MNIST feature reduction and KNN

```
library(readr)
library(class)
library(MASS)

#Reading training and test datasets
train <- read_csv("train.csv")
```

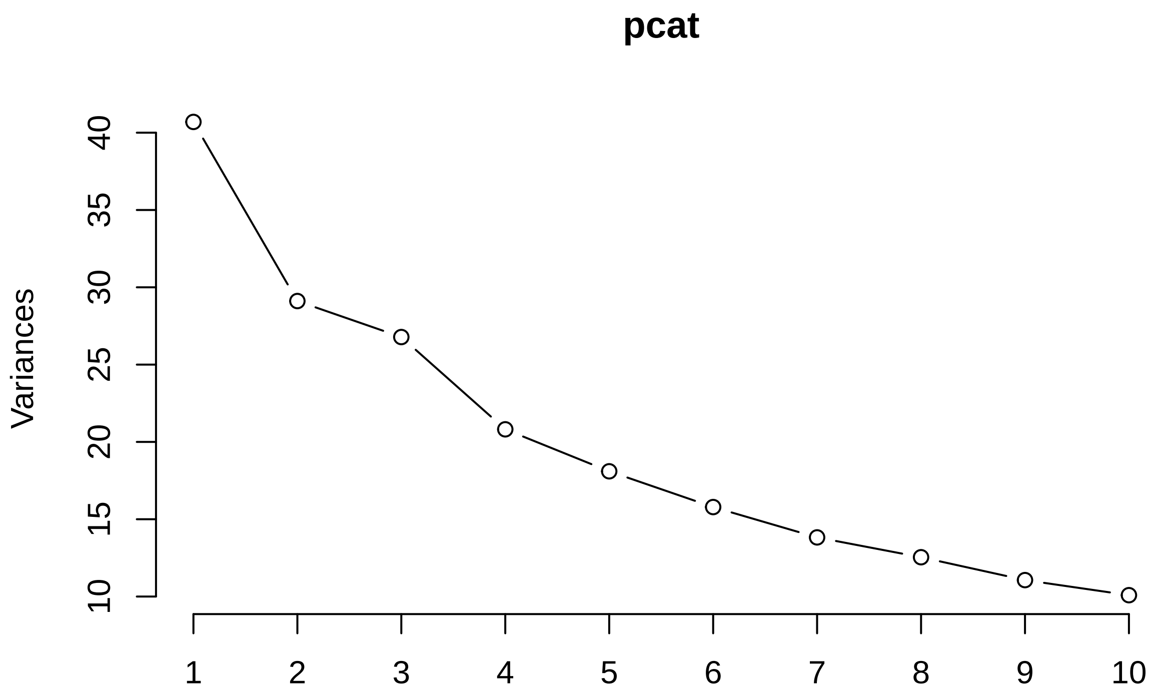
```
## Parsed with column specification:
## cols(
##   .default = col_integer()
## )
```

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

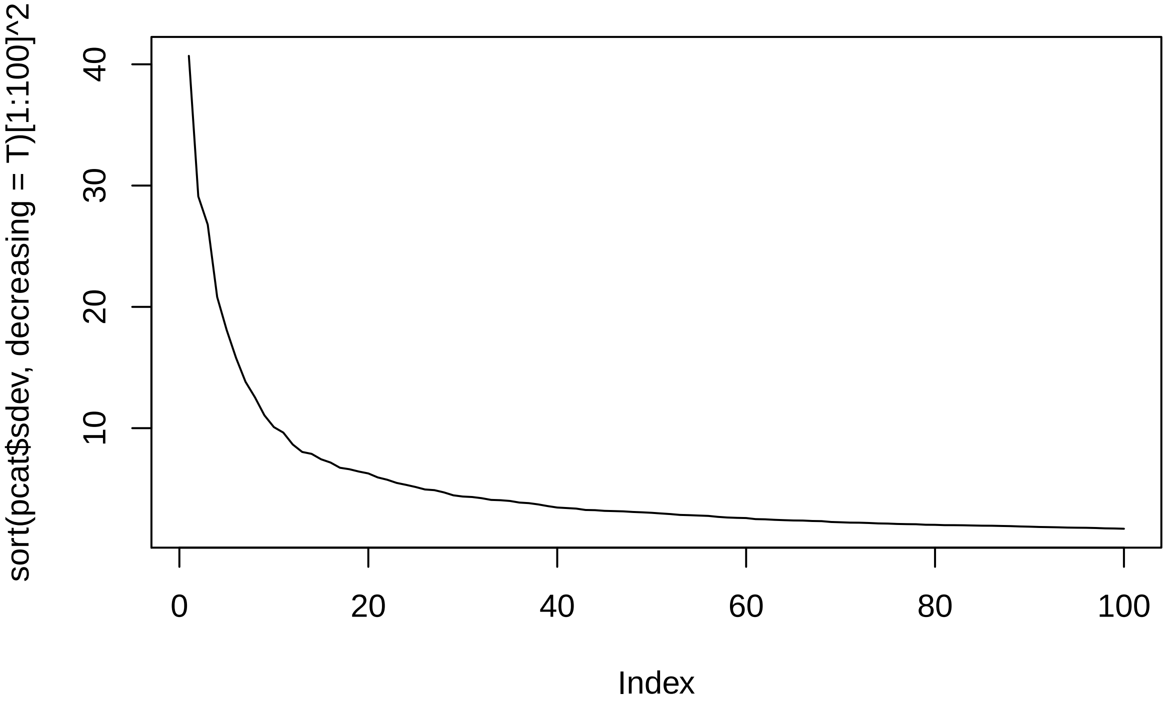
```
test <- read_csv("test.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_integer()
## )
## See spec(...) for full column specifications.
```

```
class <- as.factor(train[[1]]) #Reading training labels

cvars <- apply(train[,-1], 2, var) #Calculating variance of all training variables
train_z <- train[,c(F, cvars!=0)] #Subsetting training dataset, exlucding zero variance
pcat <- prcomp(x=train_z, center=T, scale=T) #Calculating PCA Components

#Plotting PCA components
plot(pcat, type='l')
```



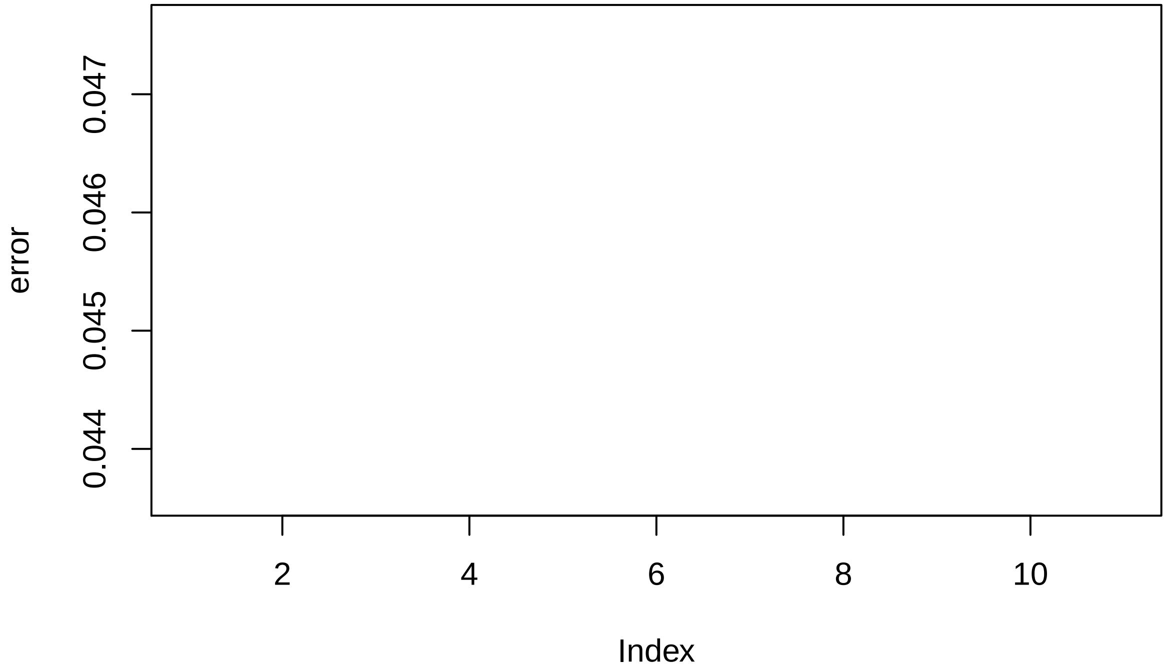
```r
plot(sort(pcat$sdev, decreasing = T)[1:100]^2, type='l')
```



```
pca.train <- pcat$x
pca.test <- predict(pcat,test)

#Subsetting first 60 components
cols <- 60

pca.train <- pca.train[,1:cols]
pca.test <- pca.test[,1:cols]

#Applying KNN

#Finding best K
error <- NULL

for(i in c(1,3,5,9,7,11))
{  k.fit <- knn.cv(pca.train, class, k=i, l=0, prob=F, use.all = T)
  error[i] <- mean(k.fit != class)
  print(i)
}
```

```
## [1] 1
## [1] 3
## [1] 5
## [1] 9
## [1] 7
## [1] 11
```

```r
plot(error, type='l')
```



```r
k.fit <- knn(pca.train, pca.test, class, k=5, l=0, prob=F, use.all = T)
submit <- data.frame(ImageId = 1:nrow(test), Label = k.fit)

write.csv(submit, "knn_submission.csv", row.names = F)
```