Oscar Scholin
179 HW 1

1. Let $\vec{y} = A\vec{x} + \vec{b}$ be random vector.

a) Let's compute the expected value $\mathbb{E}(\vec{y})$ over the arbitrary interval Ω:

$$\mathbb{E}(\vec{y}) = \int_\Omega (A\vec{x} + \vec{b})\, \mathbb{P}(\vec{x})\, dx$$

where $\mathbb{P}(x)$ is the probability density function.

since constant wrt $\vec{x}$

$$= A \underbrace{\int_\Omega \vec{x}\, P(\vec{x})\, dx}_{\mathbb{E}(\vec{x})} + \vec{b} \underbrace{\int_\Omega \mathbb{P}(\vec{x})\, dx}_{\equiv 1}$$

$$= A\,\mathbb{E}(\vec{x}) + \vec{b} \quad \square$$

b) WTS $\mathrm{cov}[\vec{y}] = A \underbrace{\mathrm{cov}[\vec{x}]}_{\equiv \Sigma} A^T$

Note $\mathrm{cov}[\vec{x}]$ is a square matrix which quantifies covariance b/n pairs of elements in $\vec{x}$.

$\therefore\ \mathrm{cov}[\vec{x}] \equiv \Sigma = \mathbb{E}\left[(\vec{x} - \mathbb{E}(\vec{x}))(\vec{x} - \mathbb{E}(\vec{x}))^T\right]$.

$\therefore$ Substituting in,

$$\mathrm{cov}[\vec{y}] = \mathbb{E}\left[\left((A\vec{x} + \vec{b}) - \mathbb{E}(A\vec{x} + \vec{b})\right)\left\{(A\vec{x} + \vec{b}) - \mathbb{E}(A\vec{x} + \vec{b})\right\}^T\right]$$

$$= \mathbb{E}\left[\left\{A\vec{x} + \vec{b} - (A\,\mathbb{E}(\vec{x}) + \vec{b})\right\}\left\{A\vec{x} + \vec{b} - (A\,\mathbb{E}(\vec{x}) + \vec{b})\right\}^T\right]$$

by part a

$$= \mathbb{E}\left[\left\{A(\vec{x} - \mathbb{E}(\vec{x}))\right\}\left\{A(\vec{x} - \mathbb{E}(\vec{x}))\right\}^T\right]$$

note $(AB)^T = B^T A^T$

$$= A\,\mathbb{E}\left[(\vec{x} - \mathbb{E}(\vec{x}))(\vec{x} - \mathbb{E}(\vec{x}))^T\right] A^T$$

$$= A\,\mathrm{cov}[\vec{x}]\, A^T \quad \square$$

2. We have the data $D = \{(0,1), (2,3), (3,6), (4,8)\}$.

a) Find $\vec{y} = \theta^T \vec{x}$ using Cramer's rule.

We are using a simple linear regression model:

$y = \theta_0 + \theta_1 x$

$\therefore$ The optimization problem is:

$$A\vec{\theta} = \vec{b}$$

the "target" y values

for $\vec{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$, $A = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix}$

col of 1s to reflect θ_0 term

col of x values

The normal equations are

$$\underbrace{A^T A}_{A'}\vec{\theta} = \underbrace{A^T\vec{b}}_{b'} \quad \text{for use in Cramer's rule}$$

$$A^T A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix}$$

$4 + 9 + 16 = 29$

$$A^T\vec{b} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 18 \\ 56 \end{bmatrix}$$

$6 + 18 + 32 = 56$

Cramer's rule says if A' non singular then solution components are:

$$x_i = \frac{\det(A'_i)}{\det(A')}$$

A_i denotes replacing i^{th} col of A' w/ $\vec{b}'$ matrix

$\vec{b}'$

$$\therefore \theta_0 = \frac{\begin{vmatrix} 18 & 9 \\ 56 & 29 \end{vmatrix}}{\begin{vmatrix} 4 & 9 \\ 9 & 29 \end{vmatrix}} = \frac{18\cdot 29 - 56\cdot 9}{29\cdot 4 - 9\cdot 9} = \frac{18}{35}$$

$$\theta_1 = \frac{\begin{vmatrix} 4 & 18 \\ 9 & 56 \end{vmatrix}}{\begin{vmatrix} 4 & 9 \\ 9 & 29 \end{vmatrix}} = \frac{4\cdot 56 - 18\cdot 9}{35}$$

$$= \frac{62}{35}$$

b) Now we use the actual solution to the normal equations, given by

$$\vec{\theta} = (A^T A)^{-1} A^T \vec{b}$$

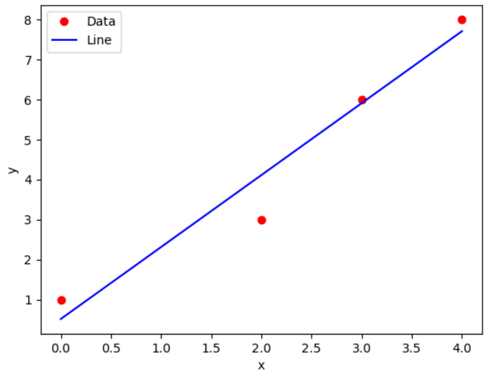
$$\begin{bmatrix} 4 & 9 \\ 9 & 29 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix}$$

$$= \frac{1}{4 \cdot 29 - 9 \cdot 9} \begin{bmatrix} 29 & -9 \\ -9 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix}$$

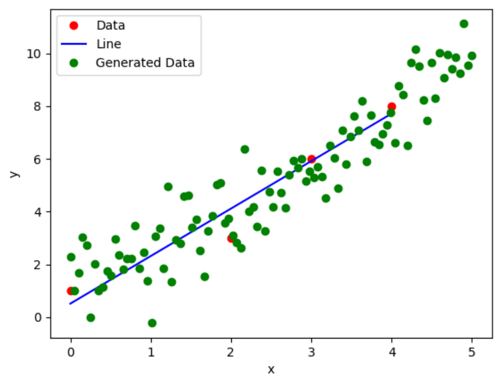
$$= \frac{1}{35} \begin{bmatrix} 29 & 11 & 2 & -7 \\ -9 & -1 & 3 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \\ 8 \end{bmatrix}$$

$$= \frac{1}{35} \begin{bmatrix} 18 \\ 62 \end{bmatrix} = \begin{bmatrix} \frac{18}{35} \\ \frac{62}{35} \end{bmatrix}$$, Same as part a. $\square$

c) Plot data w/ fit ✓



d)



Sampled 100 points from $x = 0 \rightarrow 5$, computed the exact values using $A\vec{\theta} + \vec{b}$, then added random noise term drawn from Gaussian w/ $\mu = 0$, $\sigma = 1$ for each point.