

```
#### Naive Bayes
```

```
library(e1071)
delays.df <- read.csv("FlightDelays.csv")

# change numerical variables to categorical first
delays.df$DAY_WEEK <- factor(delays.df$DAY_WEEK)
delays.df$DEP_TIME <- factor(delays.df$DEP_TIME)
delays.df$Flight_Status <- factor(delays.df$Flight_Status)
# create hourly bins departure time
delays.df$CRS_DEP_TIME <- factor(round(delays.df$CRS_DEP_TIME/100))
str(delays.df)
# Create training and validation sets.
selected.var <- c(10, 1, 8, 4, 2, 13)
train.index <- sample(c(1:dim(delays.df)[1]), dim(delays.df)[1]*0.6)
train.df <- delays.df[train.index, selected.var]
valid.df <- delays.df[-train.index, selected.var]
# run naive bayes
delays.nb <- naiveBayes(Flight_Status ~ ., data = train.df)
delays.nb
```

```
#### Table 8.5
```

```
# use prop.table() with margin = 1 to convert a count table to a proportion table,
# where each row sums up to 1 (use margin = 2 for column sums).
prop.table(table(train.df$Flight_Status, train.df$DEST), margin = 1)
```

```
#### Table 8.6
```

```
## predict probabilities
pred.prob <- predict(delays.nb, newdata = valid.df, type = "raw")
## predict class membership
pred.class <- predict(delays.nb, valid.df)

df <- data.frame(actual = valid.df$Flight_Status, predicted = pred.class, pred.prob)

df[valid.df$CARRIER == "DL" & valid.df$DAY_WEEK == 7 & valid.df$CRS_DEP_TIME == 10 &
  valid.df$DEST == "LGA" & valid.df$ORIGIN == "DCA",]
```

```
library(caret)
```

```
# training
pred.class <- predict(delays.nb, newdata = train.df)
confusionMatrix(pred.class, train.df$Flight_Status)
```

```
# validation
pred.class <- predict(delays.nb, newdata = valid.df)
confusionMatrix(pred.class, valid.df$Flight_Status)
```

```
#### Figure 8.1
```

```
library(gains)
gain <- gains(iffelse(valid.df$Flight_Status=="delayed",1,0), pred.prob[,1], groups=100)

plot(c(0,gain$cume.pct.of.total*sum(valid.df$Flight_Status=="delayed"))~c(0,gain
$cume.obs),
      xlab="# cases", ylab="Cumulative", main="", type="l")
lines(c(0,sum(valid.df$Flight_Status=="delayed"))~c(0, dim(valid.df)[1]), lty=2)
```