```
##################################
##第十二回「一般化加法モデル」
##################################


#############################
## 脊柱後弯症
#install.packages("mgcv")
#install.packages("gam")
#library(gam)
library(mgcv)
data(kyphosis,package="gam")

#一変量, 判別
plot(kyphosis$Age,kyphosis$Kyphosis=="present",xlab="Age",lwd=2,col="blue",ylab=
"発病-健常")
kfit <- gam(Kyphosis ~ s(Age,k=10,fx=F), scale=-10,data=kyphosis,family=binomial
(link=logit))

kpred<-predict(kfit,type="response", se.fit = TRUE)
upper = kpred$fit + (1.2 * kpred$se.fit)
lower = kpred$fit - (1.2 * kpred$se.fit)

res <- sort(kyphosis$Age,index.return=TRUE)
fitted <- kpred$fit[res$ix]
upper_s <- upper[res$ix]
lower_s <- lower[res$ix]

plot(fitted ~ res$x,type="l",lwd=2,col="red",ylim = c(0,1),xlab="Age",ylab="p")
lines(res$x,upper_s,type="l",lty="dashed",lwd=1.5)
lines(res$x,lower_s,type="l",lty="dashed",lwd=1.5)
points(kyphosis$Age,kyphosis$Kyphosis=="present",lwd=2,col="blue")

lm.kfit<-gam(Kyphosis ~ Age, scale=-10,data=kyphosis,family=binomial(link=logit)
)        #  線形判別

summary(kfit)
summary(lm.kfit)
# GCVは良くなっている.

# 2変量
Y = (kyphosis$Kyphosis == "absent")
xt <- kyphosis[Y,2:4]
xf <- kyphosis[!Y,2:4]
plot(xt[c(1,3)],col="red",lwd=2)
points(xf[c(1,3)],col="blue",lwd=2,pch=3)

kfit2 <- gam(Kyphosis ~ s(Age,Start,k=10,fx=F),data=kyphosis,scale=-10,family=bi
nomial(link=logit))


imsize <- 40
xim <- seq(min(kyphosis[,2]),max(kyphosis[,2]),length=imsize)
yim <- seq(min(kyphosis[,4]),max(kyphosis[,4]),length=imsize)

zim <- matrix(c(1:(imsize*imsize)),nrow=imsize)
xcount <- 0
for(xxim in xim){
    xcount <- xcount + 1
    ycount <- 0
    for(yyim in yim){
        ycount <- ycount + 1
        zim[xcount,ycount] <- predict(kfit2,newdata = data.frame(Age=c(xxim),Sta
rt=c(yyim)),type="response")
    }
}

image(xim, yim, zim, col = terrain.colors(100), axes = FALSE,xlab="Age",ylab="St
art")
par(new=T)
contour(zim, method = "edge", vfont = c("sans serif", "plain"))
summary(kfit2)
```

```
# 1変量の場合と比べ, GCVは良くなっている.
# Age,Startの交互作用は有意


#############################
## カリフォルニア住宅価格
## library(oce): プロット用
## library(maps): 地図プロット用
library(oce)
library(maps)

chouse <- read.csv("cal_house.csv",header=TRUE)
#log(MedHouseValue) (住宅価格の対数を回帰)
lmfit <- gam(log(MedHouseValue) ~ MedIncome + MedHouseAge + TotalRooms
+ TotalBedrooms + Population + Households + Latitude + Longitude, data=chouse)
qqnorm(lmfit$residuals)

limpredictions = predict(lmfit,se.fit=TRUE)

#プロット用関数の定義, 推定値と実測値の関係をプロット
plot_housepred <- function(actual_val,pred_val,pred_se){
    plot(actual_val,exp(pred_val),cex=0.1,xlab="Actual price",ylab="Predicted")
    segments(actual_val,exp(pred_val-2*pred_se),
        actual_val,exp(pred_val+2*pred_se), col="grey")
    abline(a=0,b=1,lty=2)
}

plot_housepred(chouse$MedHouseValue,limpredictions$fit,limpredictions$se.fit)
#結構外れている.
summary(lmfit)
# GCV: 0.11568
sd(lmfit$residuals)
# Residual standard error: 0.34

#加法モデルを当てはめる
addfit <- gam(log(MedHouseValue) ~ s(MedIncome)
+ s(MedHouseAge) + s(TotalRooms)
+ s(TotalBedrooms) + s(Population) + s(Households)
+ s(Latitude) + s(Longitude), data=chouse)

predictions1 = predict(addfit,se.fit=TRUE)
plot_housepred(chouse$MedHouseValue,predictions1$fit,predictions1$se.fit)
#だいぶ当てはまりは良くなっている.

summary(addfit)
# GCV: 0.08, かなりの改善
sd(addfit$residuals)
# Residual standard error: 0.29

plot(addfit,scale=0,se=2,shade=TRUE,pages=1)

res <- anova(addfit)
barplot(res$chi.sq,ylab="chi2") #寄与率
barplot(res$chi.sq/res$edf,ylab="F") #F値

#緯度と経度の交互作用を取り入れる
addfit2 <- gam(log(MedHouseValue) ~ s(MedIncome) + s(MedHouseAge)
+ s(TotalRooms) +s(TotalBedrooms) + s(Population) + s(Households)
+ s(Longitude,Latitude), data=chouse)

predictions2 = predict(addfit2,se.fit=TRUE)
plot_housepred(chouse$MedHouseValue,predictions2$fit,predictions2$se.fit)

summary(addfit2)
# GCV: 0.07. 交互作用なしと比べ良くなっている.
sd(addfit2$residuals)
# Residual standard error: 0.27

res <- anova(addfit2)
barplot(res$chi.sq,ylab="chi2") #寄与率, 緯度経度の交互作用は寄与率が高い.
barplot(res$chi.sq/res$edf,ylab="F") #F値
```

```
x <- map('county', 'california', fill = FALSE)
plot(addfit2,select=7,se=FALSE,lwd=2,main="House Prise by Latitude x Longitude")

lines(x,lwd=1.2,col="red")


predictions2 = predict(addfit2,type="terms") # type="terms"とすると各変数ごとの
予測値が得られる.
predval <- predictions2[,7] #緯度経度の交互作用のみを取り出す
colvec <- colormap(predval)
drawPalette(colvec$zlim, col = colvec$col, breaks = colvec$breaks)
plot(x,lwd=2,col=5,type="l",xlab="Longitute",ylab="Latitude") #カリフォルニア州
をプロット
points(chouse$Longitude,chouse$Latitude,bg = colvec$zcol, pch = 21, cex = 1) #緯
度経度のみで説明された住宅価格


predictions1 = predict(addfit,type="terms") # 交互作用なし, 予測値

mean((log(chouse$MedHouseValue) - (attr(predictions2,"constant") + predictions2[
,7]))^2)
mean((log(chouse$MedHouseValue) - (attr(predictions1,"constant") + predictions1[
,7] + predictions1[,8]))^2)
```