

Linear Model for IDMB data

ASM Homework 1

Ricard Monge Amalia Vradi Maria Gkotsopoulou

3/10/2019

Exploratory data analysis

```
data <- read.table(file="IMDB.csv", header = T, sep = ";")
summary(data)
```

```
##              movietitle      gross      budget
## 10 Days in a Madhouse: 1   Min.   :    8691   Min.   :   427000
## 12 Years a Slave      : 1   1st Qu.: 12549485   1st Qu.: 10000000
## 13 Going on 30        : 1   Median : 33451479   Median : 23000000
## 21 & Over             : 1   Mean    : 55476041   Mean    : 39843632
## 21 Grams              : 1   3rd Qu.: 69951824   3rd Qu.: 45000000
## 25th Hour             : 1   Max.    :760505847   Max.    :300000000
## (Other)               :391
##      duration      titleyear      directorfl      actor1fl
## Min.   : 74.0   Min.   :2000   Min.   :    0.0   Min.   :    11
## 1st Qu.: 94.0   1st Qu.:2003   1st Qu.:   13.0   1st Qu.:   834
## Median :103.0   Median :2008   Median :   59.0   Median :   2000
## Mean   :107.2   Mean   :2008   Mean   :  870.1   Mean   :   9872
## 3rd Qu.:117.0   3rd Qu.:2012   3rd Qu.:  221.0   3rd Qu.:  14000
## Max.   :216.0   Max.   :2016   Max.   :22000.0   Max.   : 640000
##
##      actor2fl      actor3fl      castfl      facenumber_in_poster
## Min.   :    8   Min.   :    3.0   Min.   :    37   Min.   : 0.000
## 1st Qu.:  431   1st Qu.:  249.0   1st Qu.:  2413   1st Qu.: 0.000
## Median :  762   Median :  495.0   Median :  4730   Median : 1.000
## Mean   : 2183   Mean   :  973.3   Mean   : 14176   Mean   : 1.743
## 3rd Qu.: 1000   3rd Qu.:  766.0   3rd Qu.: 18322   3rd Qu.: 2.000
## Max.   :21000   Max.   :19000.0   Max.   :656730   Max.   :19.000
##
##      genre
## Action: 46
## Comedy:166
## Drama :130
## Terror: 55
##
##
```

Model selection

Complete model

Optimal model

Multicollinearity

Model validation

Model interpretation