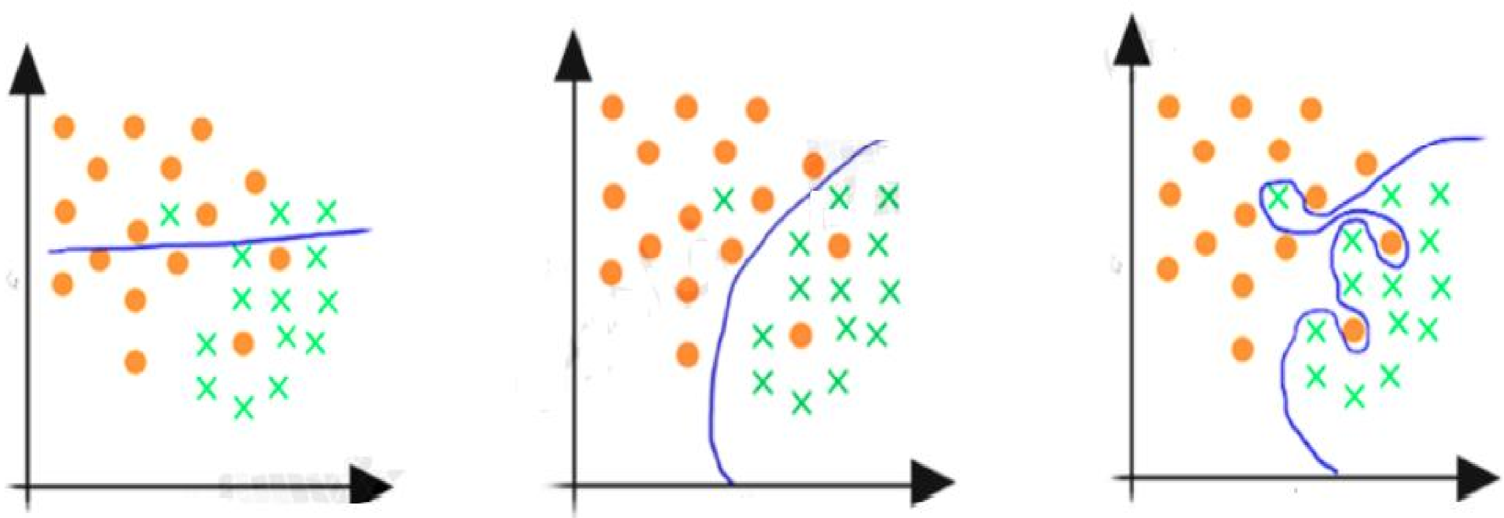
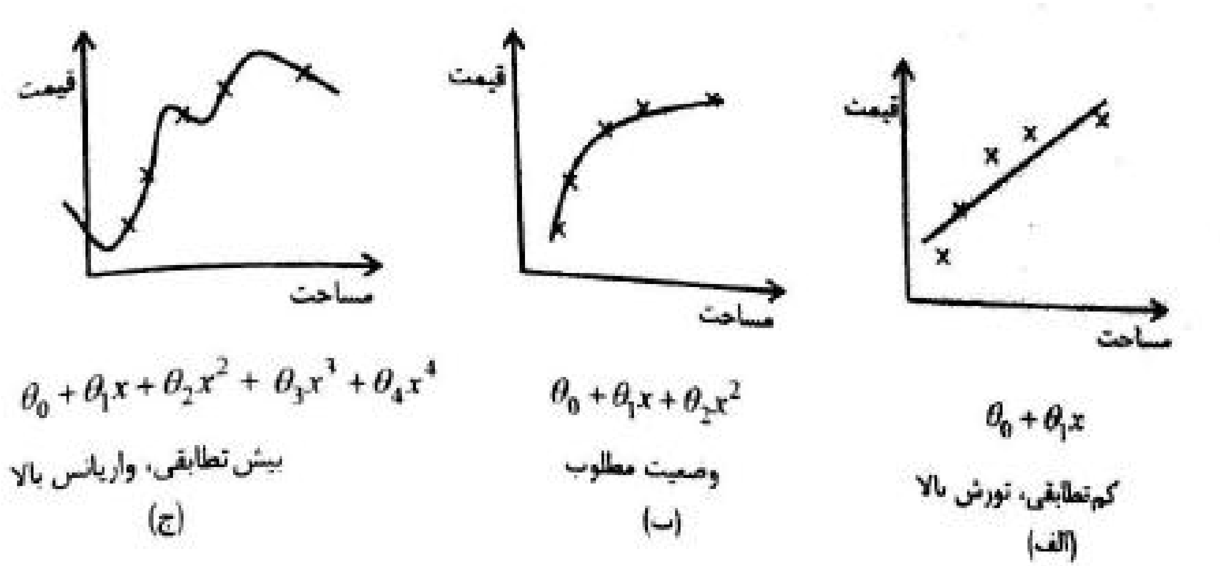
تنظیم رگرسیون

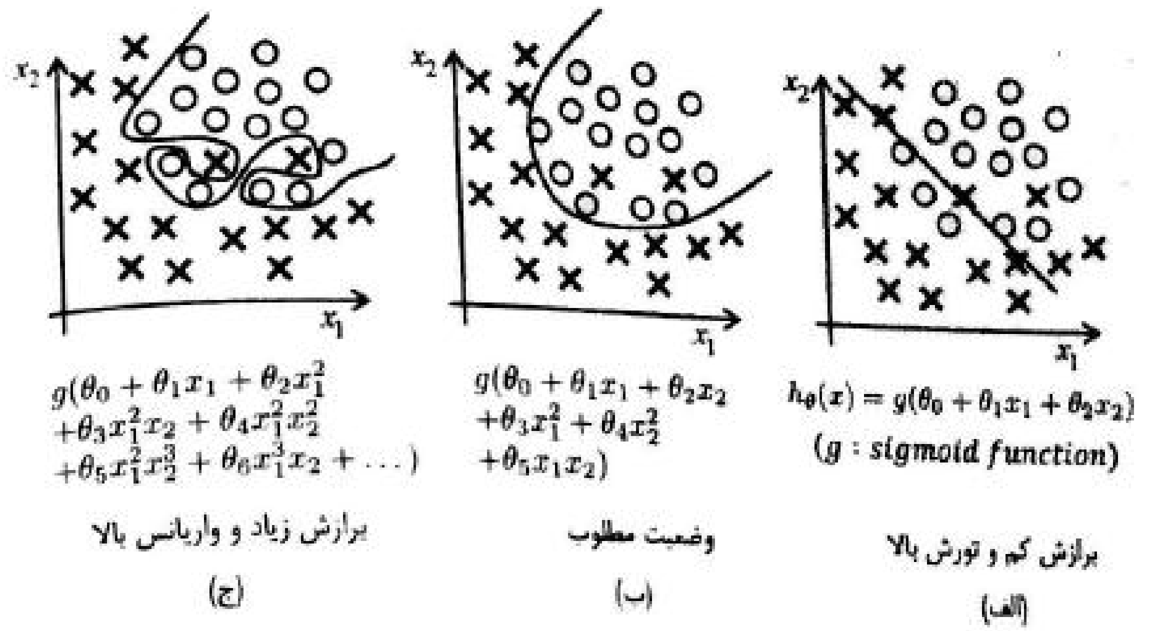


Over fitting and under fitting



بیش برازشی زمانی رخ می دهد که تعداد نمونه ها از تعداد ویژگی ها کمتر باشد.

بیش برازشی در رگرسیون لجستیک



بیش برازشی زمانی رخ می دهد که تعداد نمونه ها از تعداد ویژگی ها کمتر باشد.

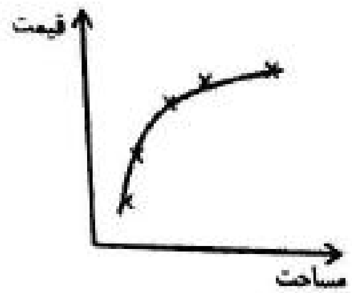
راه حل:

کاهش تعداد ویژگی ها و انتخاب درست ویژگی ها
تطبیق الگوریتم با استفاده از ویژگی های مناسب

تنظیم رگرسیون:

همه ویژگی ها حفظ شده اما اهمیت یا اندازه پارامترها θ کاهش می یابد.
زمانی مناسب است که ویژگی ها زیاد و سهم هر ویژگی یکسان و اندازه کمی در پیش بینی بازی می کند

$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_4 x^4$



$\theta_0 + \theta_1 x + \theta_2 x^2$

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 + 1000\,\theta_3^2 + 1000\,\theta_4^2$$

با این کار شکل تابع درجه ۴ به سمت تابع درجه ۲ میل می کند.

اگر همه پارامترها را جریمه کنیم سبب می شود تابع فرضیه ساده تر و تمایل به بیش تطابقی کم شود.

وقتی تعداد ویژگی ها زیاد است و نمی دانیم که کدام ویژگی باعث بیش تطابقی می شود که جریمه بیشتری شود یا حذف گردد لذا در این موارد تمام تابع هزینه به نحوی تغییر می کند که تمام پارامترها کوچک شوند.

اضافه کردن عبارت تنظیم رگرسیون به تابع هزینه قبلی

$$\lambda \sum_{j=1}^{n} \theta_j^2$$

λ ضریب تنظیم است

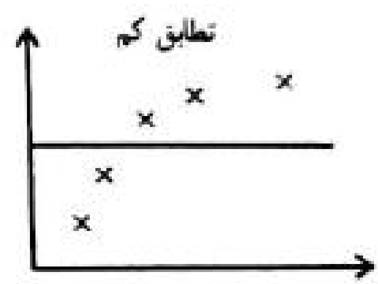
λ باید بین بخش اول که بهترین برازش و بخش دوم که کوچک نگه دانشتن مقادیر پارامترهاست تعادل برقرار نماید.

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

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

λ نزدیک به صفر بیش تطابقی و اگر بسیار بزرگ باشد همه θ ها صفر می شود و فقط θ0 باقی می ماند

$$\square(x) = \theta_0 x_0 + \theta_1 x_1 + \ldots\ldots + \theta_n x_n$$



نزول گرادیان در رگرسیون تنظیم شده

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

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

Repeat until convergence{

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

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

}

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

$$\theta_j = \theta_j(1 - \alpha\frac{\lambda}{m}) - \frac{\alpha}{m}\sum_{i=1}^{m} h_\theta(x^i) - y^i)x_j^i \quad j = 1, 2, \dots, n)$$

تفاوت نزول گرادیان تنظیم شده با الگوریتم گرادیان معمولی اضافه شدن عبارت $(1 - \alpha\frac{\lambda}{m})$ است.

این عبارت تنظیم کننده رگرسیون است.

معمولاً $(1 - \alpha\frac{\lambda}{m})<1$ است.

حل تحلیلی نزول گرادیان اصلاح شده

$$X_{m*(n+1)} = \begin{matrix} x^{1T} \\ x^{2T} \\ x^{3T} \\ . \\ x^{mT} \end{matrix} \qquad y_{m*1} = \begin{matrix} y^{1} \\ y^{2} \\ y^{3} \\ . \\ y^{m} \end{matrix} \qquad \theta_{(n+1)*1} = \begin{matrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ . \\ \theta_n \end{matrix} \qquad X_{(n+1)*1} = \begin{matrix} x_0 \\ x_1 \\ x_2 \\ . \\ x_n \end{matrix}$$

$$\theta = (X^T X)^{-1} X^T y$$

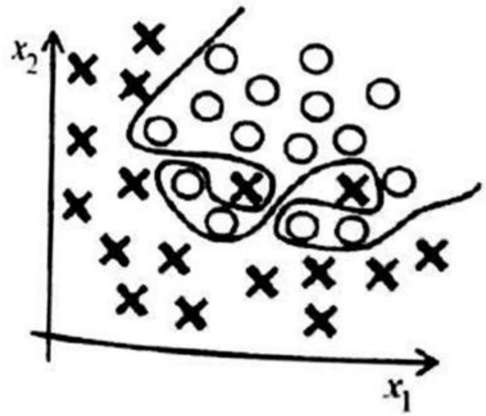
ماتریس $X^T X$ همیشه دارای معکوس است.

$$\theta = \left(X^T X + \lambda \begin{bmatrix} 0 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & \ddots & \\ & & & & & 1 \end{bmatrix}_{(n+1 \times n+1)} \right)^{-1} X^T y$$

مثال: n=2

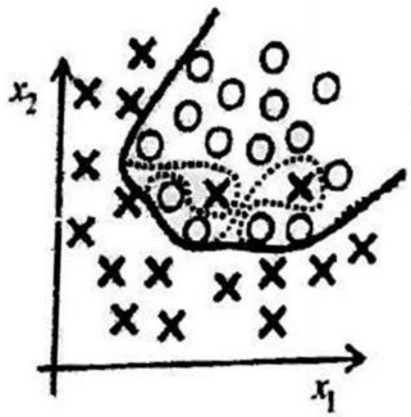
$$\theta = \left(X^T X + \lambda \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)^{-1} X^T y$$

تنظیم رگرسیون لجستیک



$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \theta_6 x_1^3 x_2^4 + \ldots)$$

(g: sigmoid function)



$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \theta_6 x_1^3 x_2 + \ldots)$$

(g: sigmoid function)

$$Cost(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & y = 1 \\ -\log(1 - h_\theta(x)) & y = 0 \end{cases}$$

$$Cost(h_\theta(x), y) = -y\log(h_\theta(x)) - (1-y)\log(1 - h_\theta(x))$$

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

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

نزول گرادیان در رگرسیون لجستیک تنظیم شده

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

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

Repeat until convergence{

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

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

}