Addiction - Additive Multinomial Logit Models

February 5, 2020

First the "Addiction"–Data from "catdata" are loaded and attached.

```
> library(catdata)
> data(addiction)
> attach(addiction)
```

For the fitting of GAMs the library "mgcv" is used.

```
> library(mgcv)
```

Now we create two data frames that will be used to predict the probabilities along the range of age, the columns are "age", "gender" and "university". We create one data frame for women and one for men.

```
> minage <- min(na.omit(age))
> maxage <- max(na.omit(age))
> ageindex <- seq(minage, maxage, 0.1)
> n <- length(ageindex)

> gender1 <- rep(1, n)
> gender0 <- rep(0, n)
> university1 <- rep(1, n)
> university0 <- rep(0, n)
> datamale <- as.data.frame(cbind(gender=gender0,age=ageindex,university=university1))
> datafemale <- as.data.frame(cbind(gender=gender1,age=ageindex,
+                                   university=university1))
```

For the hierarchical model a new reponse "ill01" is created.

```
> ill01 <-ill
> ill01[ill==1] <- 0
> ill01[ill==2] <- 1
```

Now the two GAMs for the hierarchical model are fitted, "gam1" models category "0" and category "1" versus category "2", "gam2" models category "0" versus category "1".

```
> gam1 <- gam(as.factor(ill01) ~ s(age) + gender + university, family=binomial())
> gam2 <- gam(as.factor(ill) ~ s(age) + gender + university, family=binomial(),
+             data=addiction[addiction$ill!=2,])
```

Then the corresponding summaries are printed.

```
> summary(gam1)

Family: binomial
Link function: logit

Formula:
as.factor(ill01) ~ s(age) + gender + university

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.1545     0.1486   -7.77     8e-15 ***
gender        0.0256     0.1835    0.14      0.89
university   -0.1306     0.2098   -0.62      0.53
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
       edf Ref.df Chi.sq p-value
s(age) 3.53   4.41   20.8 0.00049 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0266   Deviance explained = 3.17%
UBRE = 0.093898  Scale est. = 1          n = 682

> summary(gam2)

Family: binomial
Link function: logit

Formula:
as.factor(ill) ~ s(age) + gender + university

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.175      0.159    -1.1   0.2702
gender         0.604      0.208     2.9   0.0037 **
university     1.376      0.265     5.2    2e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
       edf Ref.df Chi.sq p-value
s(age) 4.02      5   50.7   1e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.196   Deviance explained = 16.3%
UBRE = 0.14577  Scale est. = 1          n = 516
```
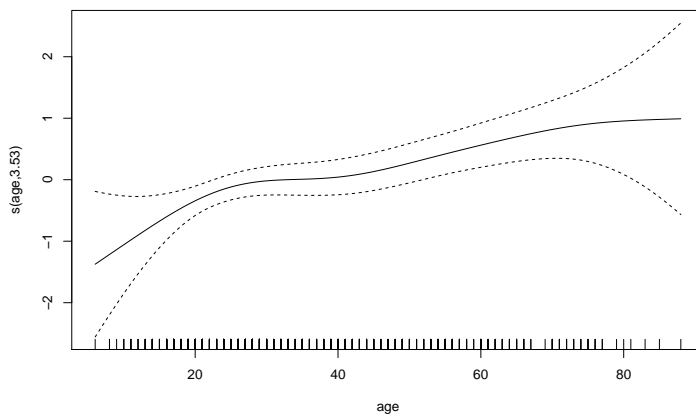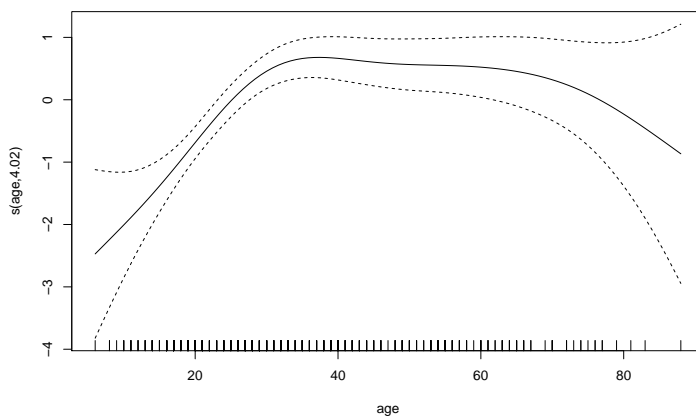
Now the smoothed effects for age are plotted.

```
> plot(gam1)
```



```
> plot(gam2)
```


For predicting the probabilities for the respective categories the probabilities from both GAMs are needed. These can be computed by the "predict"–function. First we use the data frame for men for prediction.

```
> predmale1 <- predict(gam1, newdata=datamale, type="response")
> predmale2 <- predict(gam2, newdata=datamale, type="response")

> predfemale1 <- predict(gam1, newdata=datafemale, type="response")
> predfemale2 <- predict(gam2, newdata=datafemale, type="response")
```

Then we compute the probabilities for each category, first for men, afterwards for women.

```
> p2 <- predmale1
> p1 <- predmale2 * (1-predmale1)
> p0 <- (1-predmale2) * (1-predmale1)
```

3

```
> pf2 <- predfemale1
> pf1 <- predfemale2 * (1-predfemale1)
> pf0 <- (1-predfemale2) * (1-predfemale1)
```

Now finally the probabilities for the respective categories can be plotted.

```
> par(mfrow=c(1,2), cex=1.8)
> plot(ageindex, p0, type="l", lty=1, ylim=c(0,1), main="men with university degree",
+       ylab="probabilities")
> lines(ageindex, p1, lty="dotted")
> lines(ageindex, p2, lty="dashed")
> legend("topright", legend=c("Weak-willed", "diseased", "both"),
+         lty=c("solid", "dotted", "dashed"))
> plot(ageindex, pf0, type="l", lty=1, ylim=c(0,1),
+       main="women with university degree", ylab="probabilities")
> lines(ageindex, pf1, lty="dotted")
> lines(ageindex, pf2, lty="dashed")
> legend("topright", legend=c("Weak-willed", "diseased", "both"),
+         lty=c("solid", "dotted", "dashed"))
```
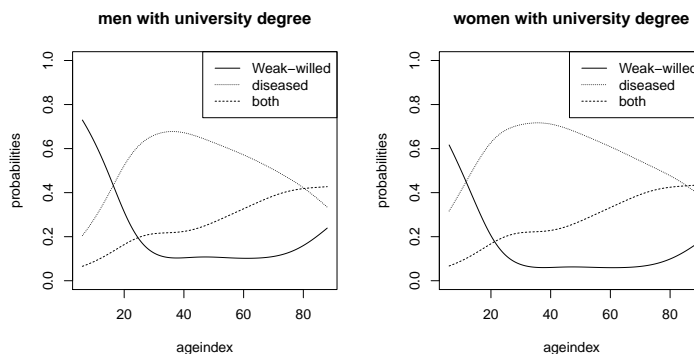

The models "gam3" and "gam4" compare category 0 versus 1 and category 0 versus 2 respectively.

```
> gam3 <- gam(as.factor(ill)~ s(age) + gender + university,
+             data=addiction[addiction$ill!=2,], family=binomial())
> gam4 <- gam(as.factor(ill)~ s(age) + gender + university,
+             data=addiction[addiction$ill!=1,], family=binomial())

> summary(gam3)

Family: binomial
Link function: logit


Formula:
as.factor(ill) ~ s(age) + gender + university


Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.175      0.159    -1.1   0.2702
```

```
gender         0.604      0.208      2.9   0.0037 **
university     1.376      0.265      5.2   2e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
        edf Ref.df Chi.sq p-value
s(age) 4.02      5   50.7   1e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.196   Deviance explained = 16.3%
UBRE = 0.14577  Scale est. = 1          n = 516

> summary(gam4)

Family: binomial
Link function: logit

Formula:
as.factor(ill) ~ s(age) + gender + university

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.572      0.179   -3.20   0.0014 **
gender         0.344      0.231    1.49   0.1363
university     0.784      0.304    2.58   0.0098 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
      edf Ref.df Chi.sq p-value
s(age) 3.9   4.86   38.6   3e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.141   Deviance explained = 12.2%
UBRE = 0.24647  Scale est. = 1          n = 367
```