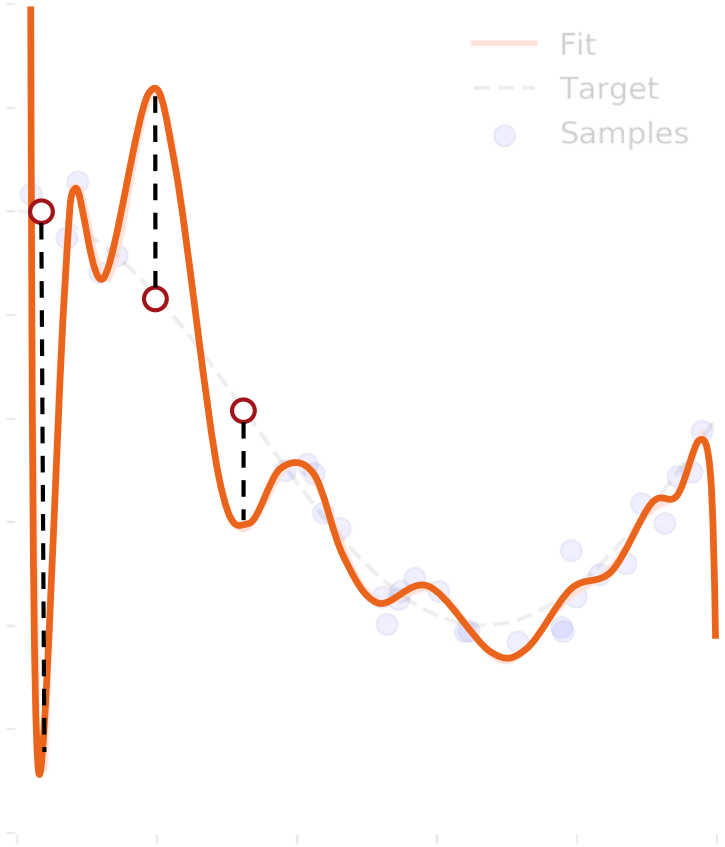
تنظیم: برخورد با بیش‌برازش

سید ناصر رضوی www.snrazavi.ir

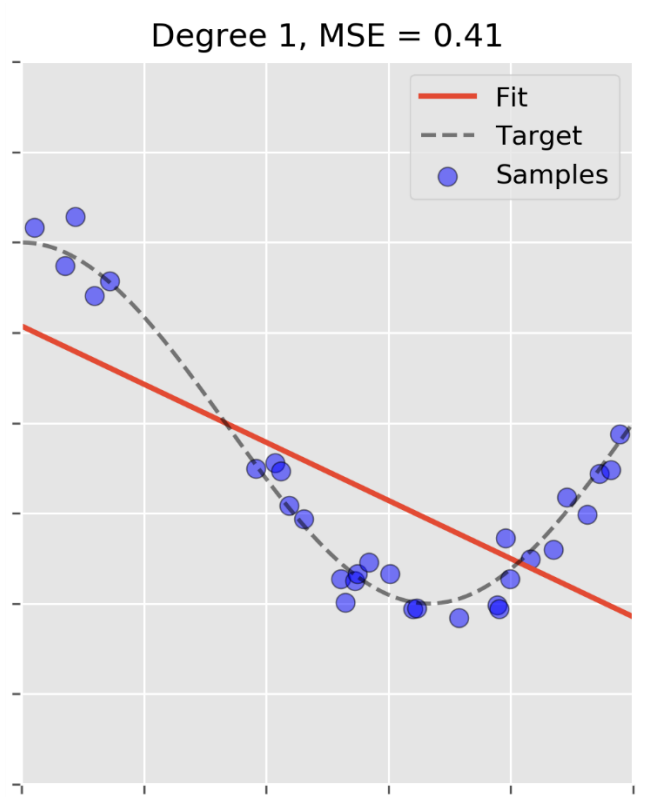
۱۳۹۷

بیش‌برازش و تنظیم

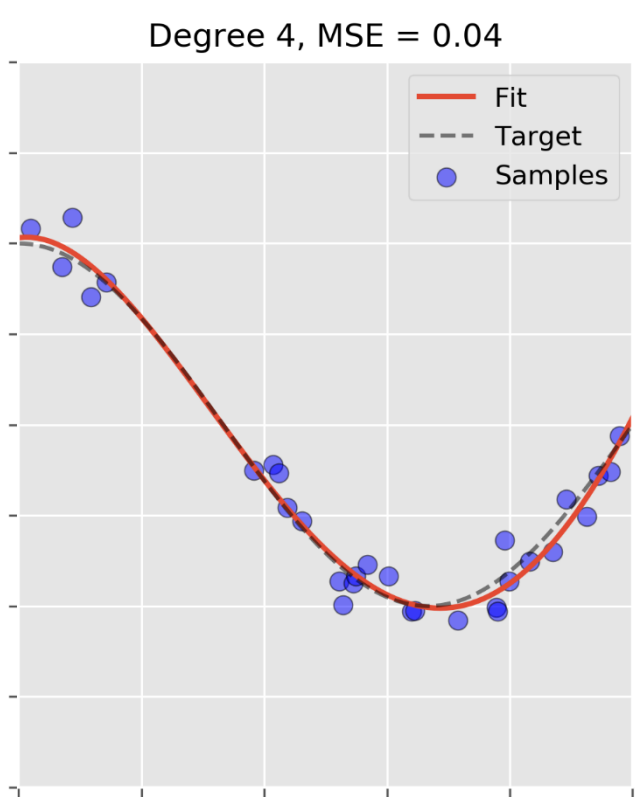


- بیش‌برازش. یک مشکل بسیار متداول در یادگیری ماشین
  - مدل بیش از حد نیاز پیچیده
    - مثلا به دلیل تعداد بسیار زیاد ویژگی‌ها
  - عملکرد بسیار خوب مدل بر روی داده‌های آموزشی
  - عملکرد بسیار بد مدل بر روی داده‌های جدید
    - عدم قابلیت تعمیم برای داده‌های جدید!
- تنظیم. یک روش مؤثر برای کاهش یا حذف بیش‌برازش.

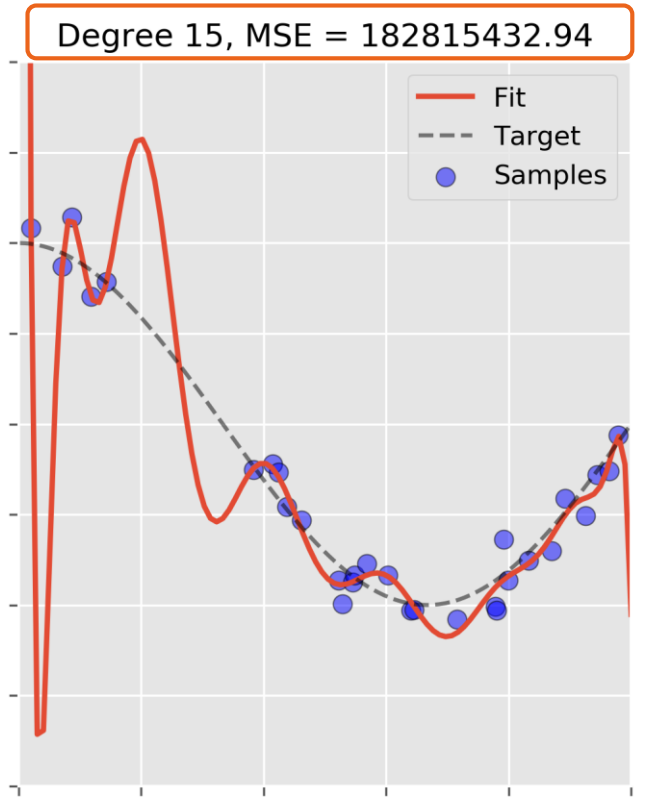
بیش‌برازش و رگرسیون



کم‌برازش

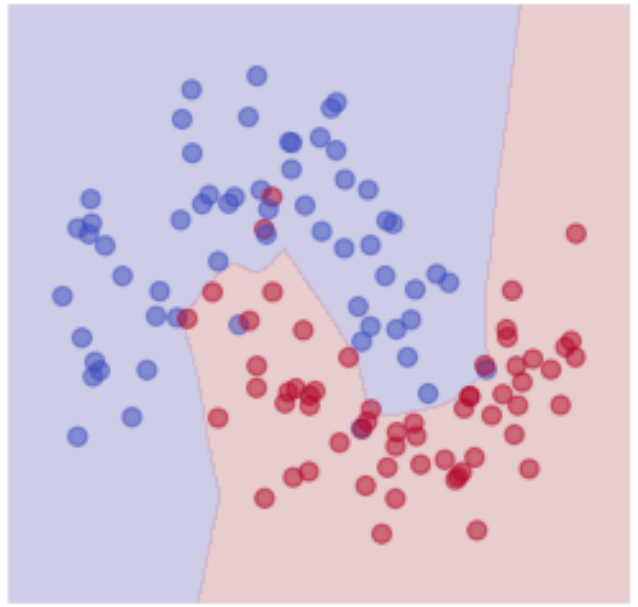


مدل درست

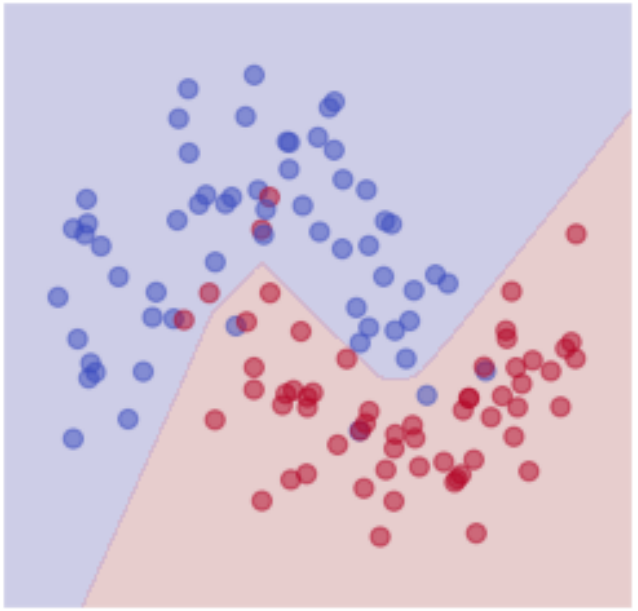


بیش‌برازش

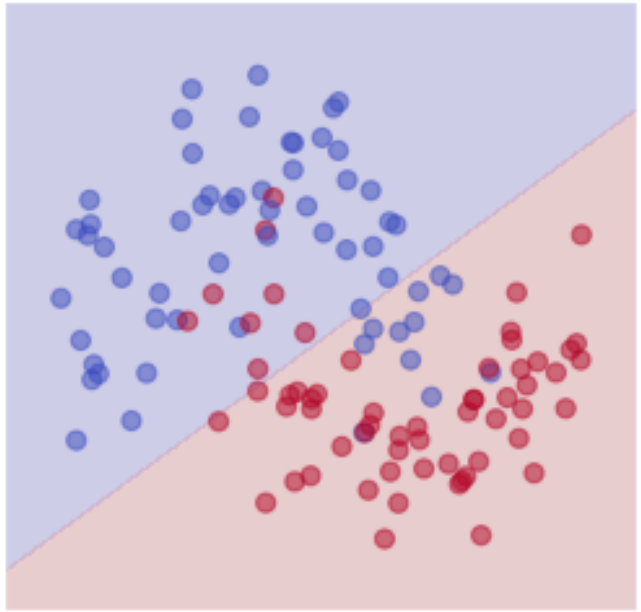
بیش‌برازش و دسته‌بندی



بیش‌برازش



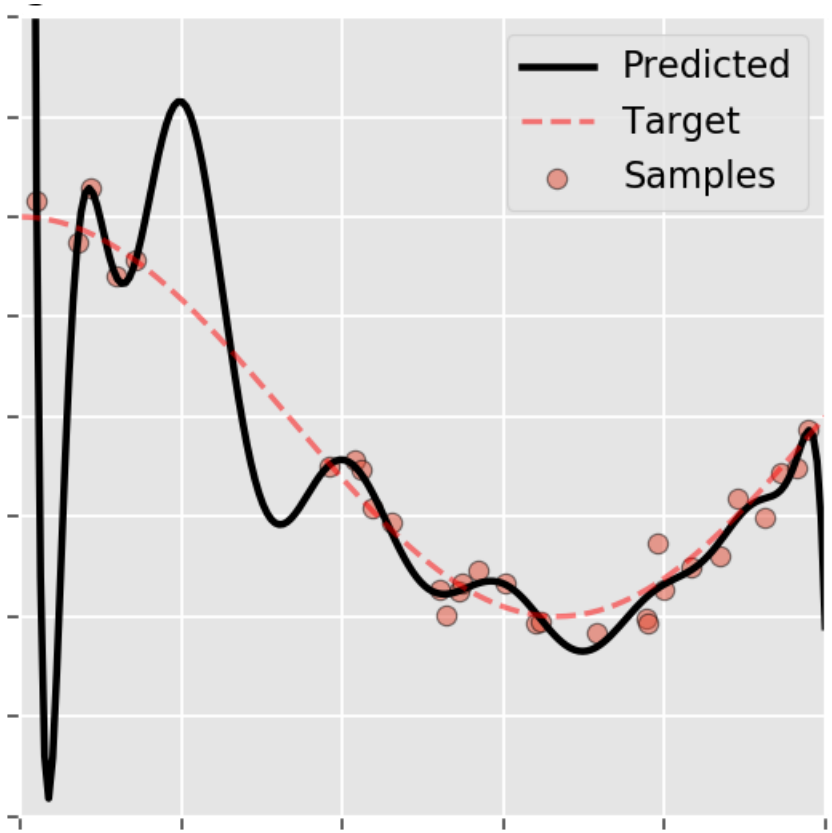
مدل درست

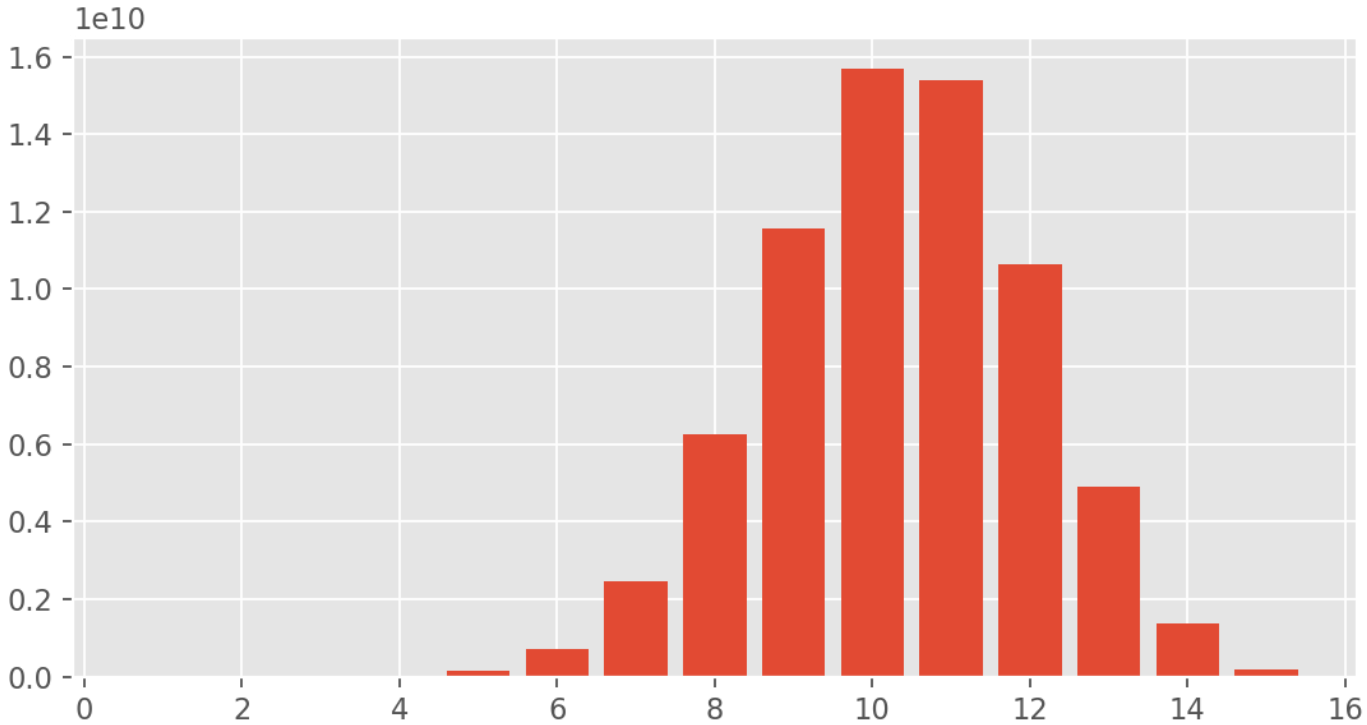


کم‌برازش

تنظیم

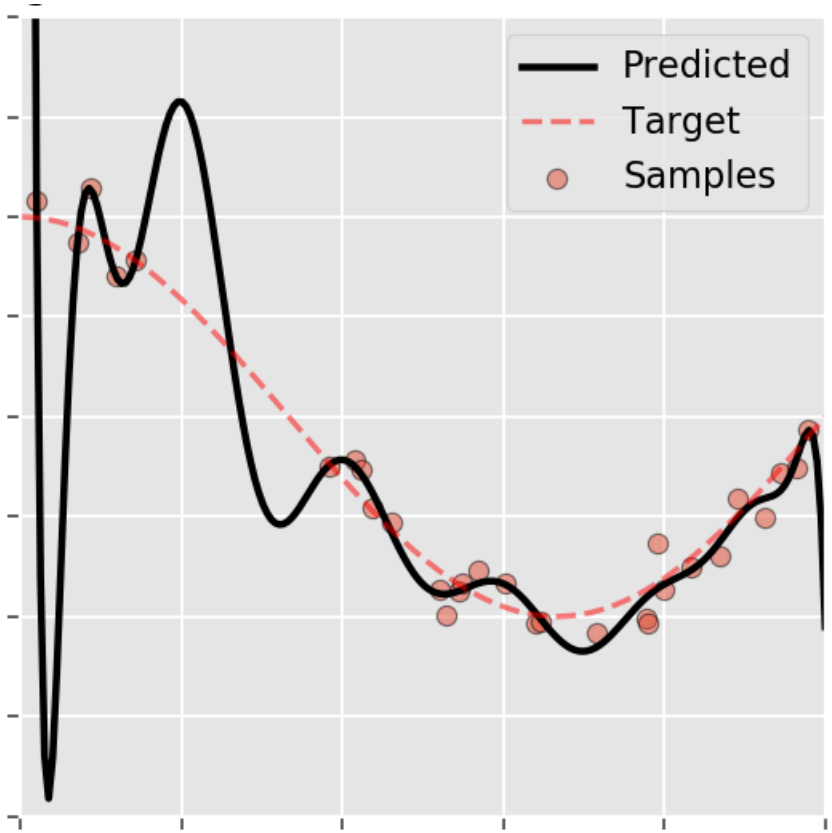
- ایده. جلوگیری از بزرگ شدن بیش از حد مقدار پارامترها با افزودن یک جمله به تابع هزینه به منظور جریمه کردن مقادیر بزرگ پارامترها.





تنظیم

- ایده. جلوگیری از بزرگ شدن بیش از حد مقدار پارامترها با افزودن یک جمله به تابع هزینه به منظور جریمه کردن مقادیر بزرگ پارامترها.



$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost\left(x^{(i)}, y^{(i)}\right) + \lambda\, R(\theta)$$

تنظیم L2

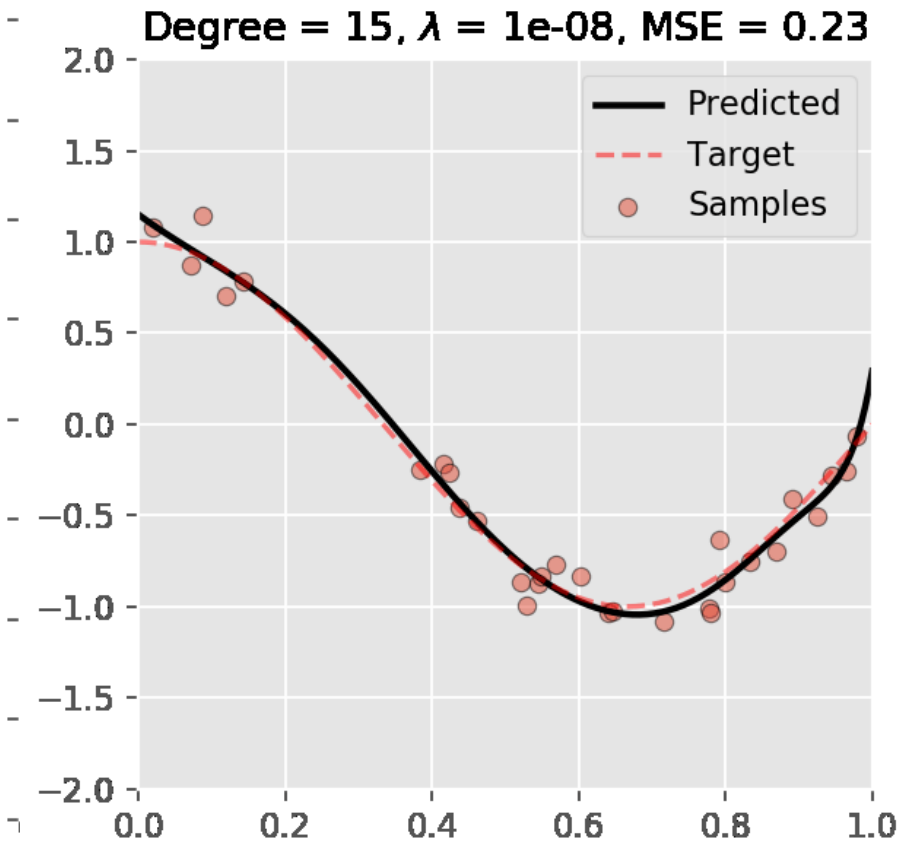
$$R(\theta) = \sum_{j=1}^{n} \theta_j^2 = \|\theta\|_2^2$$

تنظیم L1

$$R(\theta) = \sum_{j=1}^{n} |\theta_j| = \|\theta\|_1$$

تنظیم

- ایده. جلوگیری از بزرگ شدن بیش از حد مقدار پارامترها با افزودن یک جمله به تابع هزینه به منظور جریمه کردن مقادیر بزرگ پارامترها.



$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost\left(x^{(i)}, y^{(i)}\right) + \lambda\, R(\theta)$$

- ضریب تنظیم. برقرای توازن میان اهداف فوق.

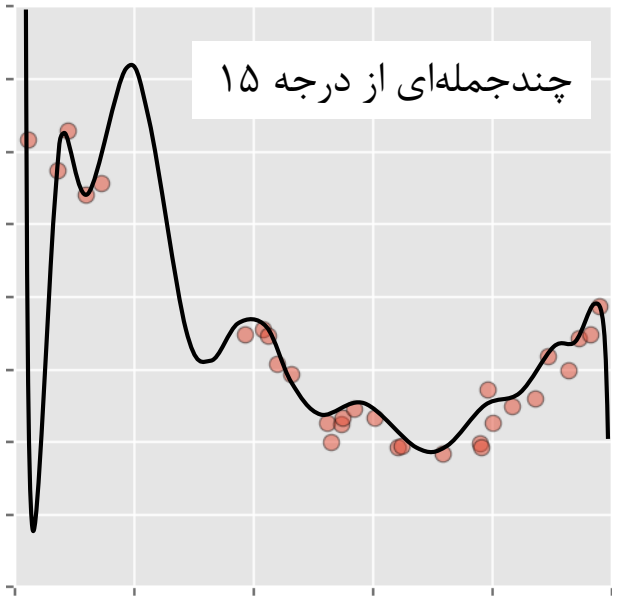
$\lambda \to 0$ ⟸ دادن اهمیت بیشتر به خطای مجموعه آموزشی

$\lambda \to \infty$ ⟸ دادن اهمیت بیشتر به خطای تعمیم

۸ تابع هزینه

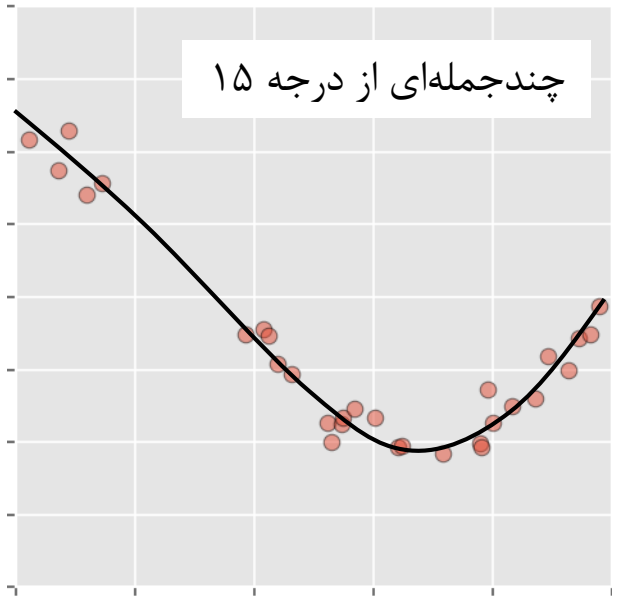
تنظیم

رگرسیون بدون تنظیم



$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost\left(x^{(i)}, y^{(i)}\right)$$

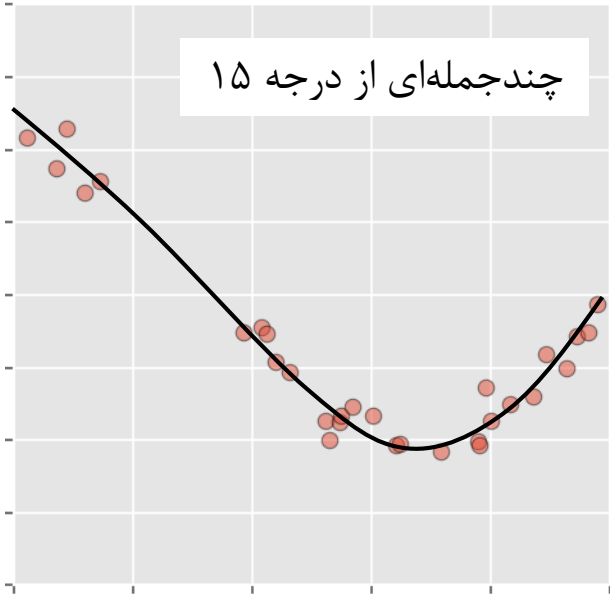
رگرسیون با تنظیم



$$\frac{1}{m}\sum_{i=1}^{m} Cost\left(x^{(i)}, y^{(i)}\right) + \lambda\, R(\theta)$$

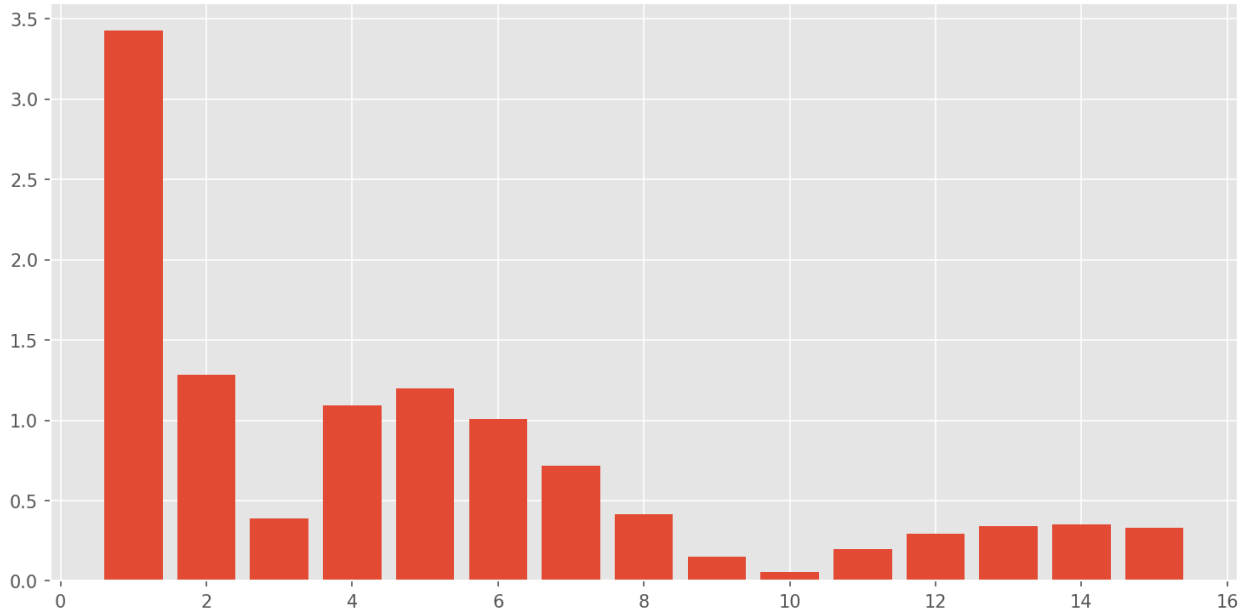
تنظیم

رگرسیون با تنظیم



$$\frac{1}{m}\sum_{i=1}^{m} Cost\left(x^{(i)}, y^{(i)}\right) + \lambda\, R(\theta)$$

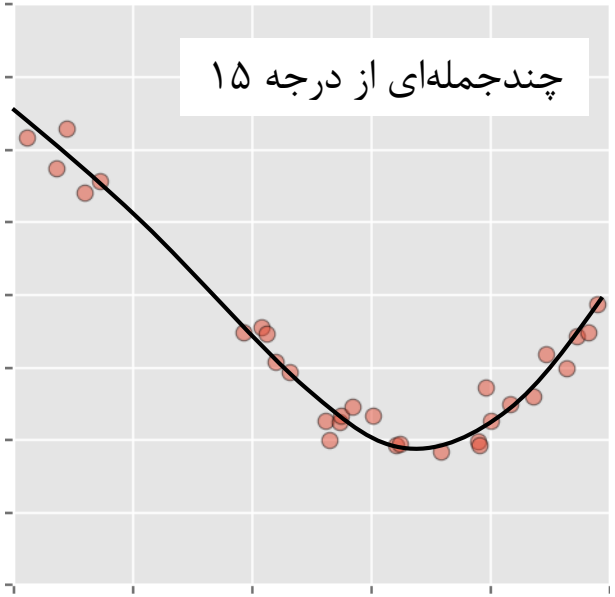
مقدار پارامترها در صورت استفاده از تنظیم



$$R(\theta) = \sum_{j=1}^{n} \theta_j^2 = \|\theta\|_2^2$$ تنظیم L2

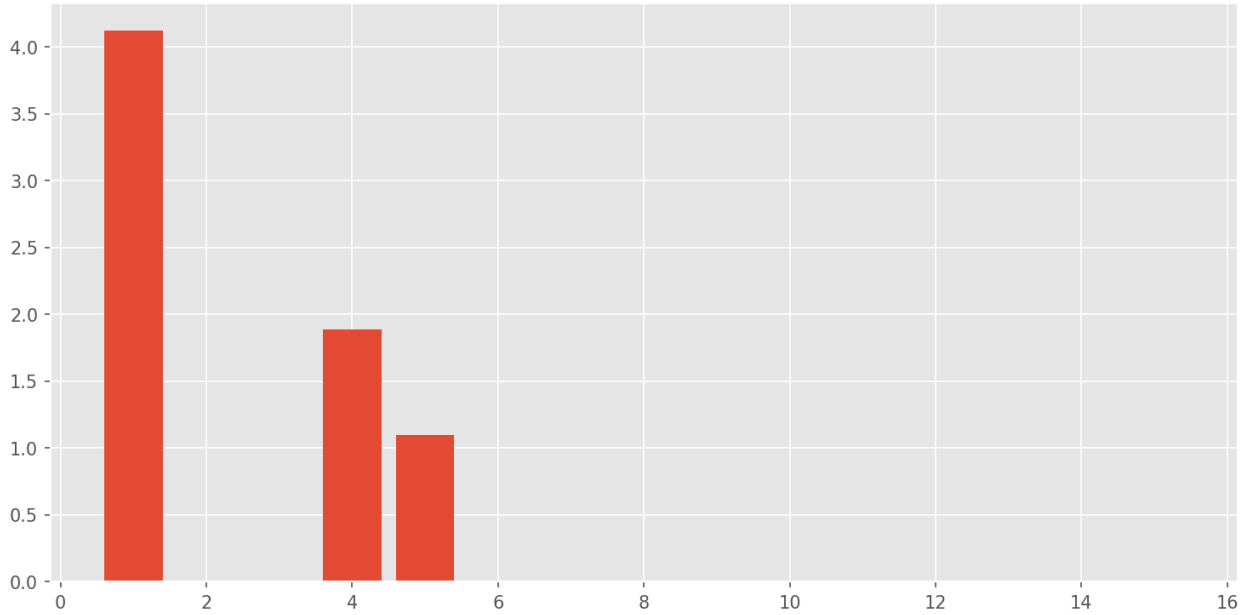
تنظیم

رگرسیون با تنظیم



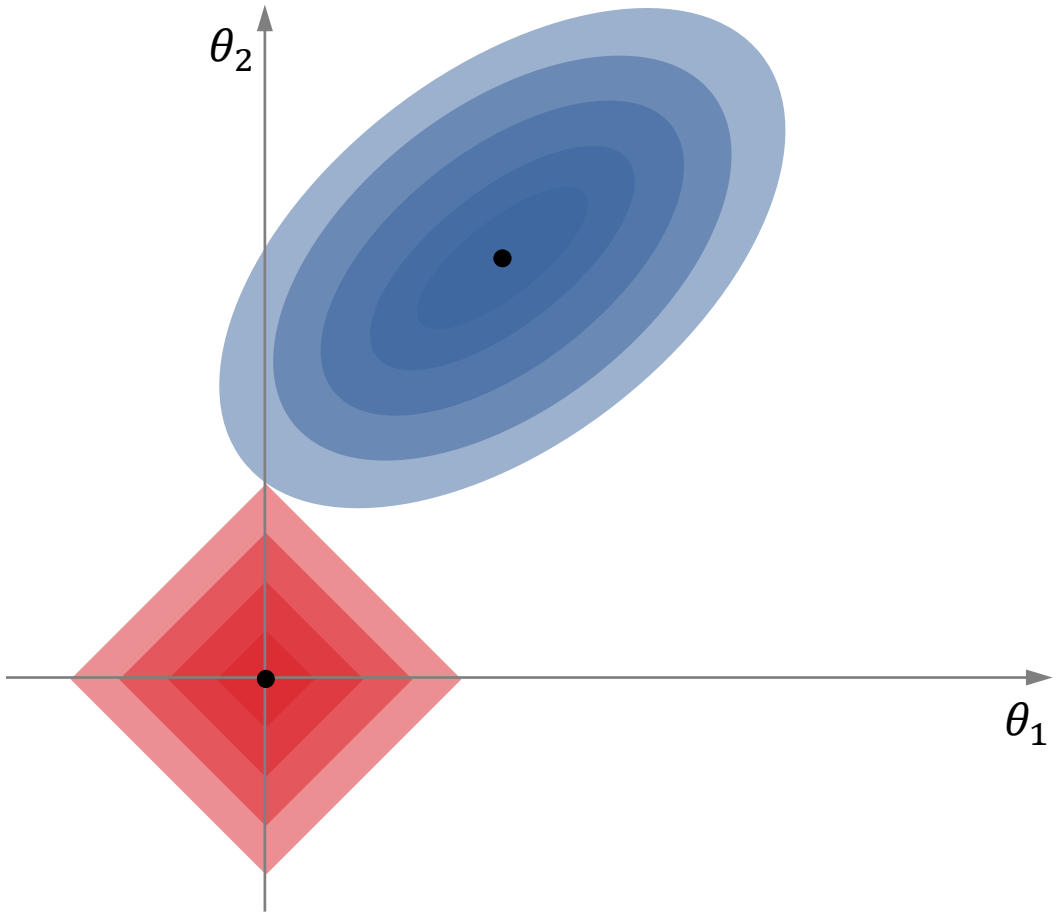
$$\frac{1}{m}\sum_{i=1}^{m} Cost\left(x^{(i)}, y^{(i)}\right) + \lambda\, R(\theta)$$

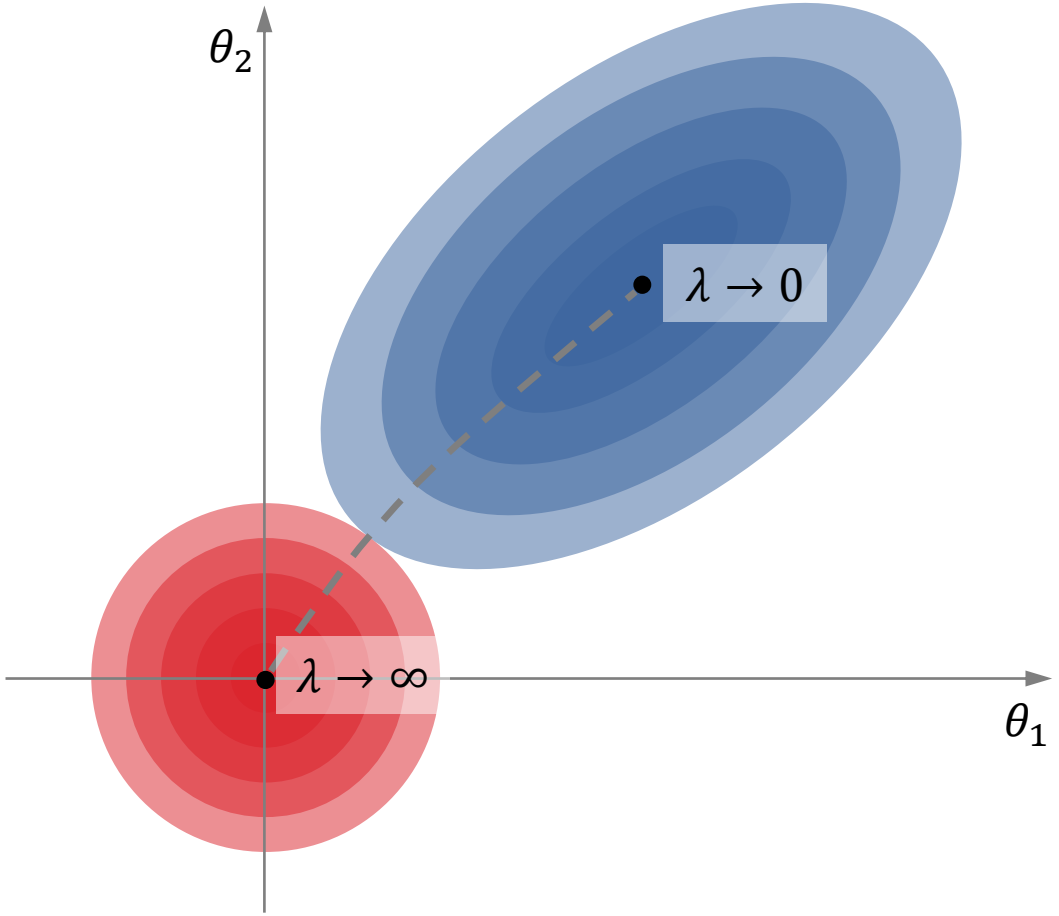
مقدار پارامترها در صورت استفاده از تنظیم



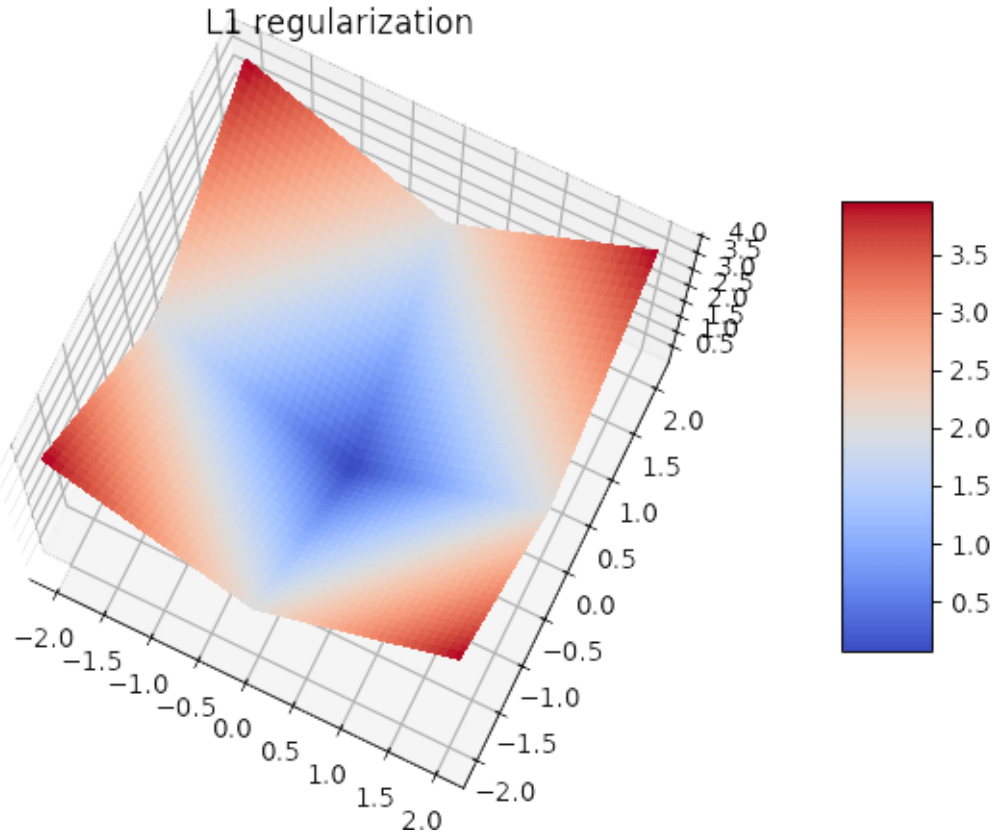
$$R(\theta) = \sum_{j=1}^{n} |\theta_j| = \|\theta\|_1$$ تنظیم L1

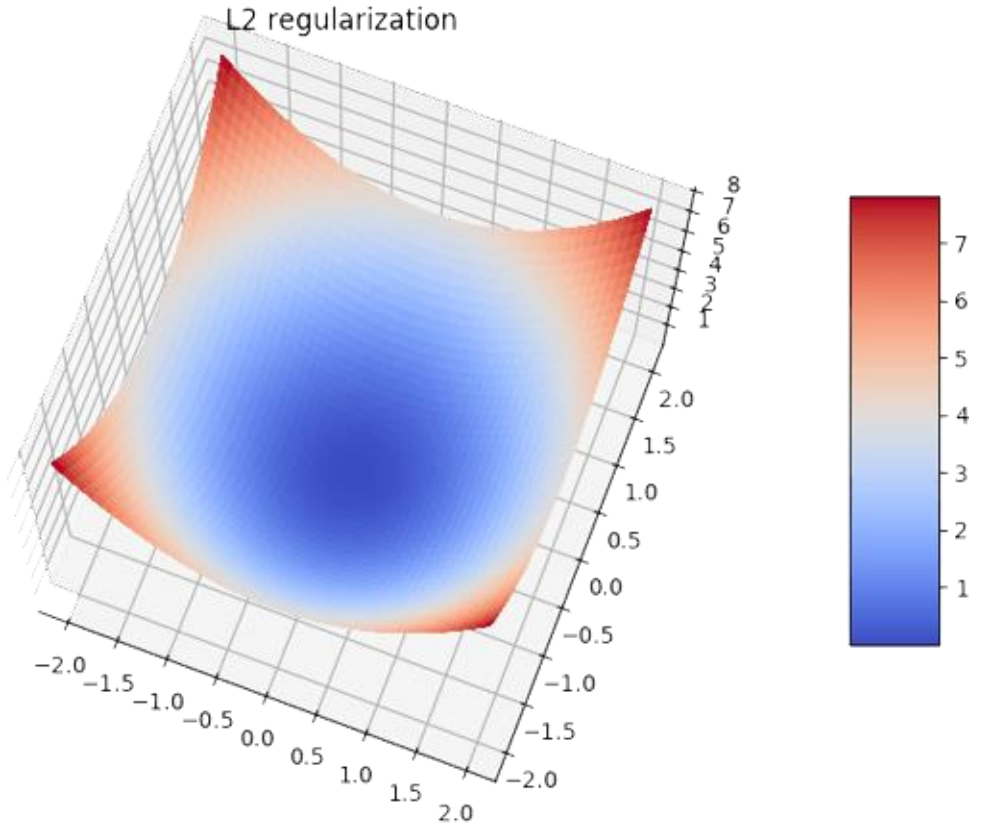
تنظیم L1 و L2





تنظیم L1 و L2


L1 regularization
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5


L2 regularization
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
8
7
6
5
4
3
2
1
7
6
5
4
3
2
1

رگرسیون خطی تنظیم شده

رگرسیون خطی تنظیم شده

- تابع هزینه.

$$J(\theta) = \frac{1}{2}\left[\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \lambda\sum_{j=1}^{n}\theta_j^2\right] \quad = \frac{1}{2}(X\theta - y)^T(X\theta - y) + \frac{1}{2}\lambda\theta^T\theta$$

- هدف. کمینه‌سازی تابع هزینه به منظور یافتن مقدار بهینه پارامترها

$$\min_\theta J(\theta)$$

گرادیان کاهشی (بدون تنظیم)

- بدون استفاده از تنظیم.

```
repeat until convergence {

}
```

$$\theta_j = \theta_j - \alpha \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_j^{(i)} \qquad (j = 0,1,2, \dots, n)$$

گرادیان کاهشی (با تنظیم)

- با استفاده از تنظیم.

repeat until convergence {

$$\theta_0 = \theta_0 - \alpha \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_0^{(i)}$$

$$\theta_j = \theta_j - \alpha \left[\sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_j^{(i)} + \lambda\theta_j\right] \qquad (j = 1,2,\dots,n)$$

}

$$\theta_j = \theta_j \underbrace{(1 - \alpha\lambda)}_{<1} - \alpha \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_j^{(i)}$$

معادله نرمال (با تنظیم)

$$J(\theta) = \frac{1}{2}(X\theta - y)^T(X\theta - y) + \frac{1}{2}\lambda\theta^T\theta$$

$$\frac{\partial J}{\partial \theta} = X^T(X\theta - y) + \lambda\theta$$

$$= X^TX\theta - X^Ty + \lambda\theta$$

$$= (X^TX + \lambda I)\theta - X^Ty = 0$$

$$(X^TX + \lambda I)\theta = X^Ty$$

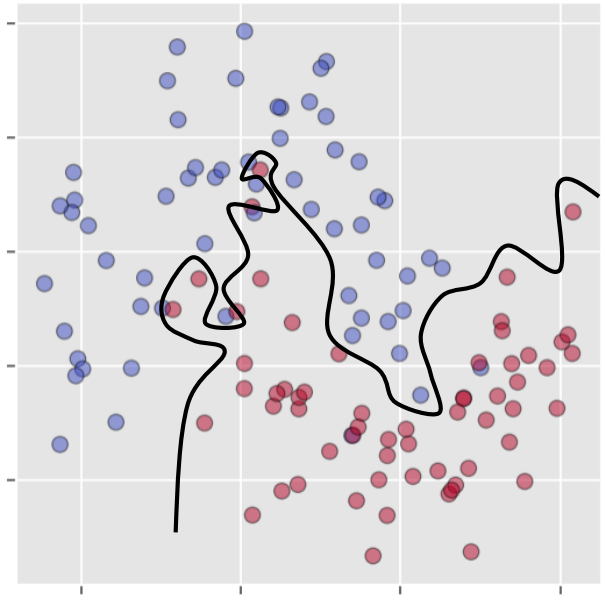
$$\theta = \underbrace{(X^TX + \lambda I)}_{\text{وارون پذیر } (\lambda > 0)}{}^{-1}X^Ty$$

$$\theta = \left(X^TX + \lambda \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \right)^{-1} X^Ty$$

رگرسیون لجستیک تنظیم شده

رگرسیون لجستیک (بدون تنظیم)

- فرضیه.

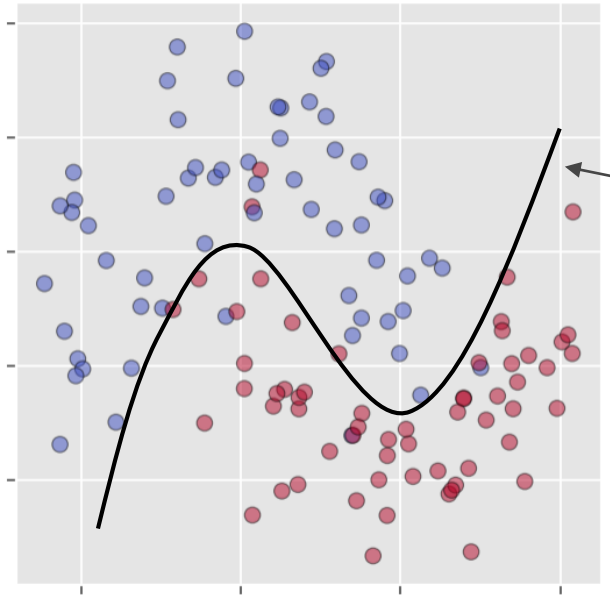


چندجمله‌ای از درجه ۱۵

- تابع هزینه.

$$J(\theta) = -\sum_{i=1}^{m} y^{(i)} \log h_\theta\left(x^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right)$$

رگرسیون لجستیک (با تنظیم)



چندجمله‌ای از درجه ۱۵

- فرضیه.

- تابع هزینه.

$$J(\theta) = -\sum_{i=1}^{m} y^{(i)} \log h_\theta\left(x^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right) + \frac{\lambda}{2} \sum_{j=1}^{n} \theta_j^2$$

گرادیان کاهشی

- با استفاده از تنظیم.

repeat until convergence {

$$\theta_0 = \theta_0 - \alpha \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_0^{(i)}$$

$$\theta_j = \theta_j - \alpha \left[\sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_j^{(i)} + \lambda\theta_j\right] \qquad (j = 1,2,\dots,n)$$

}

$$h_\theta\left(x^{(i)}\right) = \frac{1}{1 + e^{-\theta^T x^{(i)}}}$$

بهینه‌سازی پیشرفته


```
from scipy.optimize import minimize

minimize(J, x0, method='CG', jac=grads)
```

گرادیان

$$\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right)-y^{(i)}\right)\cdot x_j^{(i)}+\lambda\theta_j$$

پیاده‌سازی تابع هزینه

$$-\sum_{i=1}^{m} y^{(i)}\log h_\theta\left(x^{(i)}\right)+\left(1-y^{(i)}\right)\log\left(1-h_\theta\left(x^{(i)}\right)\right)+\frac{\lambda}{2}\sum_{j=1}^{n}\theta_j^2$$