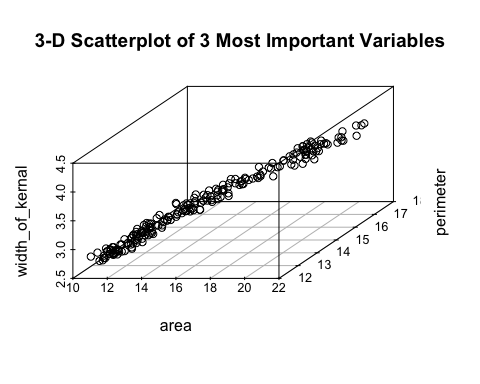
3dscatterplot\_hw5.R

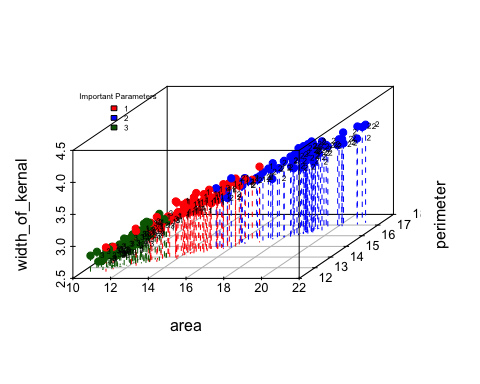
jasminedumas

Tue Mar 8 22:08:58 2016

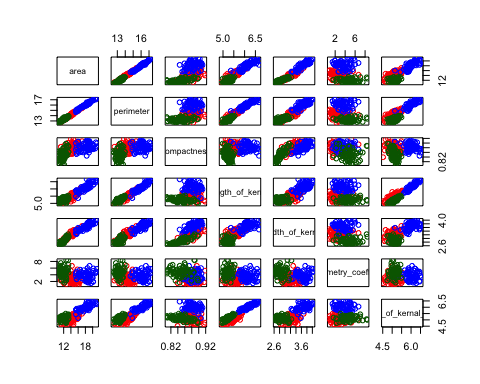
library(scatterplot3d)  
  
seeds <- read.table("/Users/jasminedumas/desktop/depaul/IS467/seeds.txt", header = T)  
  
with(seeds, {  
 scatterplot3d(area, # x axis  
 perimeter, # y axis  
 width\_of\_kernal, # z axis  
 main="3-D Scatterplot of 3 Most Important Variables")  
})



# create column indicating point color  
seeds$pcolor[seeds$class==1] <- "red"  
seeds$pcolor[seeds$class==2] <- "blue"  
seeds$pcolor[seeds$class==3] <- "darkgreen"  
  
with(seeds, {  
 s3d <- scatterplot3d(area, perimeter, width\_of\_kernal, # x y and z axis  
 color=pcolor, pch=19, # circle color indicates no. of cylinders  
 type="h", lty.hplot=2, # lines to the horizontal plane  
 scale.y=.75) # lines to the horizontal plane  
   
 s3d.coords <- s3d$xyz.convert(area, perimeter, width\_of\_kernal) # convert 3D coords to 2D projection  
 text(s3d.coords$x, s3d.coords$y, # x and y coordinates  
 labels=seeds$class, # text to plot  
 cex=.5, pos=4)   
 # add the legend  
 legend("topleft", inset=.05, # location and inset  
 bty="n", cex=.5, # suppress legend box, shrink text 50%  
 title="Important Parameters",  
 c("1", "2", "3"), fill=c("red", "blue", "darkgreen"))  
   
})



pairs(seeds[, 1:7], col=seeds$pcolor)



# linear model  
fit <- lm(class ~ area + perimeter + compactness +   
 length\_of\_kernal + width\_of\_kernal +   
 asymmetry\_coefficient + length\_of\_kernal\_grove, data = seeds)   
summary(fit)

##   
## Call:  
## lm(formula = class ~ area + perimeter + compactness + length\_of\_kernal +   
## width\_of\_kernal + asymmetry\_coefficient + length\_of\_kernal\_grove,   
## data = seeds)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.30568 -0.24785 -0.01632 0.24198 1.22362   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 53.44356 7.44511 7.178 1.32e-11 \*\*\*  
## area 1.48907 0.26133 5.698 4.25e-08 \*\*\*  
## perimeter -3.22038 0.53815 -5.984 9.77e-09 \*\*\*  
## compactness -30.67744 5.24108 -5.853 1.92e-08 \*\*\*  
## length\_of\_kernal -2.31510 0.45444 -5.094 8.01e-07 \*\*\*  
## width\_of\_kernal 0.24598 0.78571 0.313 0.755   
## asymmetry\_coefficient 0.11489 0.02257 5.089 8.19e-07 \*\*\*  
## length\_of\_kernal\_grove 2.19260 0.20358 10.770 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.415 on 202 degrees of freedom  
## Multiple R-squared: 0.7515, Adjusted R-squared: 0.7428   
## F-statistic: 87.25 on 7 and 202 DF, p-value: < 2.2e-16

# forward selection  
min.model = lm(class ~ 1, data=seeds) # aka intercept only model  
biggest = formula(lm(class ~ area + perimeter + compactness +   
 length\_of\_kernal + width\_of\_kernal +   
 asymmetry\_coefficient + length\_of\_kernal\_grove, data = seeds) ) # the full model  
model = step(min.model, direction='forward', scope=biggest)

## Start: AIC=-83.15  
## class ~ 1  
##   
## Df Sum of Sq RSS AIC  
## + asymmetry\_coefficient 1 46.654 93.346 -166.267  
## + compactness 1 39.476 100.524 -150.708  
## + width\_of\_kernal 1 25.105 114.895 -122.649  
## + area 1 16.766 123.234 -107.934  
## + perimeter 1 15.053 124.947 -105.035  
## + length\_of\_kernal 1 9.266 130.734 -95.528  
## <none> 140.000 -83.148  
## + length\_of\_kernal\_grove 1 0.083 139.917 -81.272  
##   
## Step: AIC=-166.27  
## class ~ asymmetry\_coefficient  
##   
## Df Sum of Sq RSS AIC  
## + compactness 1 18.1451 75.201 -209.66  
## + width\_of\_kernal 1 11.3020 82.044 -191.37  
## + area 1 6.7386 86.607 -180.00  
## + perimeter 1 6.0216 87.324 -178.27  
## + length\_of\_kernal 1 3.6115 89.734 -172.55  
## <none> 93.346 -166.27  
## + length\_of\_kernal\_grove 1 0.1319 93.214 -164.56  
##   
## Step: AIC=-209.66  
## class ~ asymmetry\_coefficient + compactness  
##   
## Df Sum of Sq RSS AIC  
## + length\_of\_kernal\_grove 1 1.99051 73.210 -213.29  
## <none> 75.201 -209.66  
## + length\_of\_kernal 1 0.25395 74.947 -208.37  
## + perimeter 1 0.15260 75.048 -208.09  
## + width\_of\_kernal 1 0.09095 75.110 -207.91  
## + area 1 0.02433 75.176 -207.73  
##   
## Step: AIC=-213.29  
## class ~ asymmetry\_coefficient + compactness + length\_of\_kernal\_grove  
##   
## Df Sum of Sq RSS AIC  
## + length\_of\_kernal 1 31.935 41.275 -331.64  
## + perimeter 1 25.289 47.921 -300.29  
## + area 1 20.475 52.735 -280.18  
## + width\_of\_kernal 1 15.820 57.390 -262.42  
## <none> 73.210 -213.29  
##   
## Step: AIC=-331.64  
## class ~ asymmetry\_coefficient + compactness + length\_of\_kernal\_grove +   
## length\_of\_kernal  
##   
## Df Sum of Sq RSS AIC  
## <none> 41.275 -331.64  
## + area 1 0.296664 40.978 -331.15  
## + width\_of\_kernal 1 0.146829 41.128 -330.39  
## + perimeter 1 0.040096 41.235 -329.84

summary(model)

##   
## Call:  
## lm(formula = class ~ asymmetry\_coefficient + compactness + length\_of\_kernal\_grove +   
## length\_of\_kernal, data = seeds)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.11443 -0.25620 0.00504 0.29770 1.22321   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.53844 1.25242 8.414 6.73e-15 \*\*\*  
## asymmetry\_coefficient 0.13656 0.02344 5.827 2.17e-08 \*\*\*  
## compactness -7.54868 1.52941 -4.936 1.65e-06 \*\*\*  
## length\_of\_kernal\_grove 2.58439 0.19984 12.932 < 2e-16 \*\*\*  
## length\_of\_kernal -2.92179 0.23199 -12.594 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4487 on 205 degrees of freedom  
## Multiple R-squared: 0.7052, Adjusted R-squared: 0.6994   
## F-statistic: 122.6 on 4 and 205 DF, p-value: < 2.2e-16