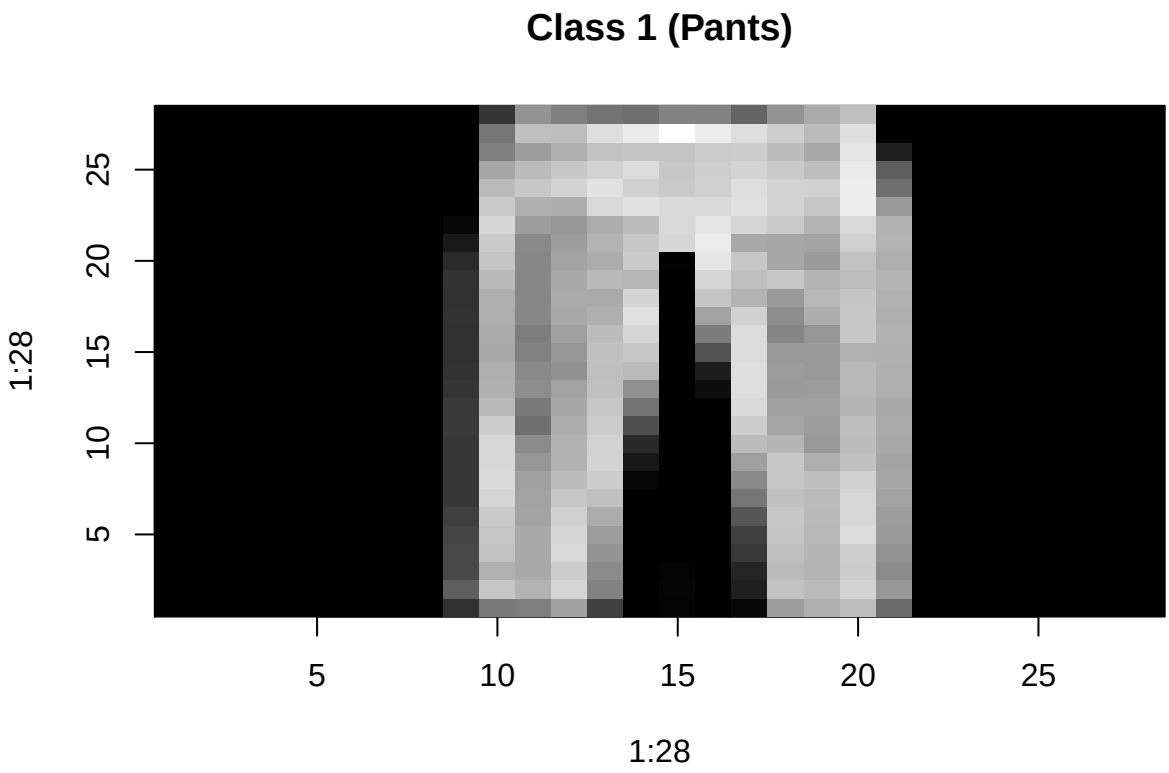



```
X0 <- matrix(as.numeric(X[2,]), ncol=28, nrow=28, byrow = TRUE)
X0 <- apply(X0, 2, rev)
image(1:28, 1:28, t(X0), col=gray((0:255)/255), main='Class 0 (Shirt)')
```




```
X1 <- matrix(as.numeric(X[8,]), ncol=28, nrow=28, byrow = TRUE)
X1 <- apply(X1, 2, rev)
```

```r
image(1:28, 1:28, t(X1), col=gray((0:255)/255), main='Class 1 (Pants)')
```



Data exploration and dimension reduction

In this section, you will experiment with representing the images in fewer dimensions than $28 * 28 = 784$. You can use any of the various dimension reduction techniques introduced in class. How can you visualize these lower dimensional representations as images? How small of dimensionality can you use and still visually distinguish images from different classes?

YOUR CODE GOES HERE

Classification task

Binary classification

In this section, you should use the techniques learned in class to develop a model for binary classification of the images. More specifically, you should split up the data into different pairs of classes, and fit several binary classification models. For example, you should develop a model to predict shoes vs shirts, shoes vs pants, and pants vs shirts.

Remember that you should try several different methods, and use model selection methods to determine which model is best. You should also be sure to keep a held-out test set to evaluate the performance of your model.

YOUR CODE GOES HERE

Multiclass classification

In this section, you will develop a model to classify all three classes simultaneously. You should again try several different methods, and use model selection methods to determine which model is best. You should also be sure to keep a held-out test set to evaluate the performance of your model. (Side question: how could you use the binary models from the previous section to develop a multiclass classifier?)

YOUR CODE GOES HERE