Ordinary Least Squares

Giri Iyengar

Ordinary Least Squares: A Linear Algebra Perspective

Ordinary Least Squares can be viewed from many different viewpoints. At this time, we'll view it from the lens of Linear Algebra and the Systems of Equations perspective. Later on in this course, and in subsequent courses, you'll encounter this again from a Machine Learning and Optimization perspective.

The set-up

- Just like a normal system of equations, you can pose OLS as the following: $\mathbf{Ax} = \mathbf{b}$

The set-up

- Just like a normal system of equations, you can pose OLS as the following: $\mathbf{A}\mathbf{x} = \mathbf{b}$
- However, instead of having m equations and m variables, we have m equations and n variables, where $m > n$

The set-up

- Just like a normal system of equations, you can pose OLS as the following: $\mathbf{Ax} = \mathbf{b}$
- However, instead of having m equations and m variables, we have m equations and n variables, where $m > n$
- More constraints than variables. Also, safely assume that $\mathbf{A}$ has a rank of n

The set-up

- Just like a normal system of equations, you can pose OLS as the following: $\mathbf{Ax} = \mathbf{b}$
- However, instead of having m equations and m variables, we have m equations and n variables, where $m > n$
- More constraints than variables. Also, safely assume that $\mathbf{A}$ has a rank of n
- No guarantee that we will be able to solve it exactly, as it is an over-constrained system

Ordinary Least Squares solution

- We cannot solve it exactly. Can we somehow convert it?

Ordinary Least Squares solution

- We cannot solve it exactly. Can we somehow convert it?
- But, we can solve $\mathbf{A^T A x} = \mathbf{A^T b}$, a related system of equations

Ordinary Least Squares solution

- We cannot solve it exactly. Can we somehow convert it?
- But, we can solve $\mathbf{A^T A x} = \mathbf{A^T b}$, a related system of equations
- This turns out to be a regular $n \times n$ system of equations

$A^T Ax = A^T b$: Insights

- We cannot solve for $\mathbf{x}$ as $\mathbf{b}$ is not in the column space of $\mathbf{A}$

$A^T Ax = A^T b$: Insights

- We cannot solve for $\mathbf{x}$ as $\mathbf{b}$ is not in the column space of $\mathbf{A}$
- Can we solve for $\hat{\mathbf{x}}$ such that $\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$ is as small as possible?

$A^T A x = A^T b$: Insights

- We cannot solve for $\mathbf{x}$ as $\mathbf{b}$ is not in the column space of $\mathbf{A}$
- Can we solve for $\hat{\mathbf{x}}$ such that $\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$ is as small as possible?
- The best we can do is to compute $\mathbf{A}\hat{\mathbf{x}} = \mathbf{p}$ where $\mathbf{p}$ is the projection of $\mathbf{b}$ in the column space

$A^TAx = A^Tb$: Insights

- We cannot solve for $\mathbf{x}$ as $\mathbf{b}$ is not in the column space of $\mathbf{A}$
- Can we solve for $\hat{\mathbf{x}}$ such that $\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$ is as small as possible?
- The best we can do is to compute $\mathbf{A}\hat{\mathbf{x}} = \mathbf{p}$ where $\mathbf{p}$ is the projection of $\mathbf{b}$ in the column space
- Can we find the optimal projection of $\mathbf{b}$ in $\mathbf{A}$'s column space? Turns out, we can

Optimal projection

- The optimal projection of $\mathbf{b}$ (i.e. $\mathbf{p}$) is the one that has the minimum $\mathbf{e} = \mathbf{b} - \mathbf{p}$

Optimal projection

- The optimal projection of $\mathbf{b}$ (i.e. $\mathbf{p}$) is the one that has the minimum $\mathbf{e} = \mathbf{b} - \mathbf{p}$
- This happens when $\mathbf{e}$ is perpendicular (orthogonal) to the column space of $\mathbf{A}$

Optimal projection

- The optimal projection of $\mathbf{b}$ (i.e. $\mathbf{p}$) is the one that has the minimum $\mathbf{e} = \mathbf{b} - \mathbf{p}$
- This happens when $\mathbf{e}$ is perpendicular (orthogonal) to the column space of $\mathbf{A}$
- This happens when $\mathbf{e}$ is the left null space of $\mathbf{A}$. That is $\mathbf{A^T e} = \mathbf{0}$

Optimal projection

- The optimal projection of $\mathbf{b}$ (i.e. $\mathbf{p}$) is the one that has the minimum $\mathbf{e} = \mathbf{b} - \mathbf{p}$
- This happens when $\mathbf{e}$ is perpendicular (orthogonal) to the column space of $\mathbf{A}$
- This happens when $\mathbf{e}$ is the left null space of $\mathbf{A}$. That is $\mathbf{A^T e} = \mathbf{0}$
- We can see that $\mathbf{A^T e} = \mathbf{A^T}(\mathbf{b} - \mathbf{A\hat{x}}) = \mathbf{0}$. Therefore, $\mathbf{A^T A \hat{x}} = \mathbf{A^T b}$

OLS Demo – I: Data

- Load up some Baseball data. Lahman package contains baseball data from 1871 to 2012

```
library(Lahman)
data(Teams)
baseball.teams <- Teams
rownames(baseball.teams) <- paste(baseball.teams$teamID,
                                  baseball.teams$yearID,
                                  sep="-")
```

OLS Demo – II: Data

```
summary(baseball.teams)
```

```
##       yearID      lgID        teamID          franchID
## Min.    :1871   AA:  85   CHN     : 139   ATL     : 139   Len
## 1st Qu.:1919   AL:1190   PHI     : 132   CHC     : 139   Cla
## Median :1962   FL:  16   PIT     : 128   CIN     : 133   Moc
## Mean   :1954   NA:  50   CIN     : 125   PIT     : 133
## 3rd Qu.:1991   NL:1414   SLN     : 123   STL     : 133
## Max.   :2014   PL:   8   BOS     : 114   PHI     : 132
##                UA:  12   (Other):2014   (Other):1966
##       Rank               G              Ghome              W
## Min.   : 1.000   Min.   :  6.0   Min.   :44.00   Min.   :
## 1st Qu.: 2.000   1st Qu.:153.0   1st Qu.:77.00   1st Qu.:
## Median : 4.000   Median :157.0   Median :80.00   Median :
## Mean   : 4.119   Mean   :150.2   Mean   :78.43   Mean   :
## 3rd Qu.: 6.000   3rd Qu.:162.0   3rd Qu.:81.00   3rd Qu.:
## Max.   :13.000   Max.   :165.0   Max.   :84.00   Max.   :
```

OLS Demo – III: Data

```r
head(baseball.teams)
```

```
##           yearID lgID teamID franchID divID Rank  G Ghome  W
## BS1-1871   1871   NA    BS1      BNA  <NA>    3 31    NA 20
## CH1-1871   1871   NA    CH1      CNA  <NA>    2 28    NA 19
## CL1-1871   1871   NA    CL1      CFC  <NA>    8 29    NA 10
## FW1-1871   1871   NA    FW1      KEK  <NA>    7 19    NA  7
## NY2-1871   1871   NA    NY2      NNA  <NA>    5 33    NA 16
## PH1-1871   1871   NA    PH1      PNA  <NA>    1 28    NA 21
##           WCWin LgWin WSWin   R   AB   H X2B X3B HR BB SO SB
## BS1-1871  <NA>     N  <NA> 401 1372 426  70  37  3 60 19 73
## CH1-1871  <NA>     N  <NA> 302 1196 323  52  21 10 60 22 69
## CL1-1871  <NA>     N  <NA> 249 1186 328  35  40  7 26 25 18
## FW1-1871  <NA>     N  <NA> 137  746 178  19   8  2 33  9 16
## NY2-1871  <NA>     N  <NA> 302 1404 403  43  21  1 33 15 46
## PH1-1871  <NA>     N  <NA> 376 1281 410  66  27  9 46 23 56
##            ER  ERA CG SHO SV IPouts  HA HRA BBA SOA  E DP
```

OLS Demo – IV: Data

- Extract a small sample to play with

```
years.00.11 <- which(baseball.teams$yearID>1999 & baseball.tea
years.90.99 <- which(baseball.teams$yearID>1989 & baseball.tea
years.12 <- which(baseball.teams$yearID==2012)
vars.interest <- c("H","HR","BB","ERA","BBA","W")

## subset to 1999-2011 and above variables
## second data set of 2012 data that we'll use for testing
baseball.teams.2k   <- baseball.teams[years.00.11,vars.interes
baseball.teams.90   <- baseball.teams[years.90.99,vars.interes
baseball.teams.test <- baseball.teams[years.12,vars.interest]
```

OLS Demo – V: Solve for $\hat{\mathbf{x}}$

- Form $\mathbf{A}$ and $\mathbf{b}$ from your training data

```
A2k <- as.matrix(baseball.teams.2k[,c("H","HR","BB","ERA","BBA
b2k <- as.matrix(baseball.teams.2k[,"W"])
A90 <- as.matrix(baseball.teams.90[,c("H","HR","BB","ERA","BBA
b90 <- as.matrix(baseball.teams.90[,"W"])
```

OLS Demo – V: Solve for $\hat{\mathbf{x}}$

- Solve for $\hat{\mathbf{x}}$. Do it as $(\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{b}$

```
x_hat2k <- solve(t(A2k) %*% A2k) %*% (t(A2k) %*% b2k)
x_hat2k
```

```
##             [,1]
## H     0.07919033
## HR    0.08542453
## BB    0.03997711
## ERA -15.45282711
## BBA  -0.00728550
```

OLS Demo – V: Solve for $\hat{\mathbf{x}}$

```
x_hat90 <- solve(t(A90) %*% A90) %*% (t(A90) %*% b90)
x_hat90
```

```
##            [,1]
## H    0.06502233
## HR   0.04776939
## BB   0.05395825
## ERA -7.47339393
## BBA -0.03375534
```

OLS Demo – VI: Evaluate the fit

```
Atest <- as.matrix(baseball.teams.test[,c("H","HR","BB","ERA",
btest <- as.matrix(baseball.teams.test[,"W"])
bpred <- Atest %*% x_hat2k
rmse <- function(obs, pred) sqrt(mean((obs-pred)^2))
rmse(btest,bpred) / mean(btest)
```

```
## [1] 0.05781108
```

```
cor(bpred,btest)
```

```
##            [,1]
## [1,] 0.9305061
```

OLS Demo – VI: Evaluate the fit

```r
bpred90 <- Atest %*% x_hat90
rmse(btest,bpred90) / mean(btest)
```

```
## [1] 0.08274209
```

```r
cor(bpred90,btest)
```

```
##           [,1]
## [1,] 0.8568212
```

- Our best model gets about 6% error. Our predictions are 94% accurate

OLS Demo – VII: Evaluate the fit