Model Building and Fitting

David T. Frazier

30/08/2017

Outline

- What is `optim` doing?
- Optimization of functions with one variable
- Optimization of functions with two variables
- Using `optim`

Optimization of functions with one variable

- Goal is to find min or max of $f(b)$, continuous in b,
- Domain of the function is $[a_0, a_1]$, values b can take
- Range of the function is $\mathbb{R}$
- *Extreme Value Theorem*
  - $f(b)$ is continuous for all b in $[a_0, a_1]$
  - then $f(b)$ obtains a min and max on $[a_0, a_1]$

Optimization: one variable

- Optimum can occur at
  - Boundary of D; i.e., a_0 or a_1
  - Point at which the function is not differentiable
  - A value $\hat{b}$ where $f'(\hat{b}) = 0$
- The point $f'(b) = 0$ is called the **"first-order condition" (FOC)**

Optimization: one variable

- Let f and f' be continuous on $[a_0, a_1]$.
- If there is a $\hat{b} \in (a_0, a_1)$ such that $f'(\hat{b}) = 0$ **and** $f''(\hat{b}) \neq 0$, $\hat{b}$ is an *optimum*
- Optimum can be max or min!
- How to differentiate?
  - if $f''(\hat{b}) > 0$, $f(b)$ has a local min at $b = \hat{b}$
  - if $f''(\hat{b}) < 0$, $f(b)$ has a local max at $b = \hat{b}$
- Sign of $f''(\hat{b})$ is called the **"second-order condition" (SOC)**

Example: polynomial

- $f(b) = 2 + 3b + 4b^2 - b^3$
- $f'(b) = 3 + 8b - 3b^2$
- Solve for $\hat{b}$ via quadratic formula!
- $0 = f'(b) = -3b^2 + 8b + 3$, implies $\hat{b} = \frac{-8 \pm \sqrt{8^2 - 4 * (-3) * 3}}{2(-3)}$
- Min or max?

Example: polynomial

- $f(b) = 2 + 3b + 4b^2 - b^3$
- $f'(b) = 3 + 8b - 3b^2$
- Solve for $\hat{b}$ via quadratic formula!
- $0 = f'(b) = -3b^2 + 8b + 3$, implies $\hat{b} = \frac{-8 \pm \sqrt{8^2 - 4*(-3)*3}}{2(-3)}$
- Min or max?
  - $f''(b) = 8 - 6b > 0$ for $b < 3/2$
  - $f''(b) = 8 - 6b < 0$ for $b > 3/2$

Example: Ordinary Least Squares (OLS) through origin

- $\min_{b_1} f(b) = \min_b \sum_i \{y_i - bx_i\}^2$
- Closed form solution
- $f'(b) = -2\sum_i x_i\{y_i - bx_i\}$
- $\hat{b} = \sum_i x_iy_i / \sum_i (x_i)^2$
- Min or max?

Example: Ordinary Least Squares (OLS) through origin

- $\min_{b_1} f(b) = \min_b \sum_i \{y_i - bx_i\}^2$
- Closed form solution
- $f'(b) = -2 \sum_i x_i \{y_i - bx_i\}$
- $\hat{b} = \sum_i x_i y_i / \sum_i (x_i)^2$
- Min or max?
  - SOC: $f''(b) = \sum_{i=1}^n x_i^2 > 0$

No Analytic Solution?

- Newton Method
- Can't solve $f'(b) = 0$ but know $f'(b)$
- Given guess $b^{(0)}$, updated guess according to

$$b^{(1)} = b^{(0)} - \frac{f'(b^{(0)})}{f''(b^{(0)})}$$

- At k-th step

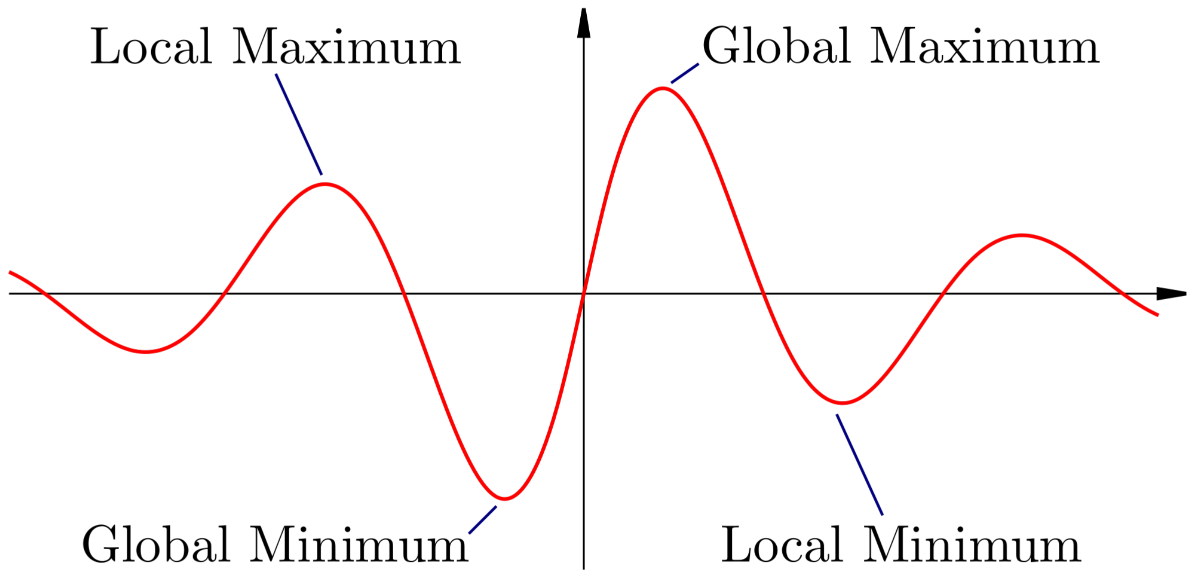
$$b^{(k)} = b^{(k-1)} - \frac{f'(b^{(k-1)})}{f''(b^{(k-1)})}$$

- `optim(f,init,method="BFGS")`, Newton-type approach

Newton: 60% of the time,

- It works ever time...
- Can fail, easily,
- What if $f''(b^{(k-1)}) = 0$?

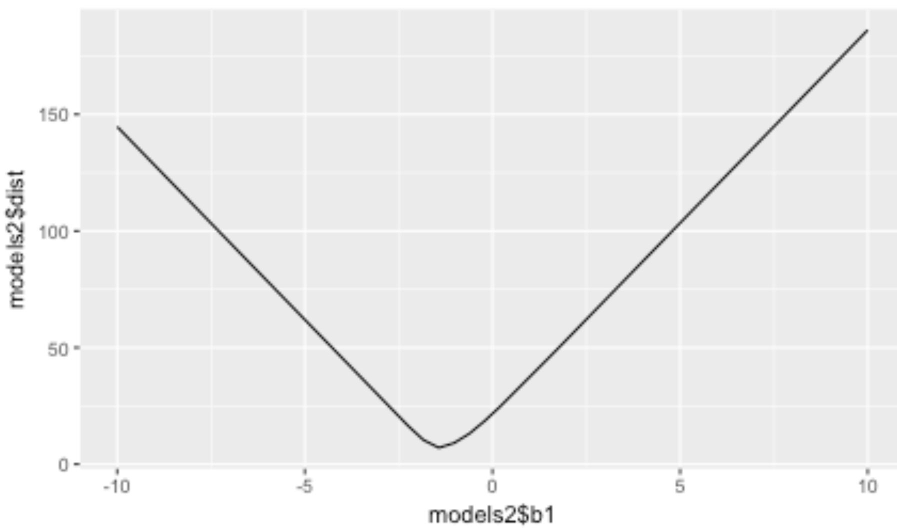
Newton: 60% of the time,



- Try multiple initial values.

What if $f'(b)$ does not exist?

- Take $\min_b f(b) = \min_b \sum_i |y_i - b \cdot x_i|$?
  - **Least Absolute Deviations**



No derivatives

- Recall, defn of derivative:

$$f'(c) = \lim_{h \to 0} \frac{f(c+h) - f(c)}{h}$$

- If $f(b) = |b|$, then at $b = 0$, $f'(0) = \lim_{h \to 0} \frac{|0+h| - |0|}{h} = \lim |h|/h$

. However,

  - Limit differs if $h > 0$ or $h < 0$
  - $\lim_{h \downarrow 0} |h|/h = 1$
  - $\lim_{h \uparrow 0} |h|/h = -1$
- Nelder-Mead algorithm doesn't use derivatives
- `optim(init,f,method="Nelder-Mead")`

14 / 25

Multivariate Optimization

- Just optimization with more than one variable
- $b = (\, b_0, b_1 \,)$
- Objective function: $f(b) = f(b_0, b_1)$

Ordinary Least Squares (OLS)

- Start with an example you've all seen...
- $\min_{b_0,b_1} \sum_i \{y_i - b_0 - b_1 x_i\}^2$
- This is multivariate optimization (in b_0, b_1)
- Closed form solutions
  - $\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}$
  - $\hat{b}_1 = \sum_i (y_i - \bar{y})(x_i - \bar{x}) / \sum_i (x_i - \bar{x})^2$
- How'd we get these?

OLS cont.

- Let $f(b_0, b_1) = \sum_i \{y_i - b_0 - b_1 x_i\}^2$
- Derivatives worked for univariate case...
  - $\frac{\partial f(b_0, b_1)}{\partial b_0} = -2 \sum_i \{y_i - b_0 - b_1 x_i\}$
  - $\frac{\partial f(b_0, b_1)}{\partial b_1} = -2 \sum_i x_i \{y_i - b_0 - b_1 x_i\}$
- Solving for zero gives us $\hat{b}_0$ and $\hat{b}_1$
- Built in function to do this in R `lm()`

Simple OLS in R

- `lm()` function fits linear models

```
  mod1 <- lm(y ~ x, data=data )
mod1$coefficients
```

```
## (Intercept)           x
##   53.013827   -2.057155
```

Simple OLS in R

Summary stats

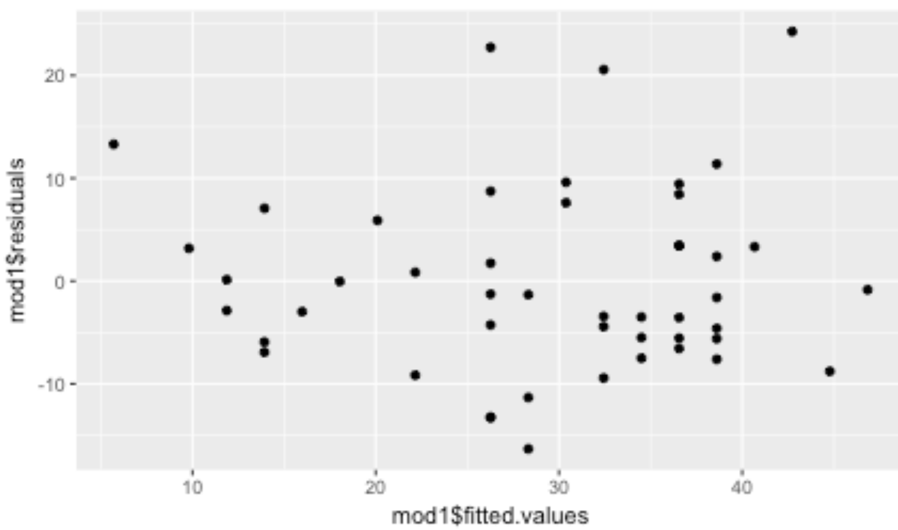
```
summary(mod1)
```

```
##
## Call:
## lm(formula = y ~ x, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.328  -5.692  -1.471   4.058  24.272
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  53.0138     3.4215  15.495  < 2e-16 ***
## x            -2.0572     0.2737  -7.516 1.54e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.103 on 46 degrees of freedom
## Multiple R-squared:  0.5512,    Adjusted R-squared:  0.5414
## F-statistic: 56.49 on 1 and 46 DF,  p-value: 1.544e-09
```

Simple OLS in R

- Predictions: $\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i$
- Residuals: $y_i - \hat{y}_i = y_i - \hat{b}_0 - \hat{b}_1 x_i$
  - Remember, want uniform residual cloud!
- `ggplot(mod1,`

`aes(x=mod1$fitted.values,y=mod1$residuals))+geom_point()`



Back to general optimization

- Multivariate case similar to univariate case
- To find optimum, say of $f(b_0, b_1)$, take partial derivatives and solve
    - $\partial f(b_0, b_1)/\partial b_0 = 0$
    - $\partial f(b_0, b_1)/\partial b_1 = 0$
- Will $\hat{b} = (\hat{b}_0, \hat{b}_1)'$ be a min or max?
- Second order conditions!

Back to optimization

- With two variables, three second order conditions to check
  - SOC1: $\partial^2 f(b_0, b_1)/\partial b_0^2$
  - SOC2: $\partial^2 f(b_0, b_1)/\partial b_1$
  - SOC3:

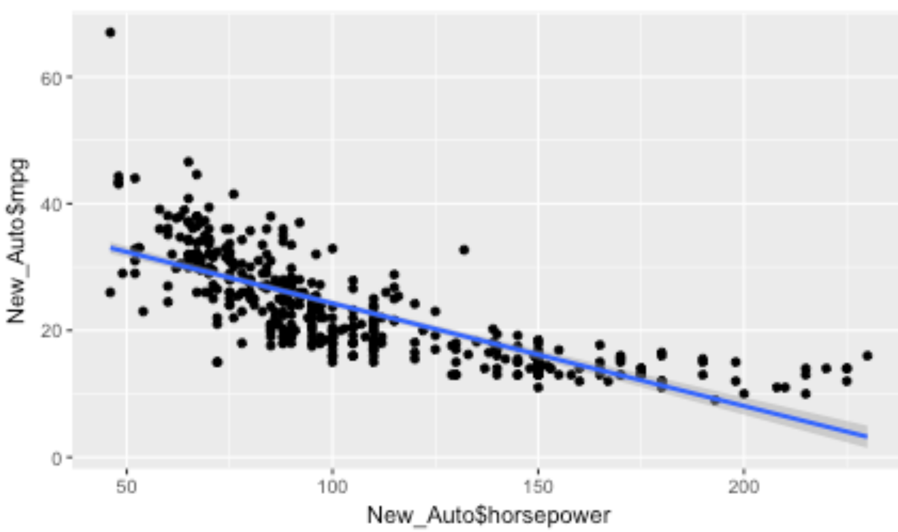
    $\left(\partial^2 f(b_0, b_1)/\partial b_0^2\right) \left(\partial^2 f(b_0, b_1)/\partial b_1^2\right) - \partial^2 f(b_0, b_1)/\partial b_0 \partial b_1$
- Maximum: $SOC1 < 0$, $SOC2 < 0$, $SOC3 > 0$
- Minimum: $SOC1 > 0$, $SOC2 > 0$, $SOC3 > 0$

Back to OLS example

- $f(b_0, b_1) = n^{-1} \sum_{i=1}^{n} \{y_i - b_0 - b_1 x_i\}^2$
  - SOC1: $\partial^2 f(b_0, b_1)/\partial b_0^2 = n^{-1} \sum_{i=1}^{n} 1 = n > 0$
  - SOC2: $\partial^2 f(b_0, b_1)/\partial b_1^2 = 2n^{-1} \sum_{i=1}^{n} x_i^2 > 0$
  - SOC3: $2n^{-1} \sum_{i=1}^{n} x_i^2 - (2n^{-1} \sum_{i=1}^{n} x_i)^2 > 0$
- What about the last SOC...

Non-Diff Example: LAD Regression

- Standard OLS fits a straight line
- What about points far away from the line?
- Automobile dataset



- Influential points can adversely affect the regression line
- $|x - y|$ less sensitive to such points than $(x - y)^2$

Least Absolute Deviation (LAD) Regression

- Build and estimate model

$$mpg_i = \beta_0 + \beta_1 Horsepower_i + \beta_2 Weight_i + \epsilon_i$$

- using $\sum_{i=1}^{n} |mpg_i - b_0 - b_1 horsepower_i - b_2 weight_i|$
- compare with OLS.

```
## [1] 43.904966400 -0.036453947 -0.005762342
```

```
##  (Intercept)   horsepower       weight
## 46.192470033 -0.051143624 -0.005809749
```