

یادگیری ماشین با پایتون

فرزاد حسینیان

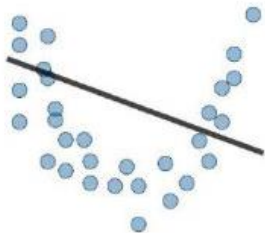
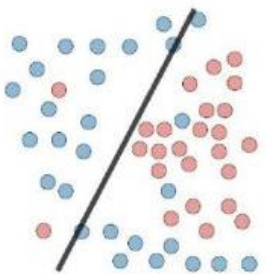
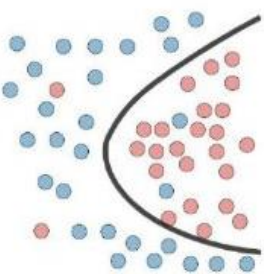
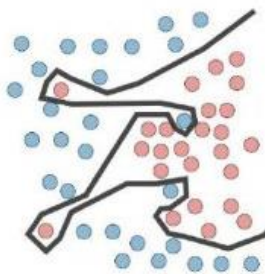
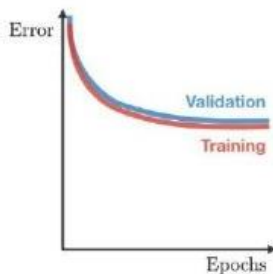
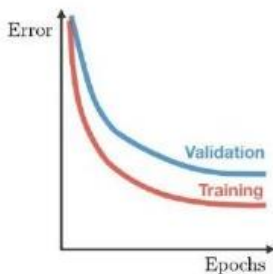
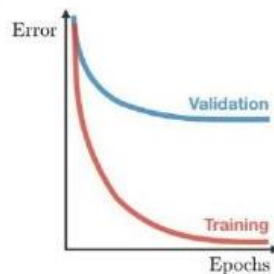
دانشگاه محقق اردبیلی

دانشکده علوم کامپیوتر

آبان 98

جلسه چهارم

رگريسيون لجستیک بخش دوم

	Underfitting	Just right	Overfitting
Symptoms	- High training error - Training error close to test error - High bias	- Training error slightly lower than test error	- Low training error - Training error much lower than test error - High variance
Regression			
Classification			
Deep learning			
Remedies	- Complexify model - Add more features - Train longer		- Regularize - Get more data

راه حل:

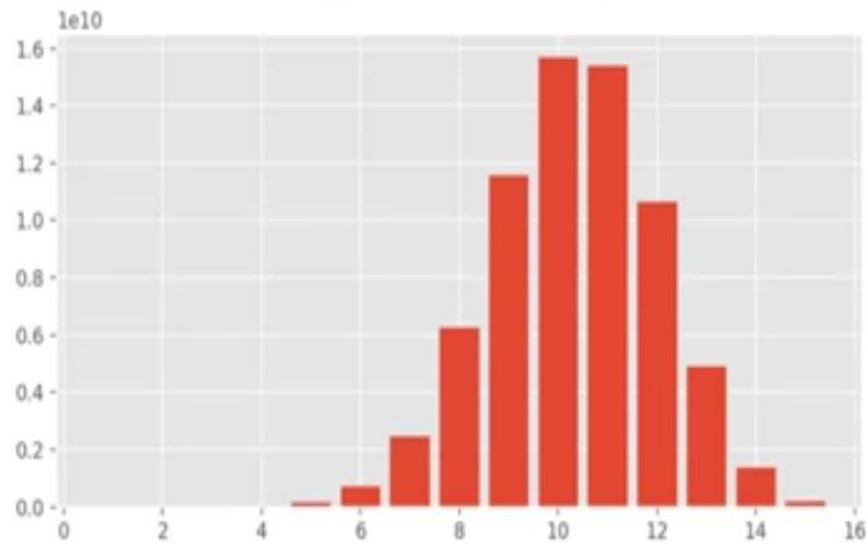
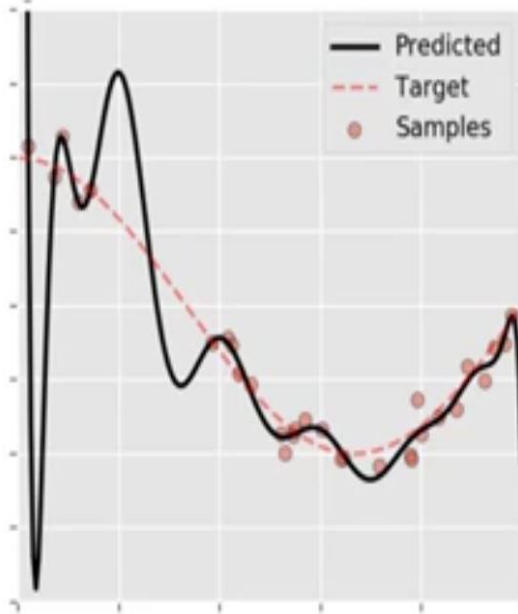
گزینش ویژگی ها (feature selection)

- دستی

- استفاده از regularization:

- . همه ویژگی ها را نگه میداریم
- . تعداد فیچر زیادی داشته باشیم

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



ضریب تنظیم. برقرای توازن میان اهداف فوق.

$\lambda \rightarrow 0$ ← دادن اهمیت بیشتر به فضای مجموعه آموزشی

$\lambda \rightarrow \infty$ ← دادن اهمیت بیشتر به فضای تعمیم

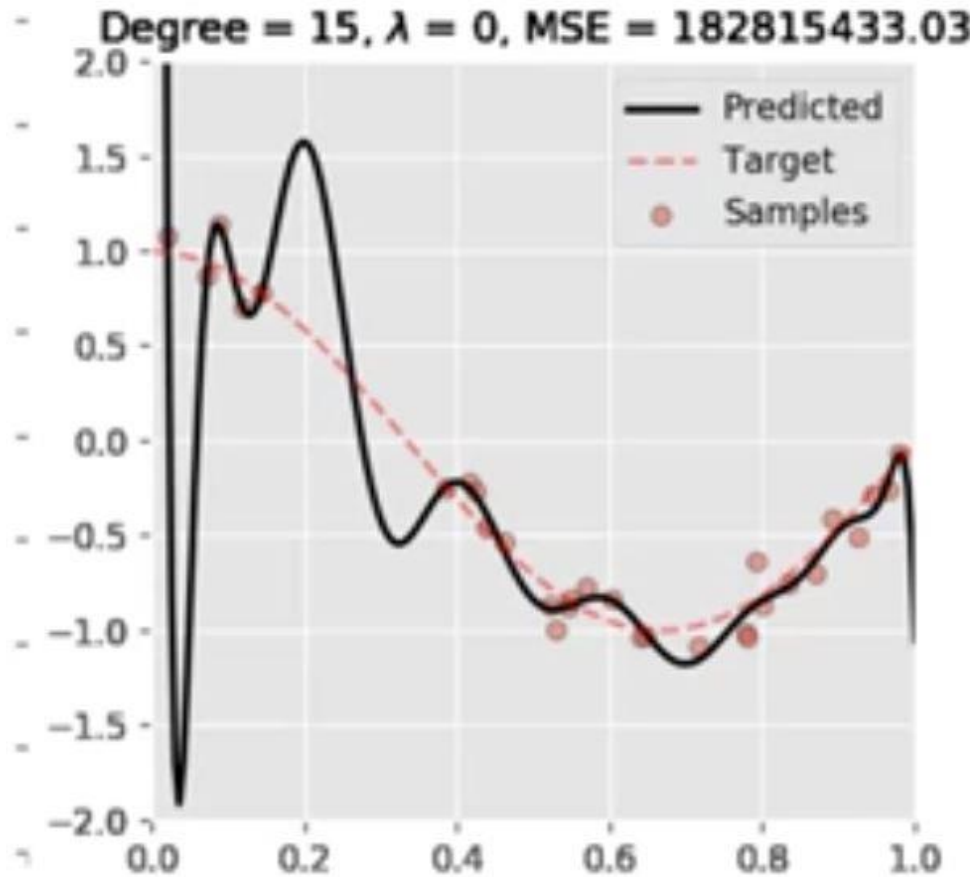
Regularized cost function

$$L(x, y) = \sum_{i=1}^n (y_i - h_{\theta}(x_i))^2$$

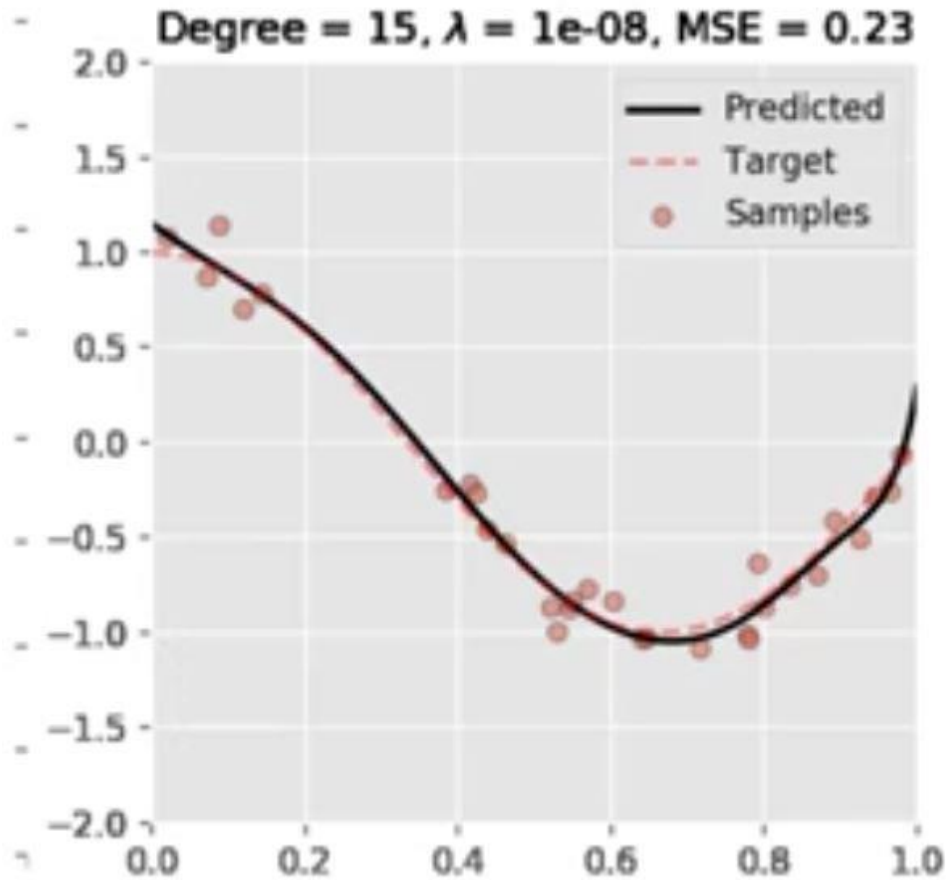
where $h_{\theta}x_i = \theta_0 + \theta_1x_1 + \theta_2x_2^2 + \theta_3x_3^3 + \theta_4x_4^4$

$$L(x, y) \equiv \sum_{i=1}^n (y_i - h_{\theta}(x_i))^2 + \lambda \sum_{i=1}^n \theta_i^2$$

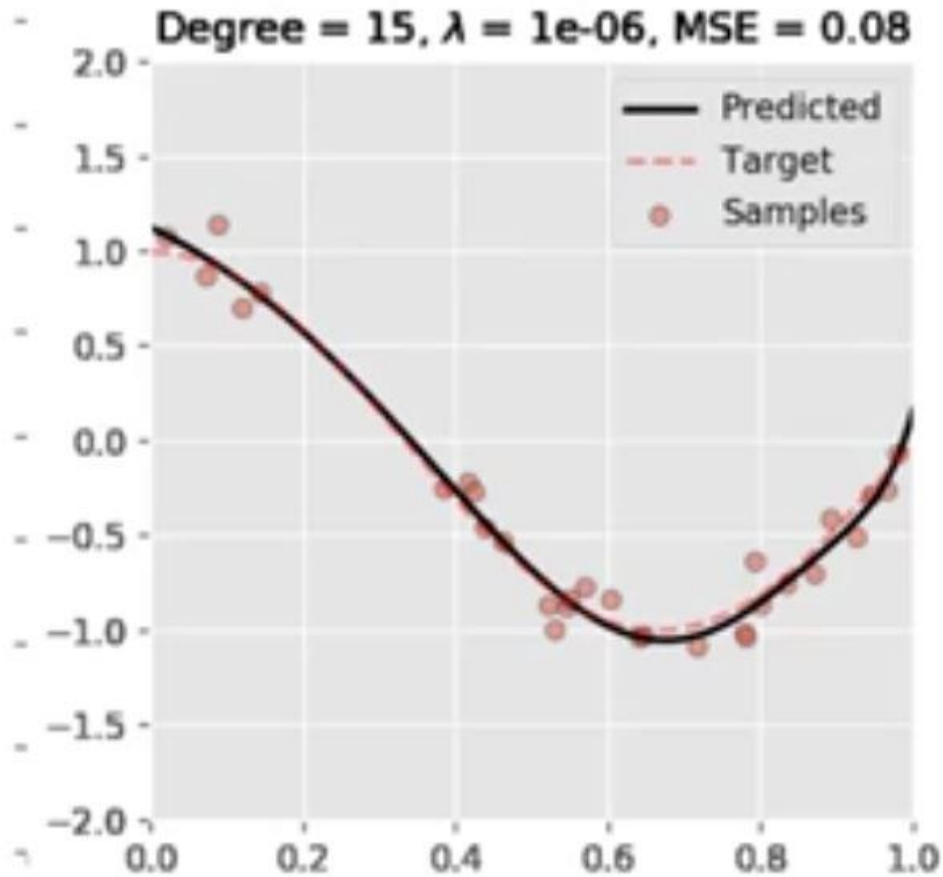
تغییرات لامبدا



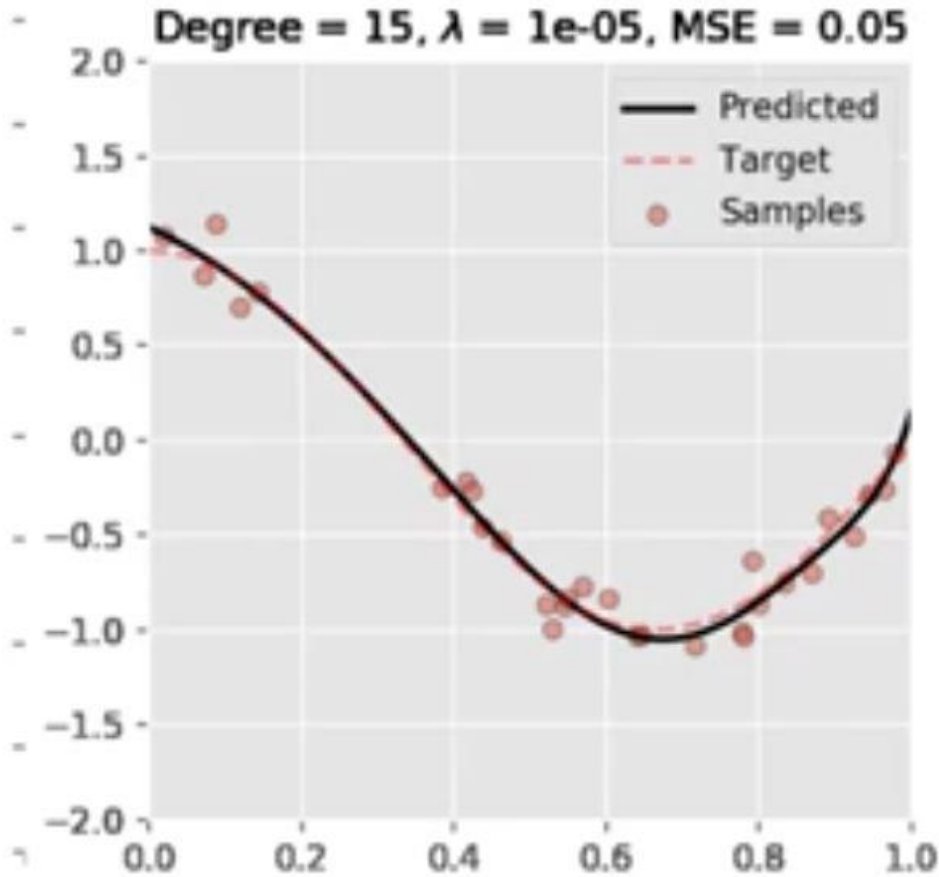
تغییرات لامبدا



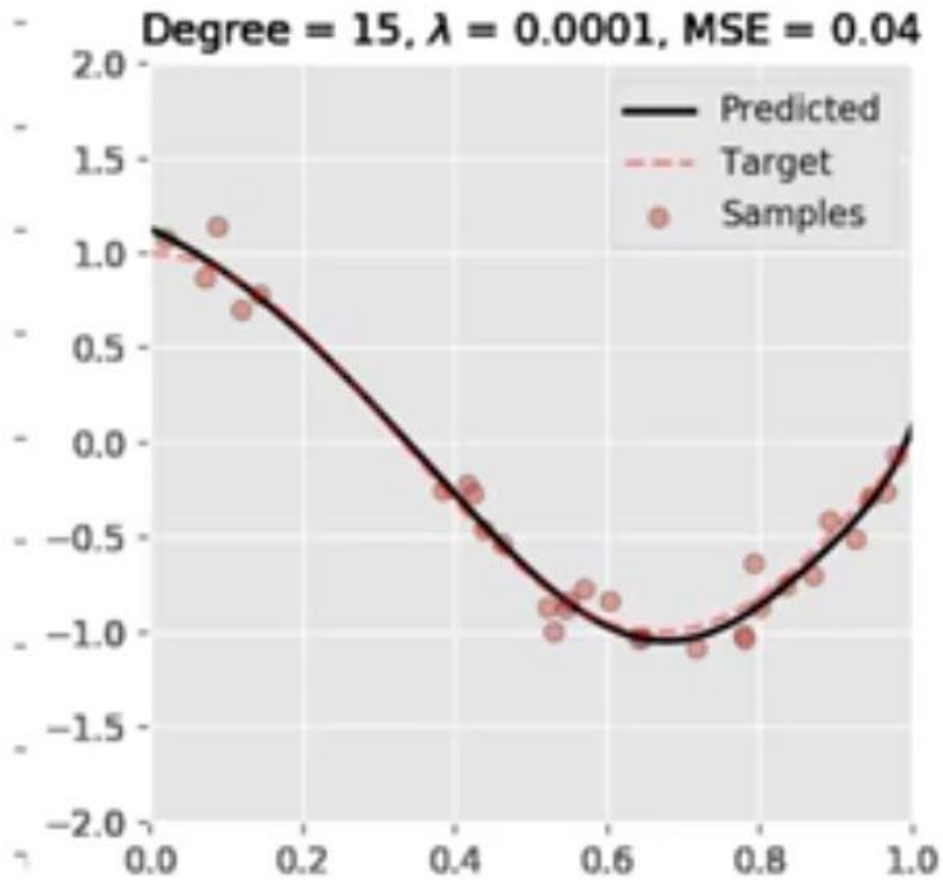
تغییرات لامبدا



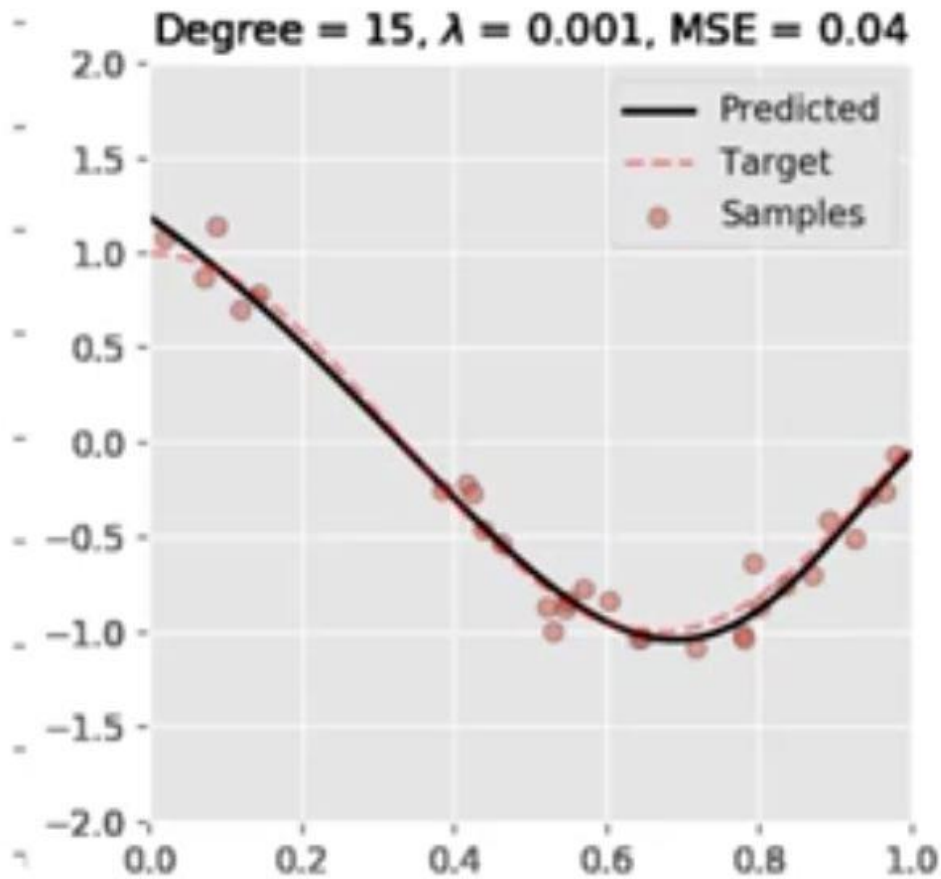
تغییرات لامبدا



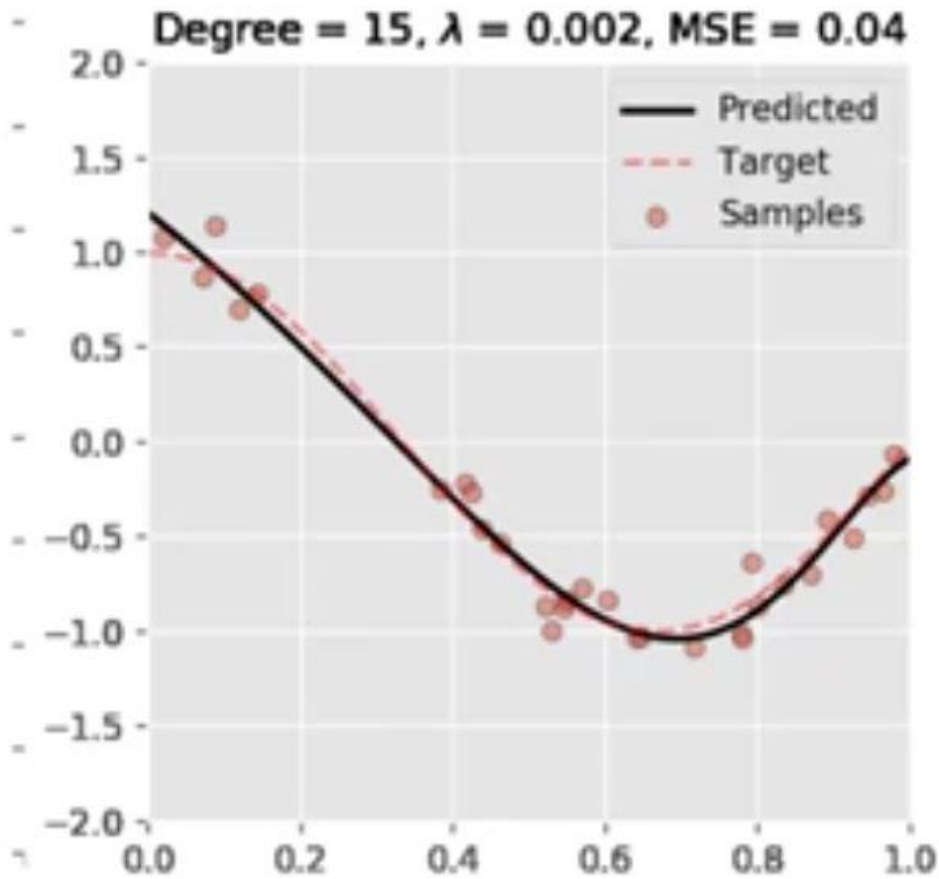
تغییرات لامبدا



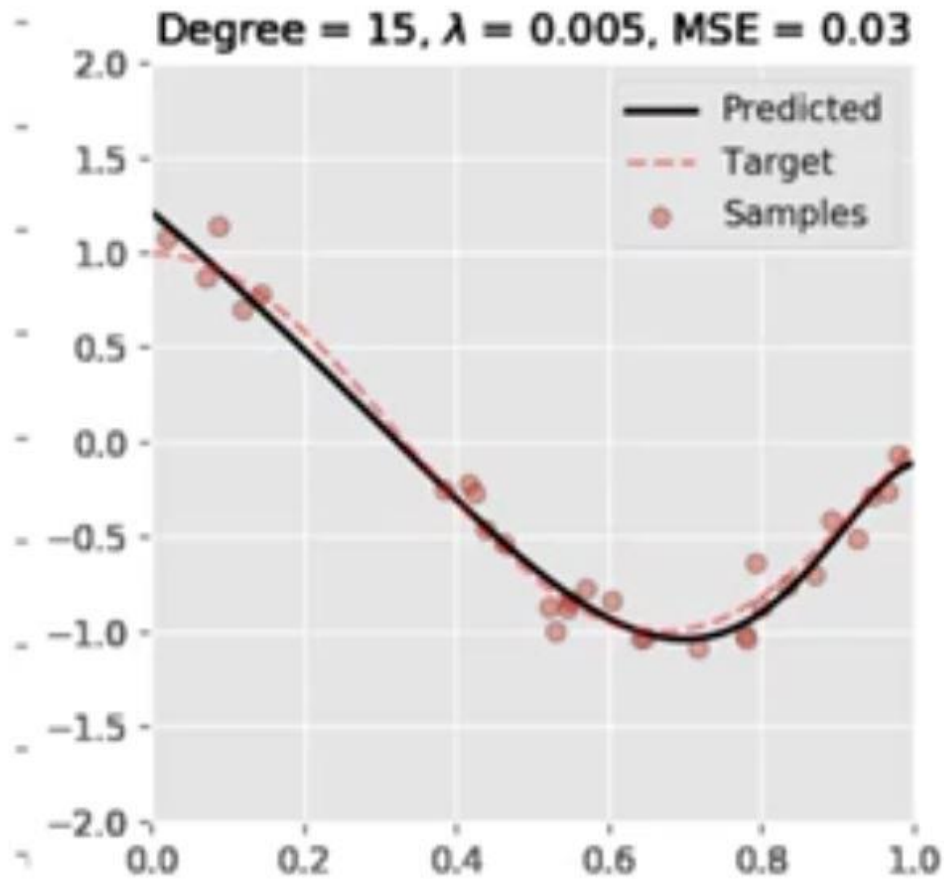
تغییرات لامبدا



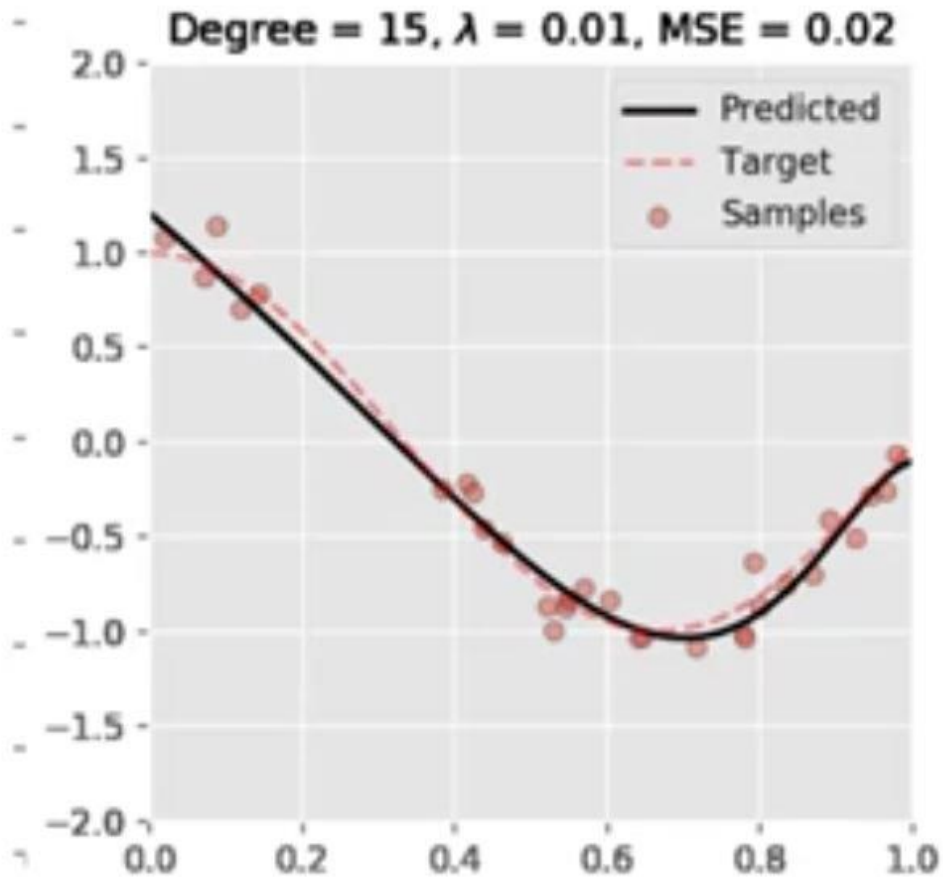
تغییرات لامبدا



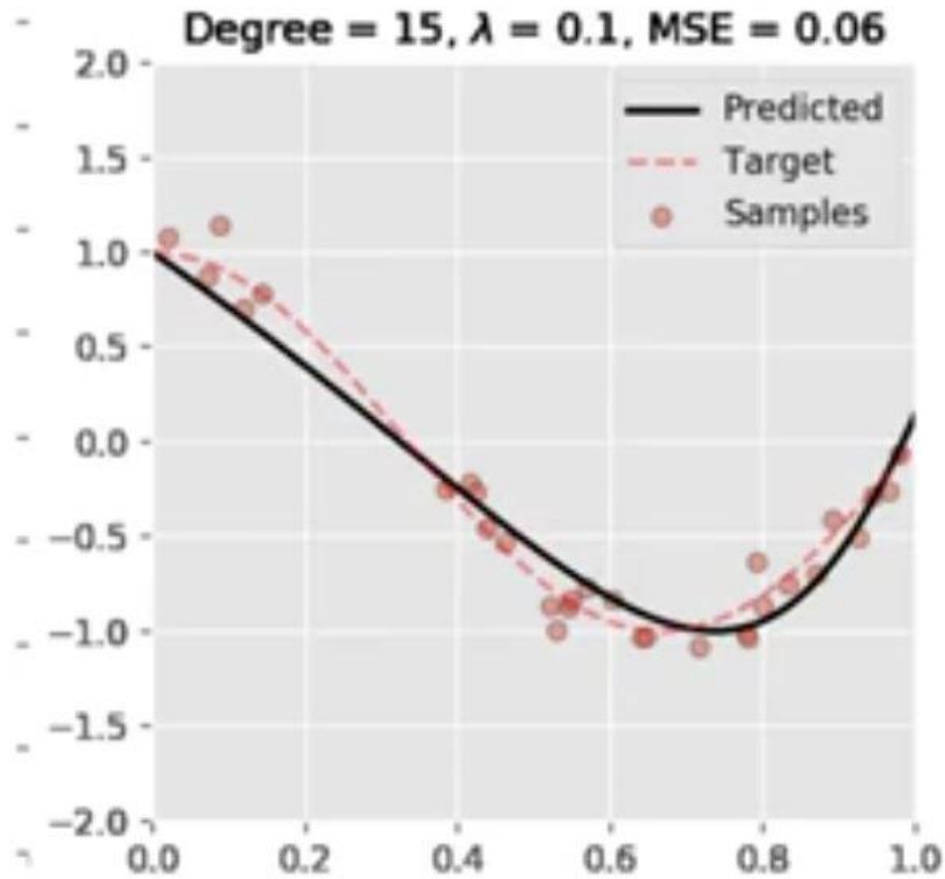
تغییرات لامبدا



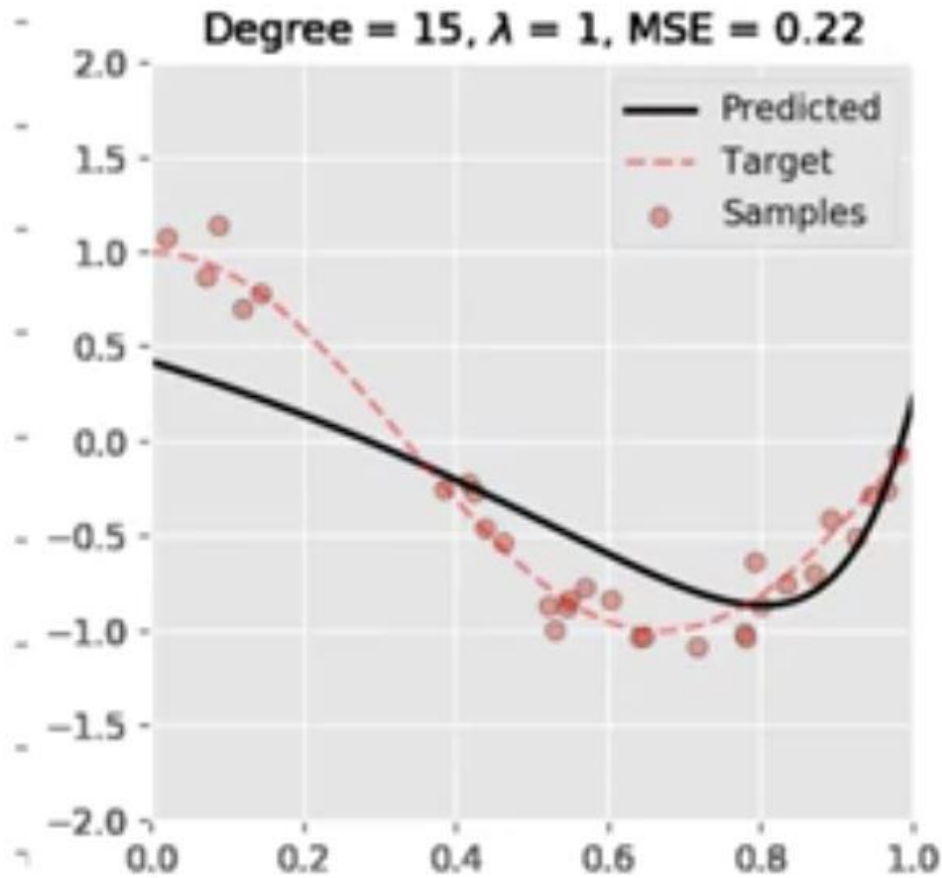
تغییرات لامبدا



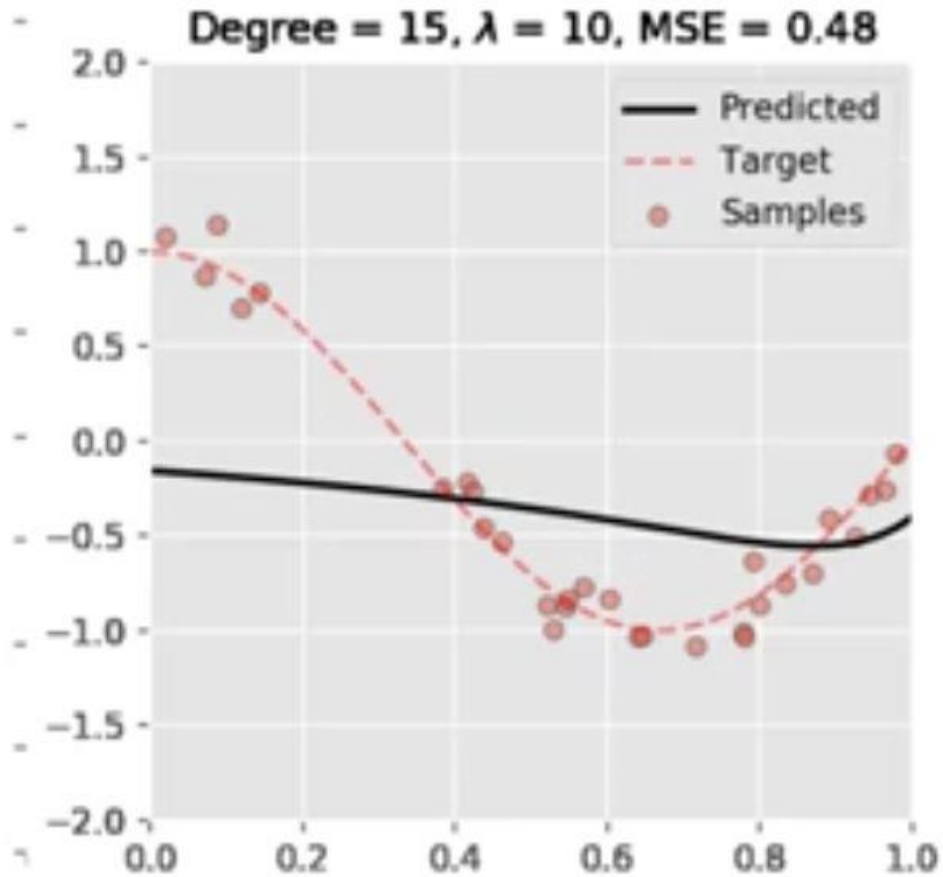
تغییرات لامبدا



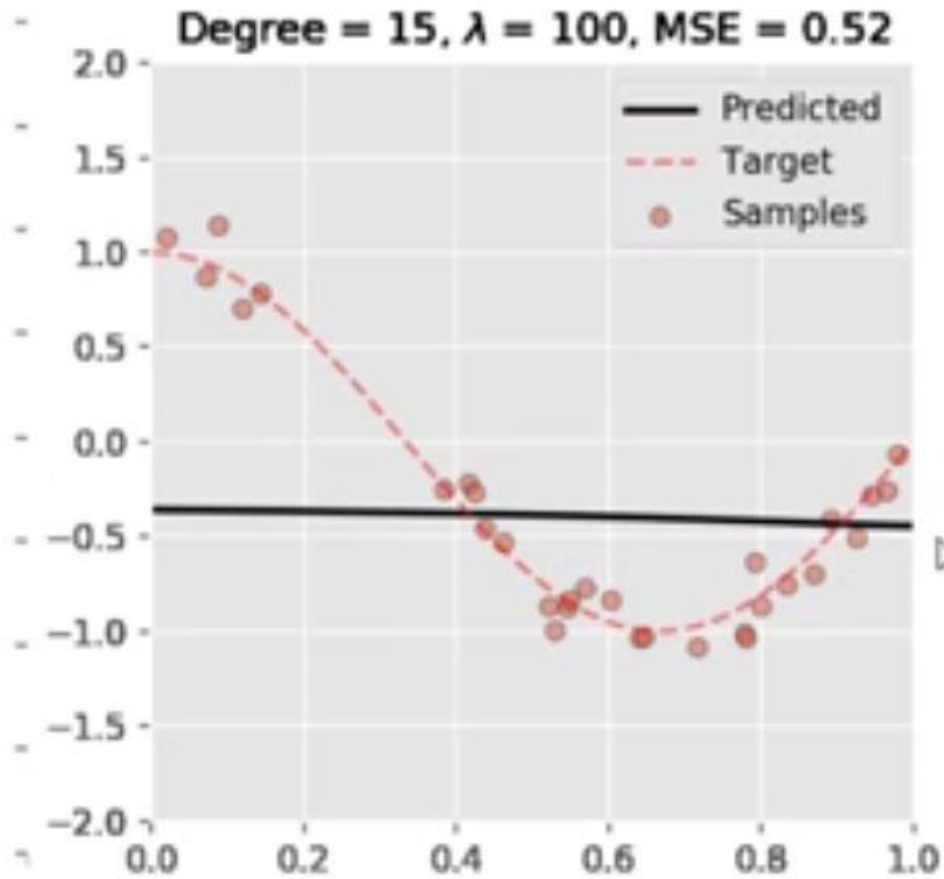
تغییرات لامبدا



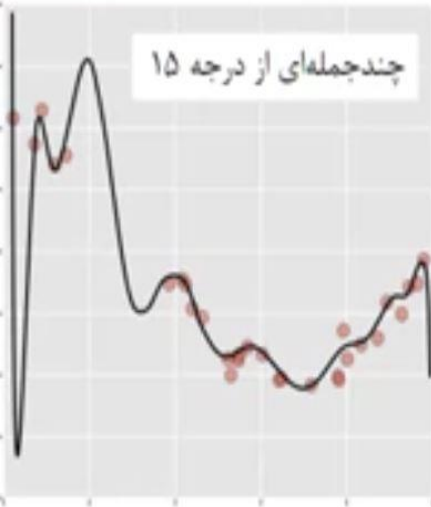
تغییرات لامبدا



تغییرات لامبدا

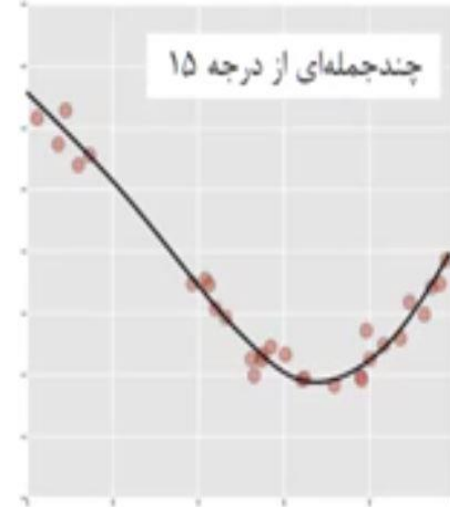


رگرسیون بدون تنظیم



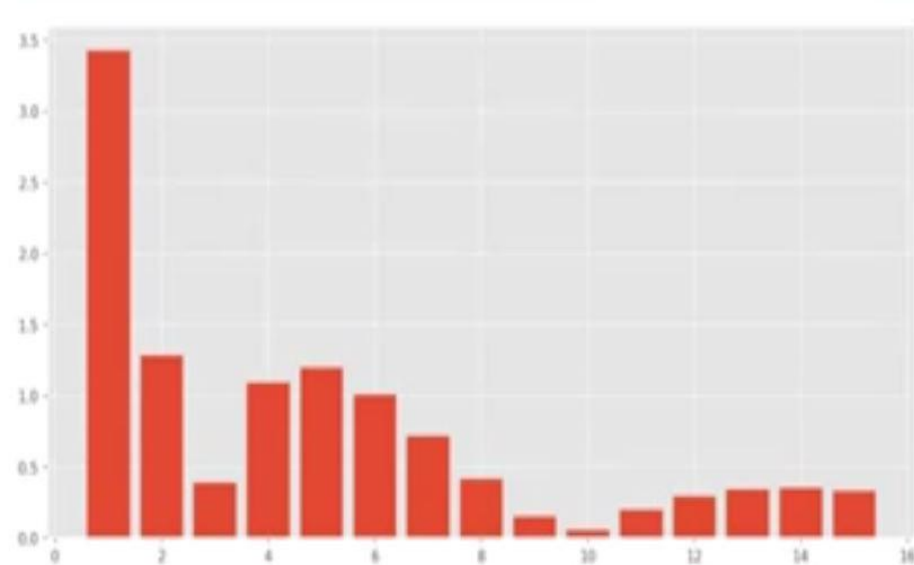
$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \text{Cost}(x^{(i)}, y^{(i)})$$

رگرسیون با تنظیم

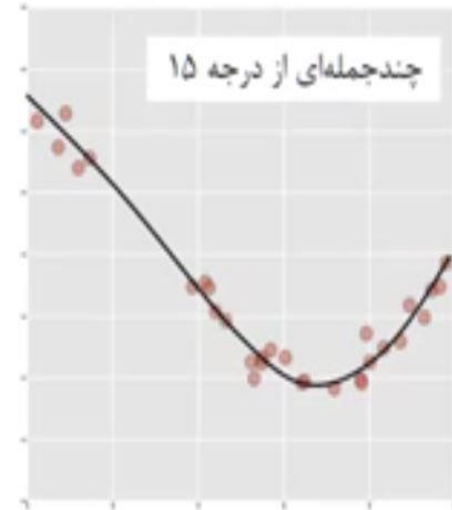


$$\frac{1}{m} \sum_{i=1}^m \text{Cost}(x^{(i)}, y^{(i)}) + \lambda R(\theta)$$

مقدار پارامترها در صورت استفاده از تنظیم



رگرسیون با تنظیم



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

Regularized logistic regression

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

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

$$\theta_j := \theta_j - \alpha \left[\frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)} + \frac{\lambda}{m} \theta_j \right]$$