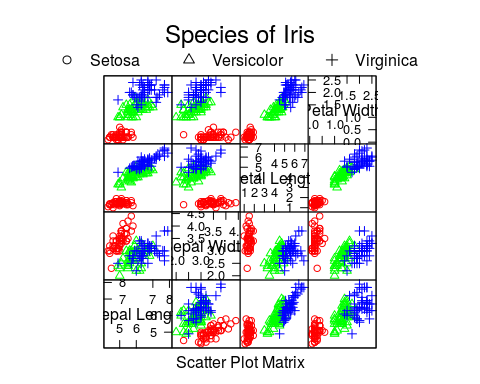
HW4.R

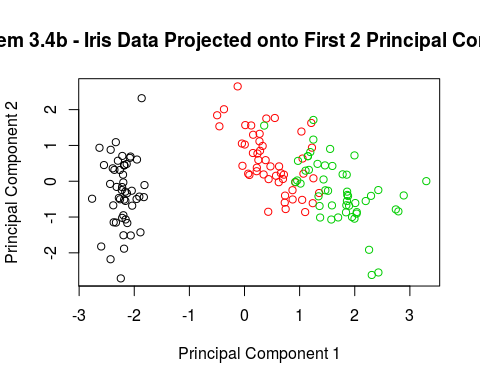
neeraj

Wed Mar 2 18:54:07 2016

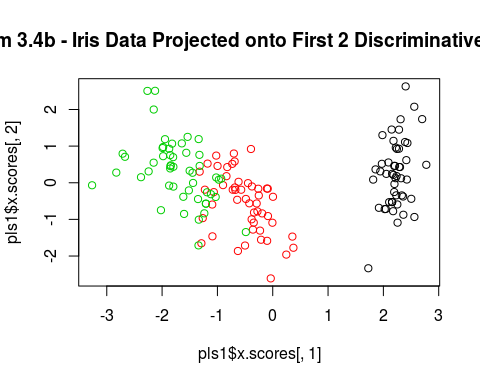
##Neeraj Asthana (nasthan2)  
##CS 498 HW4  
  
##Environment Setup  
library(lattice)  
library(plsdepot)  
library(ggplot2)  
library(scatterplot3d)  
setwd('/home/neeraj/Documents/UIUC/CS 498/CS498MachineLearning/HW4')  
  
##Problem 3.4  
#Read data  
iris\_data <- read.csv("iris.data", header=FALSE)  
names <- c("sep\_len", "sep\_width", "petal\_len", "petal\_width", "class")  
colnames(iris\_data) <- names  
  
##Problem 3.4a  
#Scatterplot matrix  
pchr <- c(1,2,3)  
colors <- c("red", "green", "blue")  
species\_names <- c("Setosa","Versicolor","Virginica")  
feature\_names <- c("Sepal Length", "Sepal Width", "Petal Length", "Petal Width")  
  
ss <- expand.grid(species=1:3)  
parset <- with( ss, simpleTheme(pch=pchr[species],col=colors[species]))  
  
splom(iris\_data[,c(1:4)], groups=iris\_data$class,  
 varnames = feature\_names,  
 par.settings=parset,  
 key=list(title="Species of Iris",  
 columns=3,  
 points=list(pch=pchr),  
 text=list(species\_names)))



##Problem 3.4b  
title34b <- "Problem 3.4b - Iris Data Projected onto First 2 Principal Components"  
iris\_features <- iris\_data[,c(1,2,3,4)]  
pca <- prcomp(iris\_features, center = TRUE, scale. = TRUE)  
plot(pca$x[,1], pca$x[,2], col=iris\_data[,5], xlab = "Principal Component 1", ylab = "Principal Component 2", main=title34b)



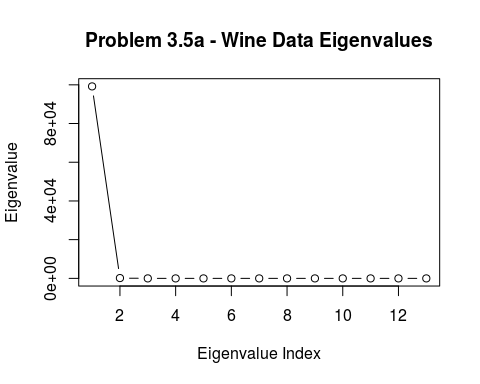
##Problem 3.4c  
title34c <- "Problem 3.4b - Iris Data Projected onto First 2 Discriminative Directions"  
hot\_vectors <- matrix(0, 150, 3)  
hot\_vectors[1:50,1] <- 1  
hot\_vectors[51:100,2] <- 1  
hot\_vectors[101:150,3] <- 1  
pls1 <- plsreg2(iris\_features, hot\_vectors)  
plot(pls1$x.scores[,1], pls1$x.scores[,2], col=iris\_data[,5], main = title34c)



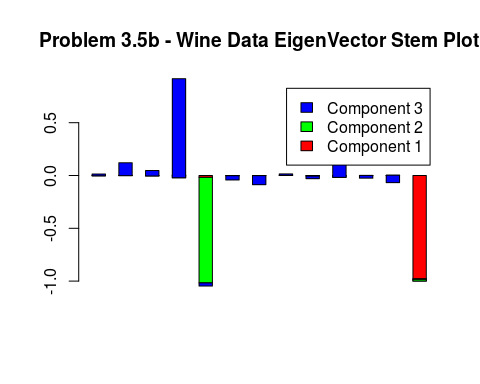
##Problem 3.5  
wine\_data <- read.csv("wine.data", header=FALSE)  
wine\_features <- wine\_data[,-1]  
wine\_labels <- wine\_data[,1]  
  
##Problem 3.5a  
title35a <- "Problem 3.5a - Wine Data Eigenvalues"  
covmat <- cov(wine\_features)  
eig <- eigen(covmat, symmetric=FALSE)  
print(eig$values)

## [1] 9.920179e+04 1.725353e+02 9.438114e+00 4.991179e+00 1.228845e+00  
## [6] 8.410639e-01 2.789735e-01 1.513813e-01 1.120968e-01 7.170260e-02  
## [11] 3.757598e-02 2.107237e-02 8.203703e-03

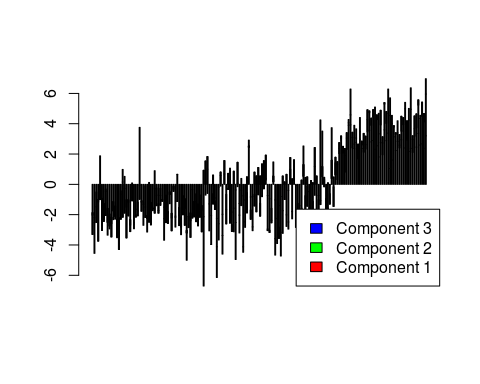
plot(eig$values, type="b", ylab="Eigenvalue", xlab = "Eigenvalue Index", main=title35a)



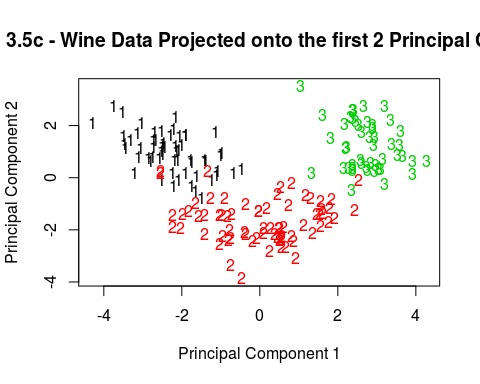
##Problem 3.5b  
title35b1 <- "Problem 3.5b - Wine Data EigenVector Stem Plot"  
counts <- t(eig$vectors[,1:3])  
barplot(counts, width=.1, space=1, col=colors, legend = c("Component 1", "Component 2", "Component 3"), main=title35b1)



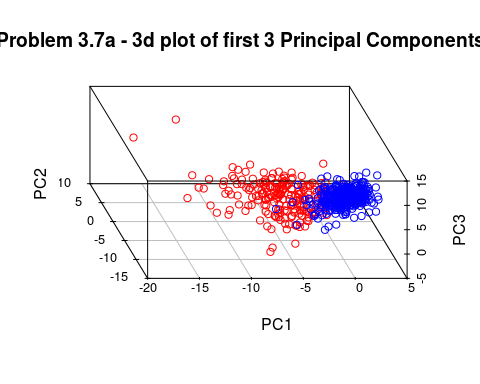
title35b2 <- "Problem 3.5b - Wine Data Stem Plot"  
princip2 <- prcomp(wine\_features, center = TRUE, scale. = TRUE, xlab="Index", main=title35b2)  
barplot(t(princip2$x[,1:3]), width=.1, space=1, col=colors, legend = c("Component 1", "Component 2", "Component 3"), args.legend = list(x="bottomright") )



##Problem 3.5c  
title35c <- "Problem 3.5c - Wine Data Projected onto the first 2 Principal Components"  
plot(princip2$x[,1:2],col="white", pch=3, xlab = "Principal Component 1", ylab = "Principal Component 2", main=title35c)  
text(princip2$x[,1:2], col=wine\_labels, labels = wine\_labels)



##Problem 3.7  
cancer\_data <- read.csv("wdbc.data", header = FALSE)  
cancer\_id <- cancer\_data[,1]  
cancer\_labels <- cancer\_data[,2]  
cancer\_features <- cancer\_data[,-c(1,2)]  
  
##Problem 3.7a  
princip3 <- prcomp(cancer\_features, center=TRUE, scale. = TRUE)  
title37a <- "Problem 3.7a - 3d plot of first 3 Principal Components"  
plot3d <- scatterplot3d(princip3$x[cancer\_labels == "M",1:3], color = "red", angle=110, main = title37a)  
plot3d$points3d(princip3$x[cancer\_labels != "M",1:3], col = "blue")



##Problem 3.7b  
cancer\_hot\_vectors <- matrix(0, 569, 2)  
cancer\_hot\_vectors[cancer\_labels == "B",1] <- 1  
cancer\_hot\_vectors[cancer\_labels == "M",2] <- 1  
cancer\_pls1 <- plsreg2(cancer\_features, cancer\_hot\_vectors, comps=3)  
title37b <- "Problem 3.7a - 3d plot of breast cancer data projected onto the first 3 Discriminative Directions"  
plot3d2 <- scatterplot3d(cancer\_pls1$x.scores[cancer\_labels == "M",1:3], color = "red", angle=40, main = title37b)  
plot3d$points3d(cancer\_pls1$x.scores[cancer\_labels != "M",1:3], col = "blue")

