

7.1.1 Generalized Additive Models

Stat 5100: Dr. Bean

Example: Baseball Dataset from 4.1.1

Let's see if we can improve upon the penalized linear regression model to predict the log of salary for professional (non-pitcher) baseball players. Note that answers will differ slightly depending on the random seed set.

```
# Set a random seed for reproducibility
set.seed(830578)

# Load data
library(stat5100)
data(baseball)

baseball_gam <-
  gam::gam(logSalary ~ gam::s(nAtBat) + gam::s(nHits) + gam::s(nHome) +
    gam::s(nRuns) + gam::s(nRBI) + gam::s(nBB) + gam::s(YrMajor) +
    gam::s(CrAtBat) + gam::s(CrHits) + gam::s(CrHome) + gam::s(CrRuns) +
    gam::s(CrRbi) + gam::s(CrBB) + gam::s(nOuts) + gam::s(nAssts) +
    gam::s(nError) + League + Division,
    data = baseball)

summary(baseball_gam)

##
## Call: gam::gam(formula = logSalary ~ gam::s(nAtBat) + gam::s(nHits) +
##   gam::s(nHome) + gam::s(nRuns) + gam::s(nRBI) + gam::s(nBB) +
##   gam::s(YrMajor) + gam::s(CrAtBat) + gam::s(CrHits) + gam::s(CrHome) +
##   gam::s(CrRuns) + gam::s(CrRbi) + gam::s(CrBB) + gam::s(nOuts) +
##   gam::s(nAssts) + gam::s(nError) + League + Division, data = baseball)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.24159 -0.44188  0.07915  0.39921  1.22576
##
## (Dispersion Parameter for gaussian family taken to be 0.3252)
##
##      Null Deviance: 207.1537 on 262 degrees of freedom
## Residual Deviance: 79.3583 on 244 degrees of freedom
## AIC: 471.2402
## 59 observations deleted due to missingness
##
## Number of Local Scoring Iterations: 2
##
## Anova for Parametric Effects
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## gam::s(nAtBat)    1 44.183   44.183 135.8476 < 2.2e-16 ***
## gam::s(nHits)     1  6.420    6.420  19.7403 1.346e-05 ***
## gam::s(nHome)     1  4.323    4.323  13.2924 0.0003257 ***
```

```

## gam::s(nRuns)      1  0.022   0.022   0.0682 0.7941343
## gam::s(nRBI)       1  2.102   2.102   6.4644 0.0116250 *
## gam::s(nBB)        1 12.286  12.286  37.7767 3.210e-09 ***
## gam::s(YrMajor)    1 52.393  52.393 161.0912 < 2.2e-16 ***
## gam::s(CrAtBat)    1  0.823   0.823   2.5291 0.1130594
## gam::s(CrHits)     1  0.180   0.180   0.5538 0.4574779
## gam::s(CrHome)     1  0.004   0.004   0.0131 0.9090009
## gam::s(CrRuns)     1  0.008   0.008   0.0252 0.8740452
## gam::s(CrRbi)      1  0.043   0.043   0.1311 0.7175796
## gam::s(CrBB)       1  1.453   1.453   4.4661 0.0355898 *
## gam::s(nOuts)      1  1.109   1.109   3.4097 0.0660247 .
## gam::s(nAssts)     1  0.000   0.000   0.0015 0.9691021
## gam::s(nError)     1  0.331   0.331   1.0184 0.3138967
## League             1  0.786   0.786   2.4175 0.1212795
## Division           1  1.327   1.327   4.0812 0.0444556 *
## Residuals          244 79.358   0.325
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```