به نام خدا
دکتر مجتبی تفاق ـ بهینه‌سازی در علوم داده

امیرحسین جوادی (۹۷۱۰۱۴۸۹)

دانشگاه صنعتی شریف
دانشکده مهندسی برق

۱ Q16.7 Minimum cost trading to achieve target sector exposures

مسئله را به صورت زیر مدل‌سازی می‌کنیم.

$$\text{Minimize:} \quad \sum_{i=1}^{n} k_i (h_i - h_i^{curr})^2 \tag{۱}$$

$$\text{Subject to:} \quad Sh = s^{des}, 1^T h = 1^T h^{curr} \tag{۲}$$

برای تبدیل این مسئله به یک مسئله مشابه درس در نظر می‌گیریم:

$$\text{Minimize:} \quad f(h) = ||Ah - b||^2 \tag{۳}$$

$$\text{Subject to:} \quad Sh = s^{des}, 1^T h = 1^T h^{curr} \tag{۴}$$

که $A_{n\times n} = diag(\sqrt{k_1}, \sqrt{k_2}, \dots, \sqrt{k_n})$ و $b = Ah^{curr}$ است. حال تابع لاگرانژ را می‌نویسیم.

$$L(h, z) = f(h) + z_1(s_1^T h - s_1^{des}) + z_2(s_2^T h - s_2^{des}) + \cdots + z_m(s_m^T h - s_n^{des}) + z_{m+1}(1^T h - 1^T h^{curr}) \tag{۵}$$

که در آن s_i^T سطر iام ماتریس $S_{m\times n}$ است. با گرفتن مشتق‌های مورد نیاز از $L(h, z)$ به معادلات زیر میرسیم.

$$\frac{\partial L(h, z)}{\partial z_1} = s_1^T h - s_1^{des} = 0 \tag{۶}$$

$$\frac{\partial L(h, z)}{\partial z_2} = s_2^T h - s_2^{des} = 0 \tag{۷}$$

$$\vdots \tag{۸}$$

$$\frac{\partial L(h, z)}{\partial z_m} = s_m^T h - s_m^{des} = 0 \tag{۹}$$

$$\frac{\partial L(h, z)}{\partial z_{n+1}} = 1^T h - 1^T h^{curr} = 0 \tag{۱۰}$$

$$\frac{\partial L(h, z)}{\partial h} = 2(A^T A)h - 2A^T b + z_1 s_1 + z_2 s_2 + \cdots + z_m s_m + z_{m+1} 1_{n\times 1} = 0 \tag{۱۱}$$

با تعریف ماتریس $C_{(m+1)\times n} = [S_{m\times n}; 1_{1\times n}]$ و آرایه‌ی $z_{(m+1)\times 1} = [z_1; z_2; \dots; z_{m+1}]$ و $d_{(m+1)\times 1} = [s^{des}; 1^T h^{curr}]$ داریم:

$$\begin{bmatrix} 2A^T A & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} h \\ z \end{bmatrix} = \begin{bmatrix} 2A^T b \\ d \end{bmatrix} \tag{۱۲}$$

با حل مسئله‌ی KKT زیر به جواب می‌رسیم.

۲ 16.11 Least distance problem

برای حل این مسئله، از روش constrained least squares با پارامترهای زیر استفاده می‌کنیم.

$$A = I, \qquad b = a \tag{۱۳}$$

پس حل این مسئله به روش KKT قابل انجام است.

$$\begin{bmatrix} 2A^TA & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} 2A^Tb \\ d \end{bmatrix} \tag{۱۴}$$

$$\begin{bmatrix} 2I & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} 2a \\ d \end{bmatrix} \tag{۱۵}$$

$$2x + C^Tz = 2a \Rightarrow x = a - \frac{C^Tz}{2} \tag{۱۶}$$

$$C(a - \frac{C^Tz}{2}) = d \Rightarrow CC^Tz = -2(d - Ca) \Rightarrow z = -2(CC^T)^{-1}(d - Ca) \tag{۱۷}$$

$$x = a + C^T(CC^T)^{-1}(d - Ca) = a - C^\dagger(Ca - d) \tag{۱۸}$$

۳ 18.6 Fitting a simple neural network model

به دنبال حل مسئله زیر و پیدا کردن مینیمم مقدار برای E هستیم.

$$E(\theta) = \sum_{i=1}^{N} (\hat{f}(x^{(i)};\theta) - y^{(i)})^2 = (\hat{f}(x;\theta) - y)^T(\hat{f}(x;\theta) - y) \tag{۱۹}$$

$$\frac{\partial E(\theta)}{\partial \theta} = 2(\hat{f}(x;\theta) - y)^T \Big[\frac{\partial}{\partial \theta}(\hat{f}(x;\theta) - y)\Big] = 2(\hat{f}(x;\theta) - y)^T \frac{\partial \hat{f}(x;\theta)}{\partial \theta} = 2(\hat{f}(x;\theta) - y)^T \nabla_\theta \hat{f}(x;\theta) \tag{۲۰}$$

۱. با تعریف ϕ و $\phi^{'}$ به شکل زیر داریم:

$$\phi(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \tag{۲۱}$$

$$\phi(x) = \frac{4}{(\exp(x) + \exp(-x))^2} \tag{۲۲}$$

$$\nabla_\theta \hat{f}(x;\theta) = \begin{cases} \phi(\theta_2 x_1 + \theta_3 x_2 + \theta_4) \\ \theta_1 x_1 \phi^{'}(\theta_2 x_1 + \theta_3 x_2 + \theta_4) \\ \theta_1 x_2 \phi^{'}(\theta_2 x_1 + \theta_3 x_2 + \theta_4) \\ \theta_1 \phi^{'}(\theta_2 x_1 + \theta_3 x_2 + \theta_4) \\ \phi(\theta_6 x_1 + \theta_7 x_2 + \theta_8) \\ \theta_5 x_1 \phi^{'}(\theta_6 x_1 + \theta_7 x_2 + \theta_8) \\ \theta_5 x_2 \phi^{'}(\theta_6 x_1 + \theta_7 x_2 + \theta_8) \\ \theta_5 \phi^{'}(\theta_6 x_1 + \theta_7 x_2 + \theta_8) \\ \phi(\theta_{10} x_1 + \theta_{11} x_2 + \theta_{12}) \\ \theta_9 x_1 \phi^{'}(\theta_{10} x_1 + \theta_{11} x_2 + \theta_{12}) \\ \theta_9 x_2 \phi^{'}(\theta_{10} x_1 + \theta_{11} x_2 + \theta_{12}) \\ \theta_9 \phi^{'}(\theta_{10} x_1 + \theta_{11} x_2 + \theta_{12}) \\ 1 \end{cases} \tag{۲۳}$$

۲. سطر i ام ماتریس $Dr(\theta)$ برابر است با $[\nabla_\theta \hat{f}(x^{(i)};\theta)]^T$

۳. برای حل این مسئله 200 جفت داده‌ی x به صورت رندوم یکنواخت در بازه‌ی $[-1,1]$ تعیین کردم و y مربوط به هر x را برابر ضرب درایه‌های آن در نظر گرفتم. مقدار λ را برابر با $1e-5$ قرار دادم و به اندازه‌ی $K_{max} = 150$ پارامترهای مسئله را بهبود بخشیدم.

$$r(\theta) = f(x,\theta) - y \approx f(x,\theta^{(k)}) + Df(\theta)(\theta - \theta^{(k)}) - y \tag{۲۴}$$

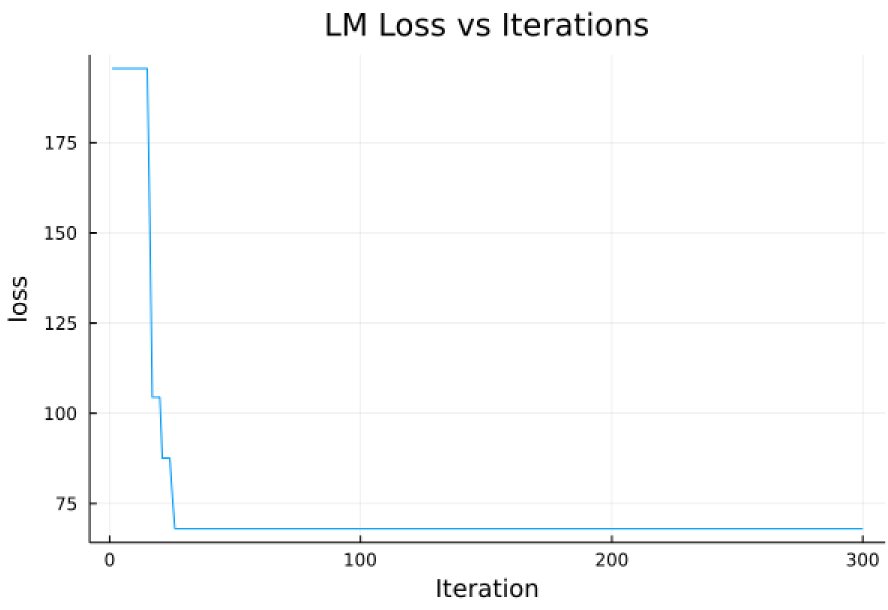
$$f(\theta) = ||r(\theta)||^2 + \lambda||\theta||^2 \tag{۲۵}$$

برای حل این مسئله مسئله را به یک مسئله‌ی بهینه سازی تبدیل میکنم.

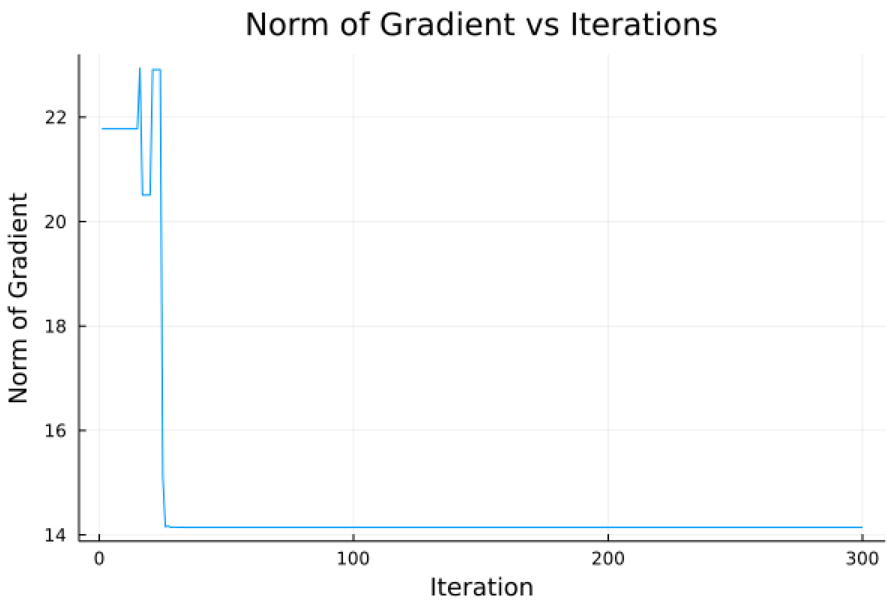
$$\begin{bmatrix} Df(\theta) \\ \sqrt{\lambda} I \end{bmatrix} \begin{bmatrix} \theta \end{bmatrix} = \begin{bmatrix} y + Df(\theta)\theta - f(x,\theta) \\ 0 \end{bmatrix} \tag{۲۶}$$

$$\Rightarrow \theta = (Df^T Df + \lambda I)^{-1} Df^T (y + Df(\theta)\theta - f(x,\theta)) \tag{۲۷}$$

نتیجه‌ی این بهینه سازی نمودارهای زیر شد:



شکل ۱: Objective function vs Iterations



شکل ۲: Gradient Norm vs Iterations

۴. برای حل مسئله به روش خطی معادله زیر را باید حل کنیم.

$$\begin{bmatrix} X^T, 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ v \end{bmatrix} = \begin{bmatrix} y \end{bmatrix} \tag{۲۸}$$

$$\Rightarrow \begin{bmatrix} \beta_1 \\ \beta_2 \\ v \end{bmatrix} = \begin{bmatrix} X^T, 1 \end{bmatrix}^{\dagger} y \tag{۲۹}$$

با حل معادله زیر میرسیم به مقادیر:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ v \end{bmatrix} = \begin{bmatrix} 0.06593416795037406 \\ -0.038839666778029074 \\ 0.006736608984711866 \end{bmatrix} \tag{۳۰}$$

خطای Rms این مدل خطی و مدل شبکه عصبی به شکل زیر است.

The RMS fitting error of linear model is 0.3372979516658594

The RMS fitting error of neural network is 0.3401538124992574

هر دو خطا تقریبا با هم برابر است اما خطای شبکه‌ی عصبی کمی بیشتر است. نباید هم انتظار می‌داشتیم که خطای شبکه عصبی خیلی کمتر شود زیر قدرت شبکه‌های عصبی معمولا در قسمت‌های غیر خطی آن است که باعث می‌شود بتواند مدل‌های پیچیده را حل کند. در این شبکه فقط ۳ تابع sigmoid داشتیم. در صورتی که در شبکه عصبی پارامترهای غیرخطی نداشته باشیم مدل نهایی یک تابع خطی خواهد بود. پس احتمالا در این مسئله اگر تعداد بیشتری پارامتر غیرخطی داشتیم دقت شبکه عصبی با اختلاف بیشتری بهتر میشد.

۴ 19.2 Portfolio optimization with downside risk

.۱

$$\text{Minimize:} \quad \frac{1}{T}\sum_{t=1}^{T}(\max\{p^{tar} - (Rw)_t, 0\}) \tag{۳۱}$$

$$\text{Subject to:} \quad 1^T w = 1, \; \frac{1}{T}1^T(Rw) = p^{tar} \tag{۳۲}$$

پس بنا به شکل بالا خواهیم داشت:

$$f(w) = \frac{1}{T}\begin{bmatrix} \max\{p^{tar} - (Rw)_1, 0\}) \\ \max\{p^{tar} - (Rw)_2, 0\}) \\ \vdots \\ \max\{p^{tar} - (Rw)_T, 0\}) \end{bmatrix} \qquad g(w) = \begin{bmatrix} 1^T w - 1 \\ \frac{1}{T}1^T(Rw) - p^{tar} \end{bmatrix} \tag{۳۳}$$

.۲

$$Df(w) = -Diag(\alpha_1, \alpha_2, \ldots, \alpha_T)R \qquad \alpha_i = \begin{cases} 0 & (Rw)_i > p^{tar} \\ 1 & (Rw)_i < p^{tar} \end{cases} \tag{۳۴}$$

.۳ در این مسئله از دیتاست asset_prices.csv از این سایت استفاده کردم. مسئله‌ی Portfolio optimization برای $p_tar = 0.2/250$ حل کردم و به downside risk با مقدار 0.004116656070326331 رسیدم. حال قصد دارم از Portfolio به دست آورده به عنوان نقطه‌ی اولیه برای مینیمم کردن تابع زیر استفاده کنم.

$$\text{Minimize:} \quad ||f(w^{(k)}) + Df(w^{(k)})(w - w^{(k)})||^2 + \lambda||w - w^{(k)}||^2 \tag{۳۵}$$

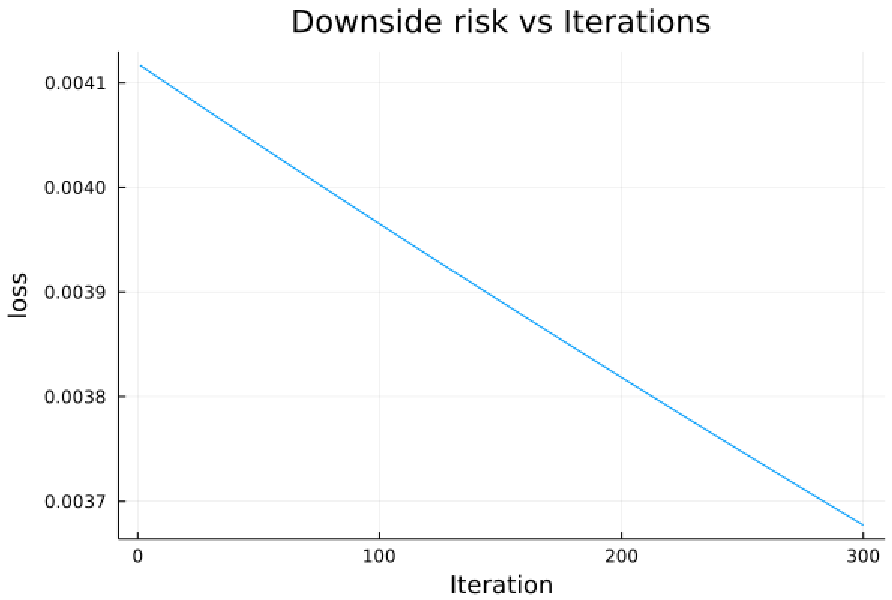
$$\text{Subject to:} \quad g(w) = 0 \tag{۳۶}$$

نتیجه نهایی به شکل زیر شد:

Final Result:

Initial downside risk is 0.004116656070326331

Downside risk of portfolio optimization after 300 iteration is 0.0036772705084149107



شکل ۳: Downside risk vs Iterations

۵ Regression

۱. برای حل این مسئله، خود تابع و مشتق آن را با نام‌های f(θ, x) و df(θ, x) مشخص کردیم و با روش Levenberg–Marquardt با پارامتر بهینه سازی $\lambda = 1e - 4$ مسئله را حل کردیم.
نتیجه نهایی به شکل زیر شد.

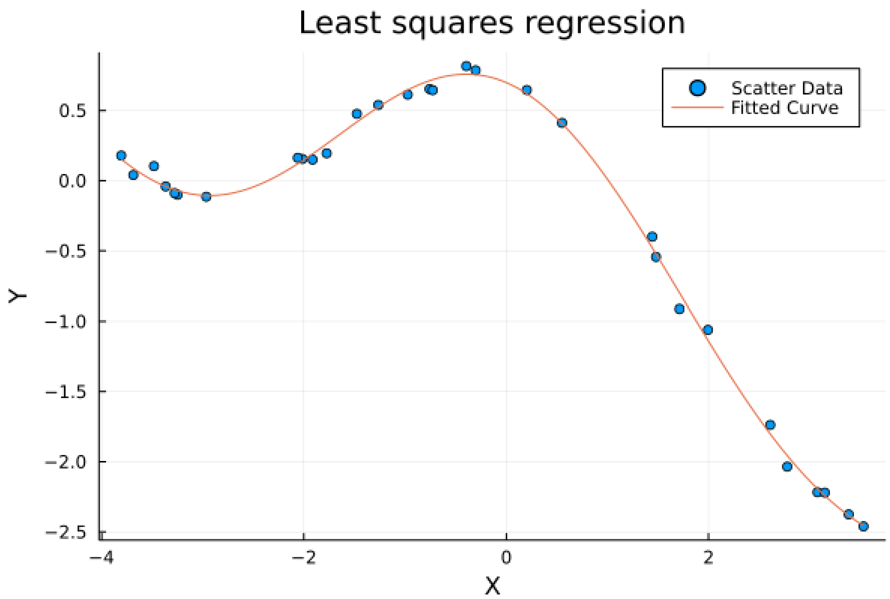
Final Result:

Theta 1 is -1.2616563948328463

Theta 2 is 0.24546488429074367

Theta 3 is 0.7341270658054024

Theta 4 is 1.2262608483148263



شکل ۴: Least Squares Regression

۲. در این مسئله سعی داریم تابع زیر را مینیمم کنیم.

$$\min \sum_{i=1}^{N} (\hat{f}(u^{(i)}, \hat{\theta}) - y^{(i)})^2 + \sum_{i=1}^{N} ||u^{(i)} - x^{(i)}||^2 \tag{۳۷}$$

با داشتن معادله بالا ماتریس f را به شکل زیر تعریف می‌کنیم.

$$F(\hat{\theta}, u)_{2N\times 1} = \begin{bmatrix} \hat{f}(u, \hat{\theta}) - y \\ u - x \end{bmatrix} = \begin{bmatrix} h(\hat{\theta}, u) \\ g(\hat{\theta}, u) \end{bmatrix} \tag{۳۸}$$

پس پارامترهای بهینه سازی در این مسئله علاوه بر θ_1، θ_2، θ_3، θ_4 بردارهای $u^{(1)}$، $u^{(2)}$، ... و $u^{(N)}$ هم هست. همه‌ی مقادیر بهینه‌سازی را در یک بردار به اسم x ذخیره می‌کنیم و ماتریس ژاکوبین را به دست می‌آوریم.

$$x = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 & u^{(1)} & u^{(2)} & \dots & u^{(30)} \end{bmatrix}^T \qquad Df(u, \hat{\theta}) = \begin{bmatrix} Dh(u, \hat{\theta}) \\ Dg(u, \hat{\theta}) \end{bmatrix} \tag{۳۹}$$

$$Dh(u, \hat{\theta}) = \begin{bmatrix} \frac{\partial \hat{f}_1}{\partial \theta_1} & \frac{\partial \hat{f}_1}{\partial \theta_2} & \frac{\partial \hat{f}_1}{\partial \theta_3} & \frac{\partial \hat{f}_1}{\partial \theta_4} & \frac{\partial \hat{f}_1}{\partial u^{(1)}} & \frac{\partial \hat{f}_1}{\partial u^{(1)}} & \cdots & \frac{\partial \hat{f}_1}{\partial u^{(30)}} \\ \frac{\partial \hat{f}_2}{\partial \theta_1} & \frac{\partial \hat{f}_2}{\partial \theta_2} & \frac{\partial \hat{f}_2}{\partial \theta_3} & \frac{\partial \hat{f}_2}{\partial \theta_4} & \frac{\partial \hat{f}_2}{\partial u^{(1)}} & \frac{\partial \hat{f}_2}{\partial u^{(1)}} & \cdots & \frac{\partial \hat{f}_2}{\partial u^{(30)}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \hat{f}_N}{\partial \theta_1} & \frac{\partial \hat{f}_N}{\partial \theta_2} & \frac{\partial \hat{f}_N}{\partial \theta_3} & \frac{\partial \hat{f}_N}{\partial \theta_4} & \frac{\partial \hat{f}_N}{\partial u^{(1)}} & \frac{\partial \hat{f}_N}{\partial u^{(1)}} & \cdots & \frac{\partial \hat{f}_N}{\partial u^{(30)}} \end{bmatrix} \tag{۴۰}$$

$$\frac{\partial \hat{f}}{\partial \theta} = \begin{bmatrix} e^{\theta_2 u^{(1)}} & \theta_1 u^{(1)} e^{\theta_2 u^{(1)}} & \cos(u^{(1)}) & 1 \\ e^{\theta_2 u^{(2)}} & \theta_1 u^{(2)} e^{\theta_2 u^{(2)}} & \cos(u^{(2)}) & 1 \\ \vdots & \vdots & \vdots & \vdots \\ e^{\theta_2 u^{(N)}} & \theta_1 u^{(N)} e^{\theta_2 u^{(N)}} & \cos(u^{(N)}) & 1 \end{bmatrix} \tag{۴۱}$$

$$\frac{\partial \hat{f}}{\partial u} = Diag(\theta_1 \theta_2 e^{\theta_2 u^{(i)}} - \theta_3 \sin(u^{(i)})) \tag{۴۲}$$

پیدا کردن $Dg(u, \hat{\theta})$ هم ساده است زیرا مشتق نسبت به θ برابر صفر و مشتق نسبت به u برابر ماتریس همانی است. نتیجه نهایی به شکل زیر شد.

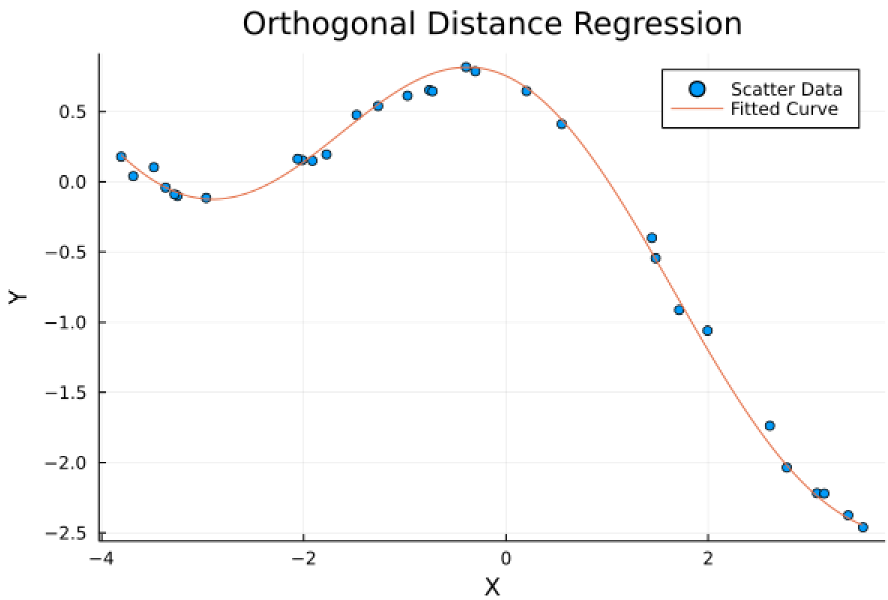
Final Result:

Theta 1 is -1.8306640691965865

Theta 2 is 0.1793195040997058

Theta 3 is 0.8220479344838888

Theta 4 is 1.7625486687139993



شکل ۵: Orthogonal Distance Regression

برای مقدار دهی اولیه این مسئله، مجهولات θ را به صورت رندوم انتخاب کردیم و مقادیر اولیه‌ی u را برابر همان مقدار x متناظر آن مشخص کردیم.