```
# Rudy Meza
# Machine Learning in R
# Michael Chang
# Spring
# '***Lab and Quiz***'
# '********************Machine Learning**********************'


library(ISLR)
attach(Wage)

# 7.8.1 Polynomial Regression and Step Functions

fit=lm(wage~poly(age ,4) ,data=Wage)
coef(summary (fit))

fit2=lm(wage~poly(age ,4, raw =T),data=Wage)
coef(summary (fit2))

fit2a=lm(wage~age+I(age ^2)+I(age ^3)+I(age ^4) ,data=Wage)
coef(fit2a)

agelims =range(age)
age.grid=seq (from=agelims [1], to=agelims [2])
preds=predict(fit ,newdata =list(age=age.grid),se=TRUE)
se.bands=cbind(preds$fit +2* preds$se.fit ,preds$fit -2* preds$se.fit)

par(mfrow =c(1,2) ,mar=c(4.5 ,4.5 ,1 ,1) ,oma=c(0,0,4,0))
plot(age ,wage ,xlim=agelims ,cex =.5, col =" darkgrey ")
title (" Degree -4 Polynomial ",outer =T)
lines(age.grid ,preds$fit ,lwd =2, col =" blue")
matlines (age.grid ,se.bands ,lwd =1, col =" blue",lty =3)

preds2 =predict(fit2 ,newdata =list(age=age.grid),se=TRUE)
max(abs(preds$fit - preds2$fit ))

fit.1= lm(wage~age ,data=Wage)
fit.2= lm(wage~poly(age ,2) ,data=Wage)
fit.3= lm(wage~poly(age ,3) ,data=Wage)
fit.4= lm(wage~poly(age ,4) ,data=Wage)
fit.5= lm(wage~poly(age ,5) ,data=Wage)
anova(fit.1, fit.2, fit.3, fit.4, fit.5)

coef(summary (fit.5))
( -11.983) ^2

fit.1= lm(wage~education +age ,data=Wage)
fit.2= lm(wage~education +poly(age ,2) ,data=Wage)
```

```
fit.3= lm(wage~education +poly(age ,3) ,data=Wage)
anova(fit.1, fit.2, fit.3)

fit=glm(I(wage >250)~poly(age ,4) ,data=Wage ,family =binomial )

preds=predict (fit ,newdata =list(age=age.grid),se=T)

pfit=exp(preds$fit )/(1+ exp( preds$fit ))
se.bands.logit = cbind(preds$fit +2* preds$se.fit , preds$fit -2*
                  preds$se.fit)
se.bands = exp(se.bands.logit)/(1+ exp(se.bands.logit))

preds=predict(fit ,newdata =list(age=age.grid),type="response",
           se=T)

plot(age ,I(wage >250) ,xlim=agelims ,type ="n",ylim=c(0 ,.2) )
points(jitter (age), I((wage >250) /5) ,cex =.5, pch ="|",
     col ="darkgrey")
lines(age.grid ,pfit ,lwd =2, col =" blue")
matlines(age.grid ,se.bands ,lwd =1, col =" blue",lty =3)

table(cut (age ,4))

fit=lm(wage~cut (age ,4) ,data=Wage)
coef(summary (fit))

# 7.8.2 Splines

library(splines )
fit=lm(wage~bs(age ,knots =c(25 ,40 ,60) ),data=Wage)
pred=predict (fit ,newdata =list(age =age.grid),se=T)
plot(age ,wage ,col = "gray ")
lines(age.grid ,pred$fit ,lwd =2)
lines(age.grid ,pred$fit+2*pred$se ,lty = "dashed")
lines(age.grid ,pred$fit-2*pred$se ,lty = "dashed")

dim(bs(age ,knots=c(25 ,40 ,60) ))

dim(bs(age ,df=6))

attr(bs(age ,df=6) ,"knots")

fit2=lm(wage~ns(age ,df =4) ,data=Wage)
pred2=predict(fit2 ,newdata =list(age=age.grid),se=T)
lines(age.grid , pred2$fit ,col ="red",lwd =2)

plot(age ,wage ,xlim=agelims ,cex =.5, col ="darkgrey")
title("Smoothing Spline")
```

```
fit=smooth.spline(age ,wage ,df =16)
fit2=smooth.spline(age ,wage ,cv=TRUE)
fit2$df

lines(fit ,col ="red",lwd =2)

lines(fit2 ,col =" blue",lwd =2)
legend("topright",legend =c("16 DF " ,"6.8 DF"),
      col=c("red "," blue"),lty =1, lwd =2, cex =.8)

plot(age ,wage ,xlim=agelims ,cex =.5, col =" darkgrey ")
title (" Local Regression ")
fit=loess(wage~age ,span =.2, data=Wage)
fit2=loess(wage~age ,span =.5, data=Wage)
lines(age.grid ,predict(fit ,data.frame(age=age.grid)),
     col ="red ",lwd =2)
lines(age.grid ,predict(fit2 ,data.frame(age=age.grid)),
     col =" blue",lwd =2)
legend("topright",legend =c("Span =0.2" , "Span =0.5") ,
      col=c("red" , " blue "),lty =1, lwd =2, cex =.8)

# 7.8.3 GAMs

gam1=lm(wage~ns(year ,4)+ns(age ,5) +education ,data=Wage)

library(gam)
gam.m3=gam(wage~s(year ,4)+s(age ,5)+education ,data=Wage)

par(mfrow =c(1,3))
plot(gam.m3, se=TRUE ,col ="blue ")

plot.Gam(gam1 , se=TRUE , col ="red ")

gam.m1=gam(wage~s(age ,5) +education ,data=Wage)
gam.m2=gam(wage~year+s(age ,5)+education ,data=Wage)
anova(gam.m1 ,gam.m2 ,gam.m3,test="F")

summary(gam.m3)

preds=predict(gam.m2,newdata =Wage)

gam.lo=gam(wage~s(year ,df=4)+lo(age ,span =0.7)+education ,
        data=Wage)
plot.Gam(gam.lo , se=TRUE , col ="green ")

gam.lo.i=gam (wage~lo(year ,age ,span =0.5) +education ,
           data=Wage)
```

```
library(akima)
plot(gam.lo.i)

gam.lr=gam(I(wage >250)~year+s(age ,df =5)+education ,
        family =binomial ,data=Wage)
par(mfrow =c(1,3))
plot(gam.lr,se=T,col =" green ")

table(education ,I(wage >250) )

gam.lr.s=gam (I(wage >250)~year+s(age ,df=5)+education ,family =
             binomial ,data=Wage ,subset =( education !="1. < HS Grad"))
plot(gam.lr.s,se=T,col = "green ")


# ''''''''''''''''''''Week 5 Quiz''''''''''''''''''''''''

# Question 1
fit1.1= lm(wage~age+education ,data=Wage)
fit2.2= lm(wage~poly(age ,2)+education ,data=Wage)
fit3.3= lm(wage~poly(age ,3)+education ,data=Wage)
anova(fit1.1, fit2.2, fit3.3)

# Model 2 AND 3
# Question 2

coef(summary(fit3.3))
# 2.12

# Question 3
table(cut(age, breaks=c(0,25,35,45,55,80)))
cut1<-cut(age, breaks=c(0,25,35,45,55,80))
fitc=lm(wage~cut1 ,data=Wage)
prefitc<-predict(fitc, newdata =list(age=age.grid),se=T)
prefitc
coef(summary(fitc))

64.493063+52.735717

117.23
# Question 4
library(splines )
fit=lm(wage~bs(age ,knots =c(25 ,40 ,60) ),data=Wage)
pred=predict (fit ,newdata =list(age =age.grid),se=T)
plot(age ,wage ,col = "gray ")
lines(age.grid ,pred$fit ,lwd =2)
lines(age.grid ,pred$fit+2*pred$se ,lty = "dashed")
lines(age.grid ,pred$fit-2*pred$se ,lty = "dashed")
```

```
# AGE 55, 104.91792
fit2=lm(wage~ns(age ,df =4) ,data=Wage)
pred2=predict(fit2 ,newdata =list(age=age.grid),se=T)
plot(age ,wage ,col = "gray ")
lines(age.grid , pred2$fit ,col ="red",lwd =2)

# AGE 55, 104.87425

fits3=smooth.spline(age ,wage ,cv=FALSE)
pred3=predict(fits3 ,newdata =list(age=age.grid),se=T)
pred3
# AGE 55,101.98295
fits4=loess(wage~age ,span =.5, data=Wage)
pred4=predict(fits4 ,newdata =data.frame(age=age.grid),se=T)
pred4
# AGE 55,101.42173
# Model 4

# Question 5
library(gam)
gam.m3=gam(wage~year+s(age ,5)+education ,data=Wage)
coef(gam.m3)
# The coefficient for year is twice as large as that of age. Both are small relative to the dummy education
variables.
# Question 6
library(gam)
education<=factor(education)
year_1<-year==2008
age_1<-age==48
age_s<-s(age_1, 5)
educ.1<-education==5
gam.m4=gam.lo.i=gam(wage~lo(year_1 ,age_s) +educ.1 ,
                    data=Wage)
coef(gam.m4)
predm4=predict(gam.m4,newdata =Wage)
0
# Question 7
library(gam)
year_1<-year==2008
age_1<-age==48
age_s<-s(age_1, 3)
educ.1<-education==5
gam.m5=gam(wage~year_1+age_s+education ,data=Wage)
coef(gam.m5)
```