# Mathematic Analysis with Matlab

Lecture 11: Linear and Non-linear Regression

Lecturer: *Dr*. Lian-Sheng Wang

*Fall Semester* 2022

Contact: lswang@xmu.edu.cn

# Outline

# Linear Regression: the problem

- Given variable x (house price) and y (rent),
- We have a series of observations $(x^1, y^1), ..., (x^i, y^i)$
- Assuming they are linearly related: y=$\theta$x
- We want to estimate $\theta$ based on these observations

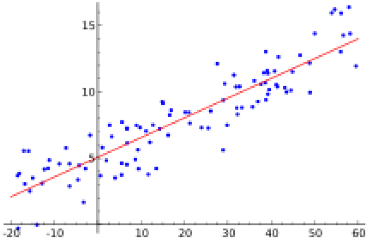



Figure: Observations (crosses in blue) and their linear regression (red line).

# Linear Regression: modeling the problem (1)

- Assumption (1): $(x^1, y^1), ..., (x^i, y^i), ..., (x^n, y^n)$ are indepent and from identical distribution (IID)
- $y^i$ and $x^i$ are related by following equation:
- We extend to multiple-variable case:
- That is: $x^i = [x_1^i, x_2^i, ...x_m^i]$

$$y^i = \theta^T x^i + \varepsilon^i, \tag{1}$$

where $\varepsilon^i$is the error

- Assumption (2): $\varepsilon^i$ follows N(0,$\sigma^2$)
- We have:

$$p(\varepsilon^i) = \frac{1}{\sqrt{2\pi\sigma}} exp(-\frac{{\varepsilon^i}^2}{2\sigma^2}) \tag{2}$$

# Linear Regression: modeling the problem (2)

$$y^i = \theta^T x^i + \varepsilon^i,$$

where $\varepsilon^i$ is the error

- Assumption (2): $\varepsilon^i$ follows N(0,$\sigma^2$)
- We have:

$$p(\varepsilon^i) = \frac{1}{\sqrt{2\pi\sigma}} exp(-\frac{{\varepsilon^i}^2}{2\sigma^2})$$

- This implies:

$$p(y^i|x^i;\theta) = \frac{1}{\sqrt{2\pi\sigma}} exp(-\frac{(y^i - \theta^T x^i)^2}{2\sigma^2}) \quad (3)$$

# Linear Regression: modeling the problem (3)

- Remember "Assumption (1)": all $(x^i, y^i)$s are IID;
- We want to maximize the probability:

$$L(\theta) = \prod_1^n p(y^i|x^i;\theta)$$
$$= \prod_1^n \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{(y^i - \theta^T x^i)^2}{2\sigma^2})$$

- Take $\log(\cdot)$ on both sides:

$$log(L(\theta)) = log \prod_1^n \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{(y^i - \theta^T x^i)^2}{2\sigma^2})$$
$$= \sum_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} - \sum_{i=1}^n \frac{(y^i - \theta^T x^i)^2}{2\sigma^2}$$

# Linear Regression: modeling the problem (4)

- We have:

$$log(L(\theta)) = \sum_{1=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} - \sum_{i=1}^{n} \frac{(y^i - \theta^T x^i)^2}{2\sigma^2} \tag{4}$$

- Maximizing $log(L(\theta))$ is equavlent to minimizing:

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{n} (y^i - \theta^T x^i)^2 \tag{5}$$

- This is a least-square problem
- We are searching for $\theta$ that minimizes $J(\theta)$

# Linear Regression: solve the problem (1)

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{n}(y^i - \theta^T x^i)^2 \tag{6}$$

- We define H:

$$\begin{aligned} H = X\theta - \overrightarrow{y} &= \begin{bmatrix} (x^1)^T\theta \\ . \\ . \\ . \\ (x^n)^T\theta \end{bmatrix} - \begin{bmatrix} (y^1) \\ . \\ . \\ . \\ (y^n) \end{bmatrix} \\ &= \begin{bmatrix} (x^1)^T\theta - (y^1) \\ . \\ . \\ . \\ (x^1)^T\theta - (y^n) \end{bmatrix} \end{aligned} \tag{7}$$

# Linear Regression: solve the problem (2)

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{n}(y^i - \theta^T x^i)^2$$

- Given H:

$$H = X\theta - \overrightarrow{y} = \begin{bmatrix} (x^1)^T\theta - (y^1) \\ . \\ . \\ . \\ (x^1)^T\theta - (y^n) \end{bmatrix}$$

- J is written as:

$$J(\theta) = \frac{1}{2}(X\theta - \overrightarrow{y})^T(X\theta - \overrightarrow{y}) \tag{8}$$

# Linear Regression: solve the problem (2)

- To minimize $J(\theta)$, we search for its derivatives with respect to $\theta$

$$\begin{aligned}\nabla_\theta J(\theta) &= \nabla_\theta \frac{1}{2}(X\theta - \overrightarrow{y})^T(X\theta - \overrightarrow{y}) \\ &= \frac{1}{2}\nabla_\theta(\theta^T X^T X\theta - \theta^T X^T \overrightarrow{y} - \overrightarrow{y}^T X\theta + \overrightarrow{y}^T \overrightarrow{y}) \\ &= \frac{1}{2}(X^T X\theta + X^T X\theta - 2X^T \overrightarrow{y}) \\ &= X^T X\theta - X^T \overrightarrow{y}\end{aligned} \tag{9}$$

- Set $\nabla_\theta J(\theta)$ to $\mathbf{0}$, we have

$$\begin{aligned}X^T X\theta - X^T \overrightarrow{y} &= 0 \\ \Rightarrow X^T X\theta &= X^T \overrightarrow{y} \\ \Rightarrow \theta &= (X^T X)^{-1} X^T \overrightarrow{y}\end{aligned} \tag{10}$$

# Linear Regression: closed form solution

- Given $X = \{x_1, x_2, \cdots, x_n\}$ and
- $y = \theta_1 x + \theta_0$
- Try to solve $\theta$ with closed form solution we obtained

$$\Rightarrow \theta = (X^T X)^{-1} X^T \overrightarrow{y} \tag{11}$$

# Linear Regression: numerical solution (1)

- Recall the least square problem:

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x^i) - y^i)^2 \tag{12}$$

- Gradient descent solve $\theta$ by performing update:

$$\theta_j = \theta_j - \alpha\frac{\partial J(\theta)}{\partial \theta_j}, \tag{13}$$

where $\alpha$ is a constant.

# Linear Regression: numerical solution (2)

- Try to solve $\frac{\partial J(\theta)}{\partial \theta_j}$

$$
\begin{aligned}
\frac{\partial J(\theta)}{\partial \theta_j} &= \frac{\partial}{\partial \theta_j} \frac{1}{2} (h_\theta(x) - y)^2 \\
&= 2 \cdot \frac{1}{2} (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j} (h_\theta(x) - y) \\
&= (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j} (\sum_{i=0}^{n} \theta_i x_i - y) \\
&= (h_\theta(x) - y) \cdot x_j
\end{aligned} \tag{14}
$$

# Linear Regression: numerical solution (3)

- Now we have:

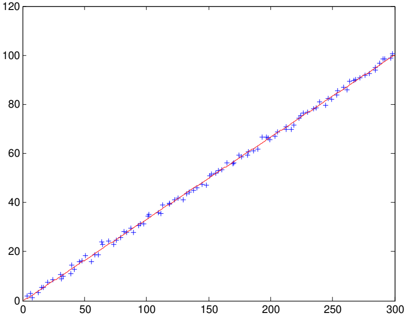
$$\theta_j := \theta_j - \alpha(h_\theta(x^i) - y^i) \cdot x_j^i \tag{15}$$

- Given $X = \{x_{i=1}^m\}$, the gradient descent procedure is:
  1. Repeat
  2. For j=1 $\cdots$ D
  3. $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_\theta(x^i) - y^i) \cdot x_j^i$
  4. End-for
  5. untile convergence

- Course project: implement above procedure in C and call it in Matlab:
  1. axi=[1:1:100]
  2. X=rand(1,100)*2+2*axi;
  3. Y=rand(1,100)*2+axi;
  4. plot(X,Y,'+');
  5. theta=linRegress(X, Y)

# Linear Regression: Matlab solution

- Given x and y:
  1. axi=[1:1:100]
  2. X=rand(1,100)*2+2*axi;
  3. Y=rand(1,100)*2+axi;
  4. n=length(Y)
  5. x=[ones(n,1),X']
  6. [b, bint, r, rint, s]=regress(Y',x)
  7. syms u v;
  8. v=[1:1:300];
  9. u=b(2)*v+b(1);
  10. plot(X,Y,'+',v,u,'r-');



# Outline

# Non-linear Regression (1)

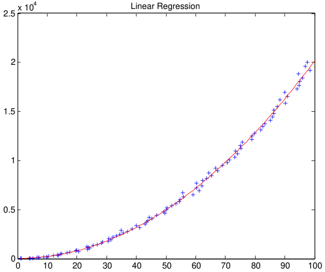
- Given x and y
- Fit x and y with following polynomial:

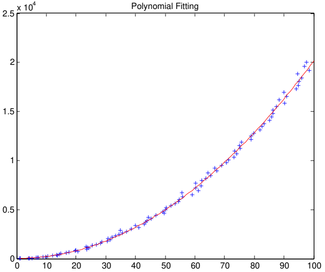
$$y = b_0 + b_1 x + b_2 x^2 + \varepsilon \tag{16}$$

- Solution 1: via linear regression
  1. axi=[1:1:100]
  2. X=randn(1,100)+axi;
  3. Y=randn(1,100)+2*axi.^2;
  4. x1=X; x2=X.^2;
  5. n=length(X);
  6. x=[ones(n,1),x1', x2'];
  7. [b, bint, r, rint, s]=regress(Y',x);

- Solution 2: via polyfit
  1. axi=[1:1:100]
  2. X=randn(1,100)+axi;
  3. Y=randn(1,100)+2*axi.^2;
  4. a=polyfit(X, Y, 2)

# Non-linear Regression (1) the results




- *Comments*: two different solutions generate the same outputs

# Exercise (1)

- The following table gives (in U.S. dollars) the estimated public expenditures on education, per inhabitant, in the indicated years in Europe and North America.

| Year | Europe | US |
|---|---|---|
| 1970 | 90 | 317 |
| 1975 | 197 | 474 |
| 1980 | 335 | 816 |
| 1985 | 394 | 1101 |

- Given the expenditure and time follow a linear equation f(x)=cx+b
- Please work out c and b that best fit to data from Europe and US
- Hints:
  - $c \cdot 0 + b$ should be close to 90
  - $c \cdot 5 + b$ should be close to 197
  - ...

# Exercise (2)

| Year | Europe | US |
|---|---|---|
| 1970 | 90 | 317 |
| 1975 | 197 | 474 |
| 1980 | 335 | 816 |
| 1985 | 394 | 1101 |

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 5 \\ 1 & 10 \\ 1 & 15 \end{bmatrix} \qquad f_1 = \begin{bmatrix} 90 \\ 197 \\ 335 \\ 394 \end{bmatrix},$$

```
A = [1 0;1 5;1 10;1 15]
f = [90 197 335 394]';
r = inv(A'*A)*A'*f;
```

Answers: [96.5, 21.0]

# Exercise (3)

| Year | Europe | US |
|---|---|---|
| 1970 | 90 | 317 |
| 1975 | 197 | 474 |
| 1980 | 335 | 816 |
| 1985 | 394 | 1101 |

- Think about fit the data with $f(x) = A + Bx + Cx^2$
- Work out A, B and C

# Exercise (4)

| Year | Europe | US |
|---|---|---|
| 1970 | 90 | 317 |
| 1975 | 197 | 474 |
| 1980 | 335 | 816 |
| 1985 | 394 | 1101 |

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 5 & 25 \\ 1 & 10 & 100 \\ 1 & 15 & 225 \end{bmatrix} \qquad f_1 = \begin{bmatrix} 90 \\ 197 \\ 335 \\ 394 \end{bmatrix},$$

```
A = [1 0 0;1 5 25;1 10 100;1 15 225]
f = [90 197 335 394]';
r = inv(A'*A)*A'*f;
```

Answers: [84.5, 28.2, -0.48]

# Non-linear Regression (2): linear or non-linear

- Depends on $\theta$ instead of $x$
- Given following form
- $f_\theta(x) = \theta_0 + \theta_1 * g_1(x) + \theta_2 * g_2(x) + \cdots$
- It is linear
- Given following form
- $f_\theta(x) = g_1(x) * \theta^1 + g_2(x) * \theta^2 + \cdots$
- It is non-linear
- Now make your judgement

$$f_\theta(x) = \theta_1 sin(x) + \theta_2 cos(x) \quad f_\theta(x) = \frac{\theta_1}{\theta_2} sin(x) + {\theta_2}^2 cos(x)$$

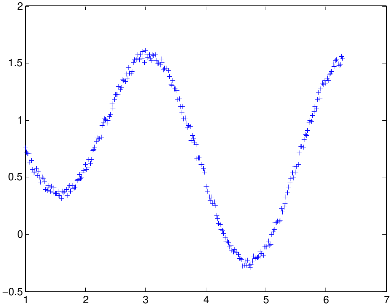
$$f_\theta(x) = \theta_1 x^2 + {\theta_2}^2 x \qquad f_\theta(x) = \theta_1 x^2 + 5\theta_2 x$$

# Non-linear Regression (2-1): linear or non-linear

- Given following function

- $f_\theta(x) = t_1 * sin(x) + t_2 * cos(x)^2$
- x is in the range of $[1, 2\pi]$
- Estimate '$t_1$' and '$t_2$' by '???'

1. A=load('reg_data2.txt');
2. x=A(:,1);y=A(:,2);
3. plot(x,y); hold on;
4. ...



# Non-linear Regression (2-2): linear or non-linear

- Given following function
- $f(x) = t_1 * sin(x) + t_2 * cos(x)^2$
- x is in the range of $[1,2\pi]$
- Estimate ‘$t_1$’ and ‘$t_2$’ by ‘regress’

```
A = load(‘reg_data2.txt’);
x = A(:,1); y=A(:,2);
plot(x,y,’b+’); hold on;
X = [sin(x),(cos(x)^2)];
[b int] = regress(y,X);
u = b(1)*sin(x)+b(2)*cos(x)^2;
plot(x, u, ’r-’);
```

# Non-linear Regression (3)

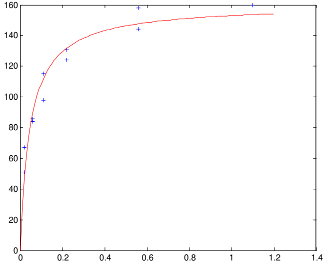
- Given x=[0.02 0.02 0.06 0.06 0.11 0.11 0.22 0.22 0.56 0.56 1.10]
- Given y=[67 51 84 86 98 115 131 124 144 158 160]
- Fit x and y with function:

$$y = \frac{\beta_1 x}{\beta_2 + x} \tag{17}$$

- Solution 1: via linear regression
  1. $\frac{1}{y} = \frac{1}{\beta_1} + \frac{\beta_2}{\beta_1}\frac{1}{x}$
  2. $\frac{1}{y} = \theta_1 + \theta_2 \frac{1}{x}$
  3. call 'regress'

- Solution 2: via polyfit
  1. b0=[143 0.03]
  2. fun=inline('b(1)*x./(b(2)+x)','b','x');
  3. [b,R,J]=nlinfit(x,y,fun,b0);
  4. xx=0:0.01:1.2;
  5. yy=b(1)*xx./(b(2)+xx);
  6. plot(x, y, 'b+', xx, yy 'r-');

# Non-linear Regression (2) the results



# Food for thought: how about no model is predefined??

- How about a series of X, Y are given
- However we have no idea about the target function

# Q & A

# Thanks for your attention!