Model Selection

This is an R markdown document.

```
library(ISLR)
summary(Hitters)
```

There are some missing values here, so before we proceed we will remove them:

```
Hitters=na.omit(Hitters)
with(Hitters, sum(is.na(Salary)))
```

Best subset regression

Best subset regression looks through all possible regression models of all different subset sizes and looks for the best of each size. And so produces a sequence of models which is the best subset for each particular size.

```
library(leaps)
regfit.full=regsubsets(Salary~.,data=Hitters)
summary(regfit.full)
```

It gives by default best subsets up to size 8; lets increase that to 19

```
regfit.full=regsubsets(Salary~., data=Hitters, nvmax=19)
reg.summary=summary(regfit.full)
names(reg.summary)
plot(reg.summary$cp, xlab="number of Variables", ylab="Cp")
which.min(reg.summary$cp)
points(10, reg.summary$cp[10], pch=20, col="red")
```

There is a plot method for the `regsubsets` object

```
plot(regfit.full, scale="Cp")
coef(regfit.full, 10)
```

Forward Stepwise Selection