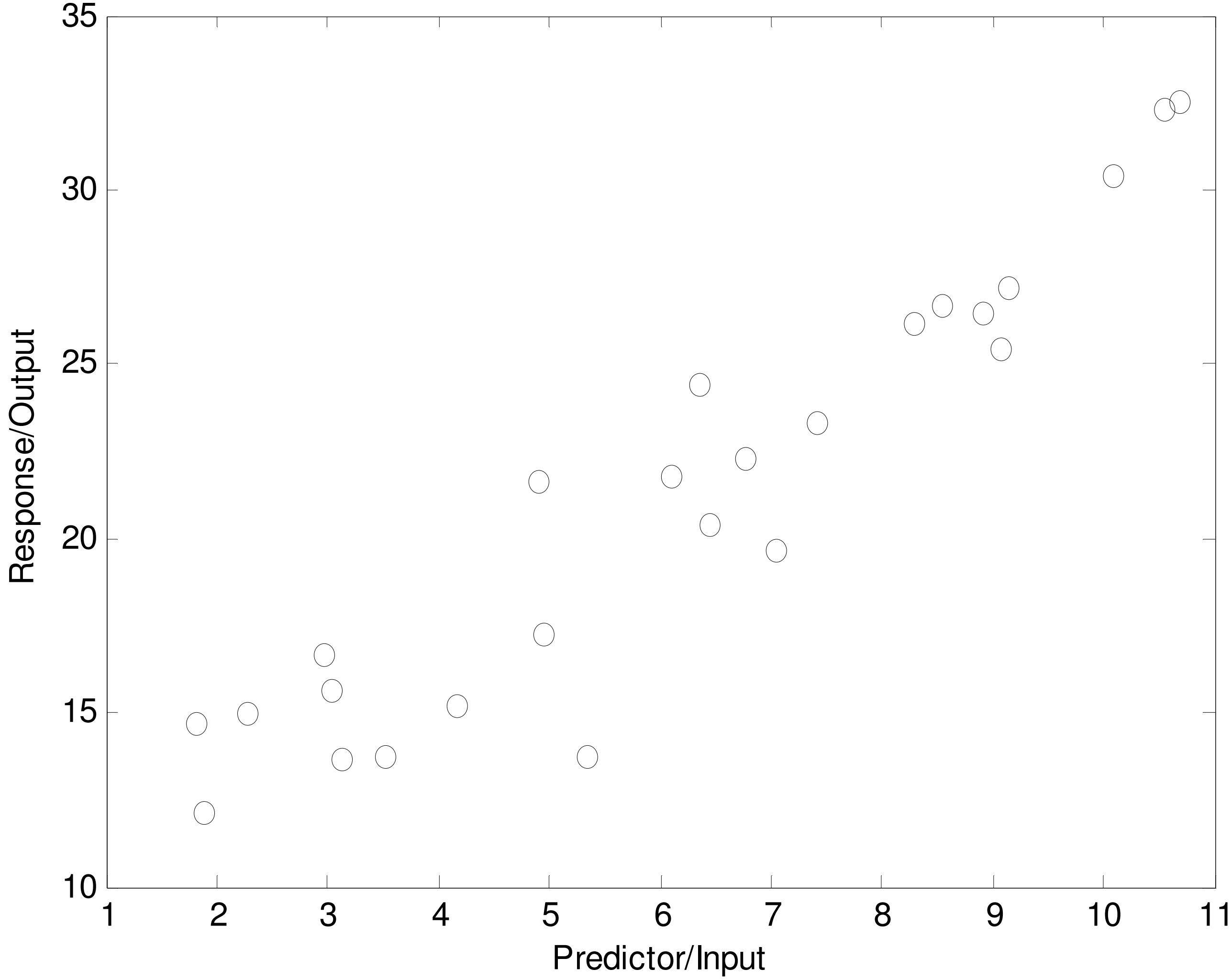
Regression Introduction and Estimation Review

Dr. Frank Wood

Quick Example – Scatter Plot



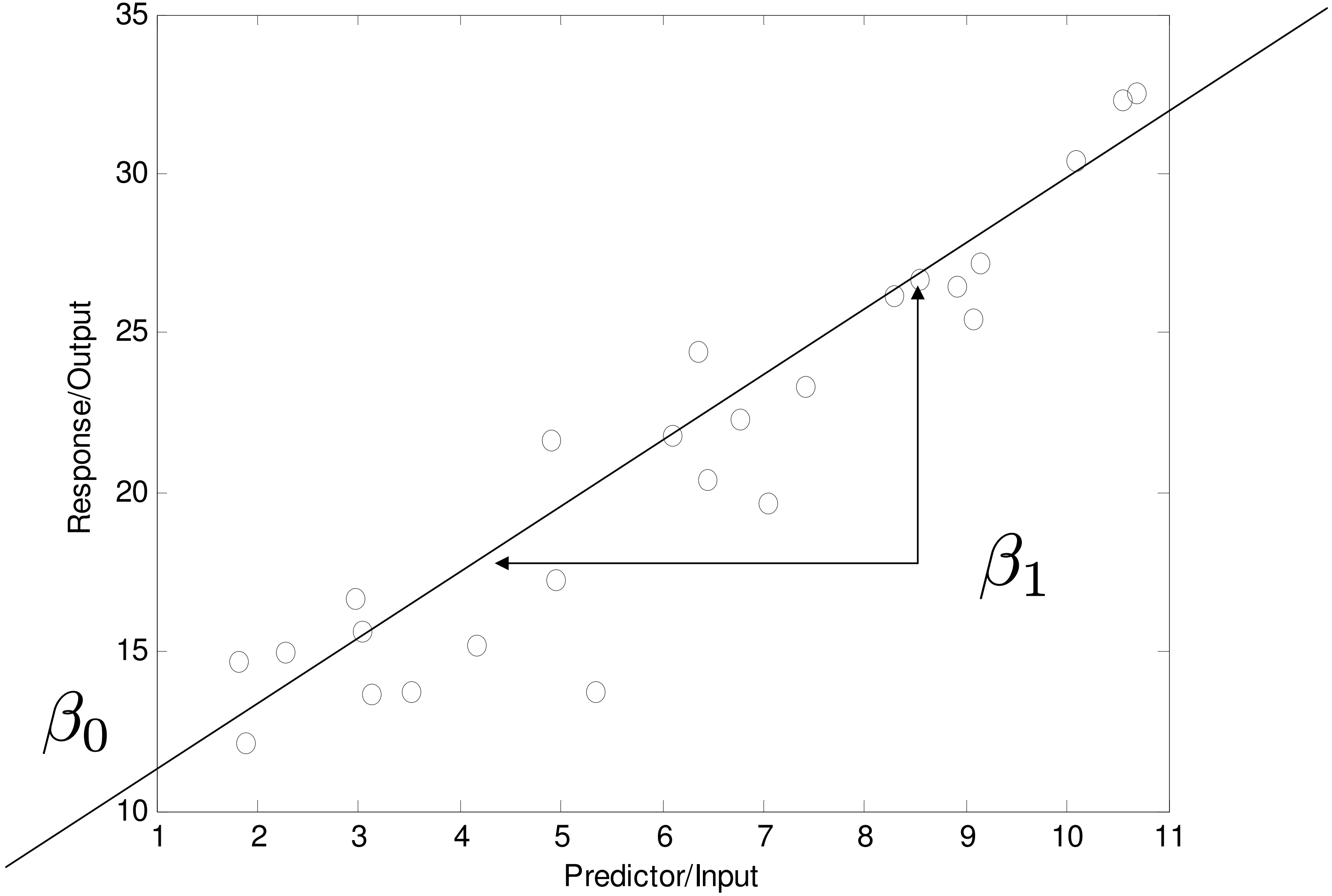
Linear Regression

- Want to find parameters for a function of the form

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

- Distribution of error random variable not specified

Quick Example – Scatter Plot



Formal Statement of Model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

- Y_i value of the response variable in the i^{th} trial
- β_0 and β_1 are parameters
- X_i is a known constant, the value of the predictor variable in the i^{th} trial
- ϵ_i is a random error term with mean $E(\epsilon_i)$ and variance $V(\epsilon_i) = \sigma^2$
- $i = 1,\ldots,n$

Properties

- The response Y_i is the sum of two components
  - 1) Constant term $\beta_0 + \beta_1 X_i$
  - 2) Random term ϵ_i

- The expected response is

$$
\begin{aligned}
E(Y_i) &= E(\beta_0 + \beta_1 X_i + \epsilon_i) \\
&= \beta_0 + \beta_1 X_i + E(\epsilon_i) \\
&= \beta_0 + \beta_1 X_i
\end{aligned}
$$

Expectation Review

- Definition

$$E(X) = E(X) = \int XP(X)dX,\ X \in \mathcal{R}$$

- Linearity property

$$\begin{aligned} E(aX) &= aE(X) \\ E(aX + bY) &= aE(X) + bE(Y) \end{aligned}$$

- Obvious from definition

Example Expectation Derivation

$$P(X) = 2X, 0 \leq X \leq 1$$

draw distro on board

- Expectation

$$E(X) = \int_0^1 XP(X)dX$$

Expectation of a Product of Random Variables

- If X,Y are random variables with joint distribution j(X,Y) then the expectation of the product is given by

$$\mathrm{E}(XY) = \int_{X,Y} XY\, j(X,Y)\, dX\, dY.$$

Expectation of a product of random variables

- What if X and Y are independent?
  - If X and Y are independent with density functions f and g respectively then

$$
\begin{aligned}
\mathrm{E}(XY) &= \int_{X,Y} XY\, f(X) g(Y)\, dX\, dY = \int_X \int_Y XY\, f(X) g(Y)\, dX\, dY \\
&= \int_X X f(X) \left[\int Y g(Y)\, dY \right] dX = \int_X X f(X) \mathrm{E}(Y)\, dX = \mathrm{E}(X)\mathrm{E}(Y)
\end{aligned}
$$

Regression Function

- The response Y_i comes from a probability distribution with mean

$$E(Y_i) = \beta_0 + \beta_1 X_i$$

- This means the regression function is

$$E(Y) = \beta_0 + \beta_1 X$$

Since the regression function relates the means of the probability distributions of Y for a given X to the level of X

Error Terms

- The response Y_i in the i^{th} trial exceeds or falls short of the value of the regression function by the error term amount ϵ_i
- The error terms ϵ_i are assumed to have constant variance σ^2

Response Variance

- Responses Y_i have the same constant variance

$$
\begin{aligned}
V(Y_i) &= V(\beta_0 + \beta_1 X_i + \epsilon_i) \\
&= V(\epsilon_i) \\
&= \sigma^2
\end{aligned}
$$

Variance (2^{nd} central moment) Review

- Continuous distribution

$$V(X) = E((X - E(X))^2) = \int (X - E(X))^2 P(X) dX, \; X \in \mathcal{R}$$

- Discrete distribution

$$V(X) = E((X - E(X))^2) = \sum_i (X_i - E(X))^2 P(X_i), \; X \in \mathcal{Z}$$

Alternative Form for Variance

$$\begin{aligned} V(X) &= E((X-E(X))^2) \\ &= E((X^2 - 2XE(X) + E(X)^2)) \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 \\ &= E(X^2) - 2E(X)^2 + E(X)^2 \\ &= E(X^2) - E(X)^2. \end{aligned}$$

Example Variance Derivation

$$P(X) = 2X, 0 \leq X \leq 1$$

Same as before

$$V(X) \quad = \quad E((X - E(X))^2) = E(X^2) - E(X)^2$$

Variance Properties

$$
\begin{aligned}
V(aX) &= a^2 V(X) \\
V(aX + bY) &= a^2 V(X) + b^2 V(Y) \, \mathrm{if} \, X \perp Y
\end{aligned}
$$

- More generally

$$
V(\sum X_i) = \sum_i \sum_j \mathrm{Cov}(X_i, X_j)
$$

Covariance

- The covariance between two real-valued random variables X and Y, with expected values E(X) = μ and E(Y) = ν is defined as

$$\mathrm{Cov}(X,Y) = \mathrm{E}((X-\mu)(Y-\nu)),$$

- Which can be rewritten as

$$\mathrm{Cov}(X,Y) = \mathrm{E}(X \cdot Y - \mu Y - \nu X + \mu\nu),$$
$$\mathrm{Cov}(X,Y) = \mathrm{E}(X \cdot Y) - \mu \mathrm{E}(Y) - \nu \mathrm{E}(X) + \mu\nu,$$
$$\mathrm{Cov}(X,Y) = \mathrm{E}(X \cdot Y) - \mu\nu.$$

Covariance of Independent Variables

- If X and Y are independent, then their covariance is zero. This follows because under independence

$$E(X \cdot Y) = E(X) \cdot E(Y) = \mu\nu.$$

and then

$$\mathrm{Cov}(X, Y) = \mu\nu - \mu\nu = 0.$$

Least Squares Linear Regression

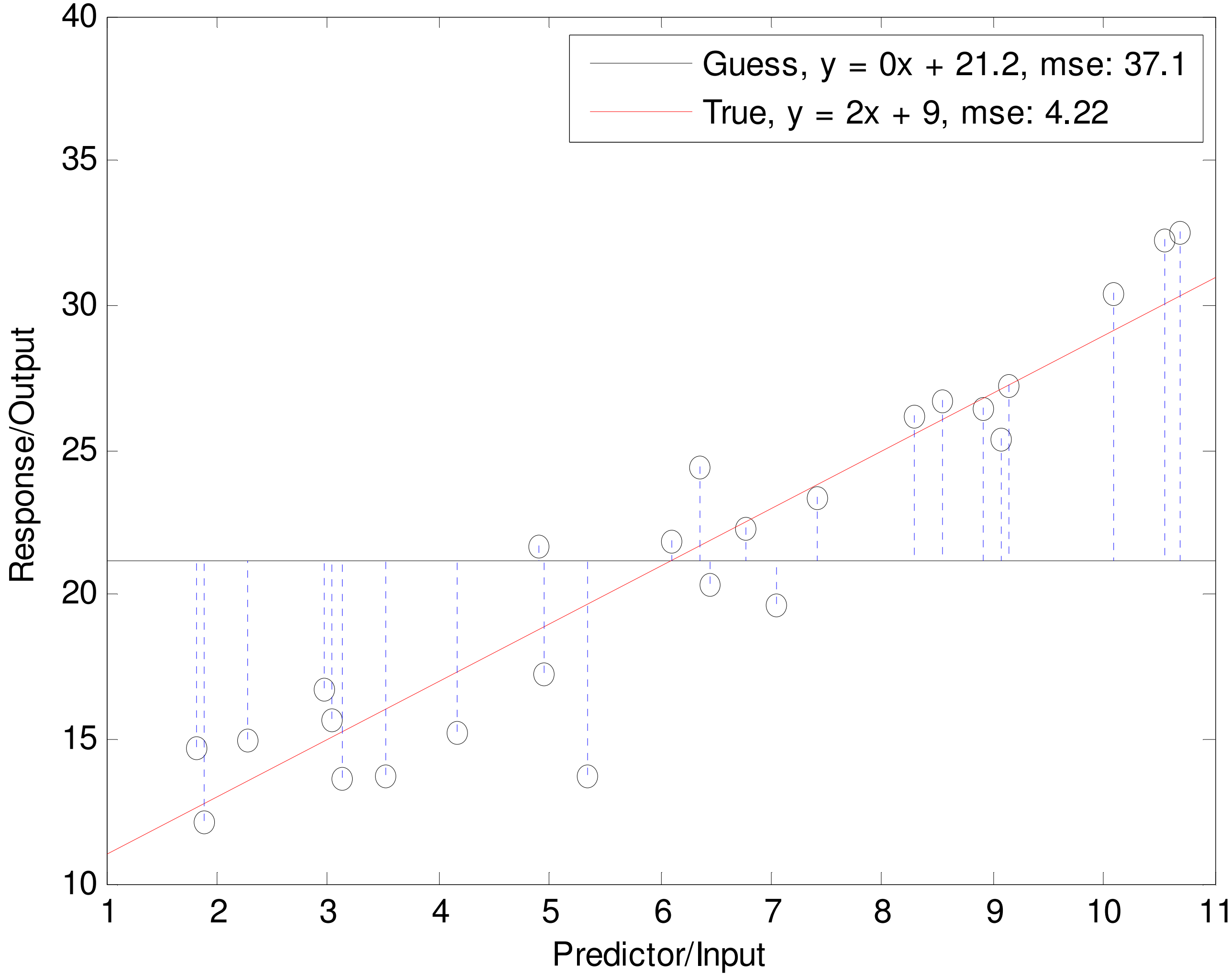
- Seek to minimize

$$Q = \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 X_i))^2$$

- By careful choice of b_0 and b_1 where b_0 is a point estimator for β_0 and b_1 is the same for β_1

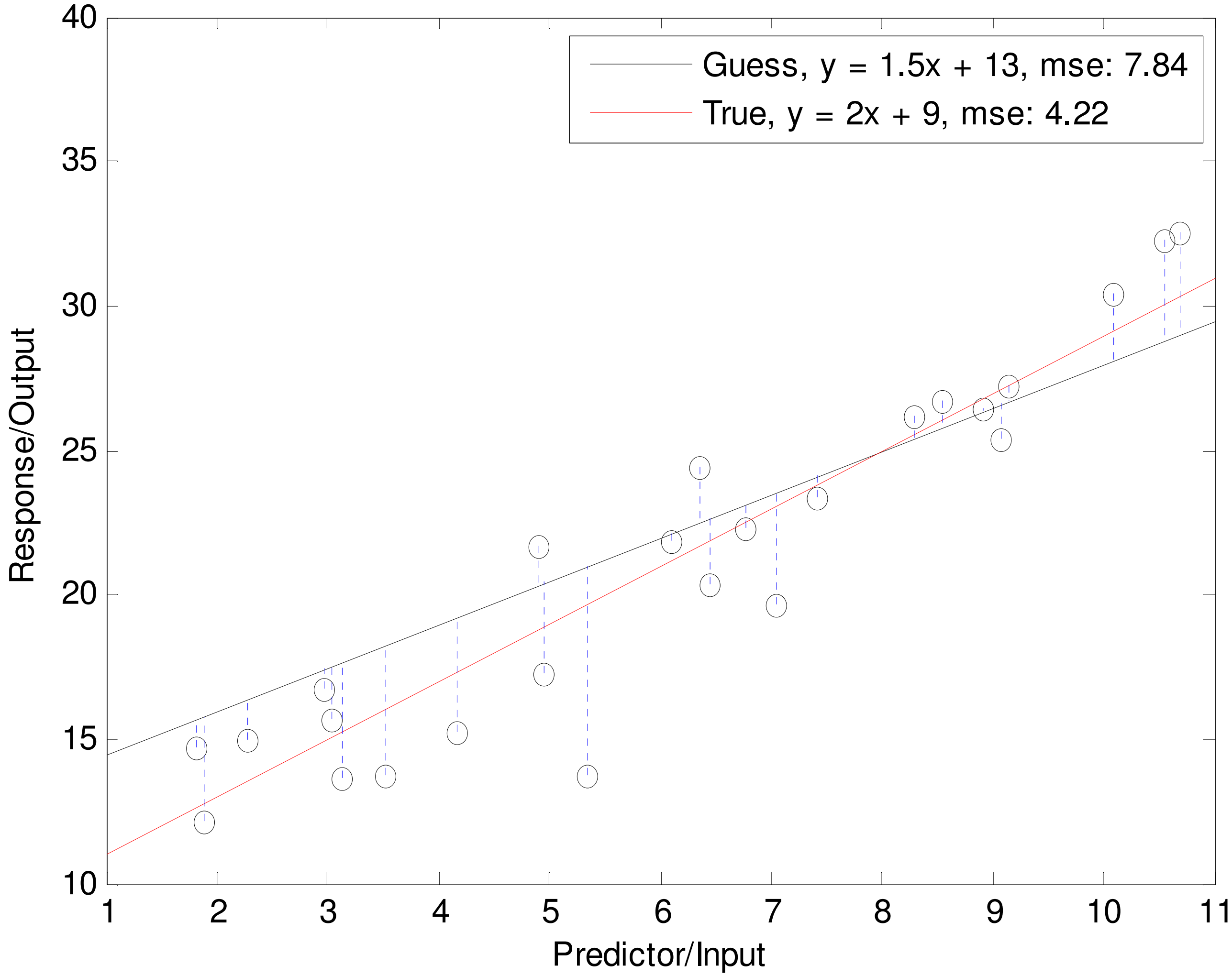
How?

Guess #1



Guess, y = 0x + 21.2, mse: 37.1
True, y = 2x + 9, mse: 4.22
Response/Output
Predictor/Input
40
35
30
25
20
15
10
1
2
3
4
5
6
7
8
9
10
11

Guess #2



Guess, y = 1.5x + 13, mse: 7.84
True, y = 2x + 9, mse: 4.22
40
35
30
25
20
15
10
1
2
3
4
5
6
7
8
9
10
11
Response/Output
Predictor/Input

Function maximization

- Important technique to remember!
  1. Take derivative
  2. Set result equal to zero and solve
  3. Test second derivative at that point

- Question: does this always give you the maximum?

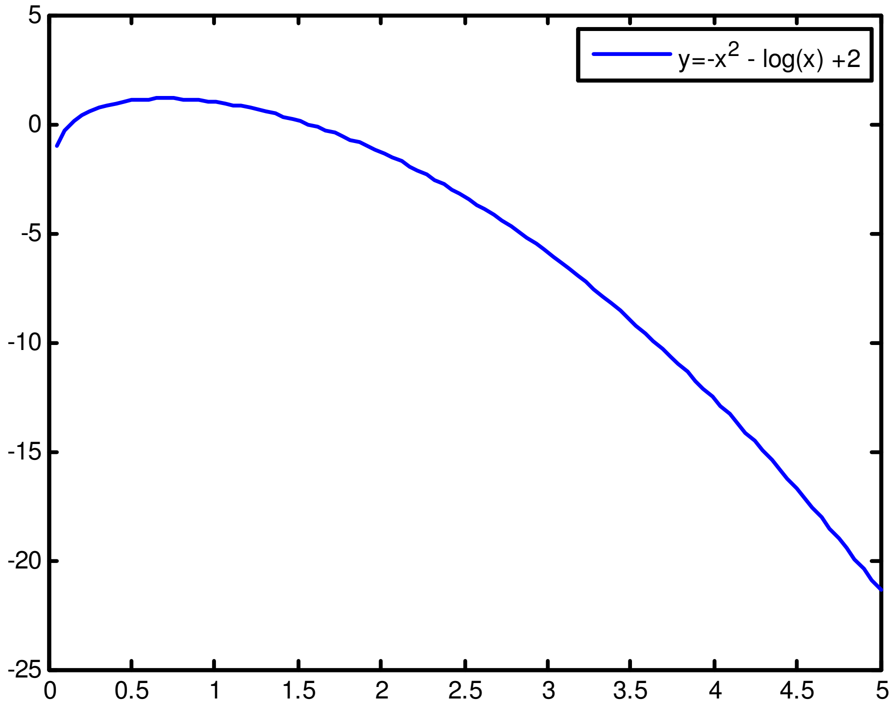
Draw some pictures

- Going further: multiple variables, convex optimization

Function Maximization

- Find the maximum value of x that satisfies the function

$$-x^2 + ln(x) = a, x > 0$$



do derivative
on board

Least Squares Max(min)imization

- Function to minimize w.r.t. β_0, β_1

$$Q = \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 X_i))^2$$

- Minimize this by maximizing –Q
- Find partials and set both equal to zero

$$\begin{aligned} \frac{dQ}{d\beta_0} &= 0 \\ \frac{dQ}{d\beta_1} &= 0 \end{aligned}$$

go to board

Normal Equations

- The result of this maximization step are called the normal equations. b_0 and b_1 are called point estimators of β_0 and β_1 respectively

$$\begin{aligned}
\sum Y_i &= nb_0 + b_1 \sum X_i \\
\sum X_i Y_i &= b_0 \sum X_i + b_1 \sum X_i^2
\end{aligned}$$

- This is a system of two equations and two unknowns. The solution is given by…

Solution to Normal Equations

- After a lot of algebra one arrives at

$$
\begin{aligned}
b_1 &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \\
b_0 &= \bar{Y} - b_1 \bar{X} \\
\bar{X} &= \frac{\sum X_i}{n} \\
\bar{Y} &= \frac{\sum Y_i}{n}
\end{aligned}
$$