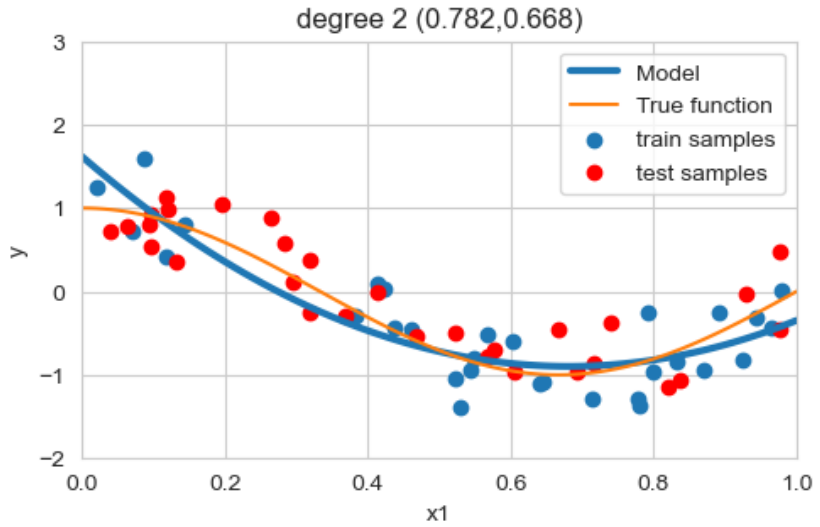
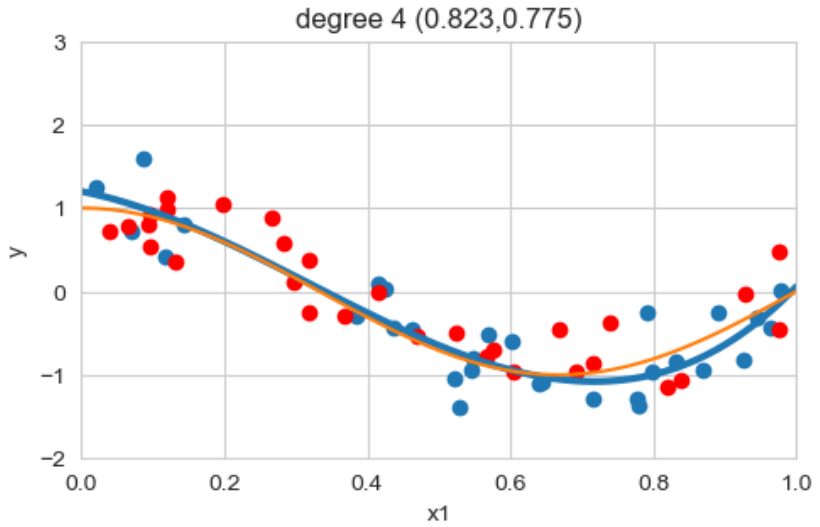
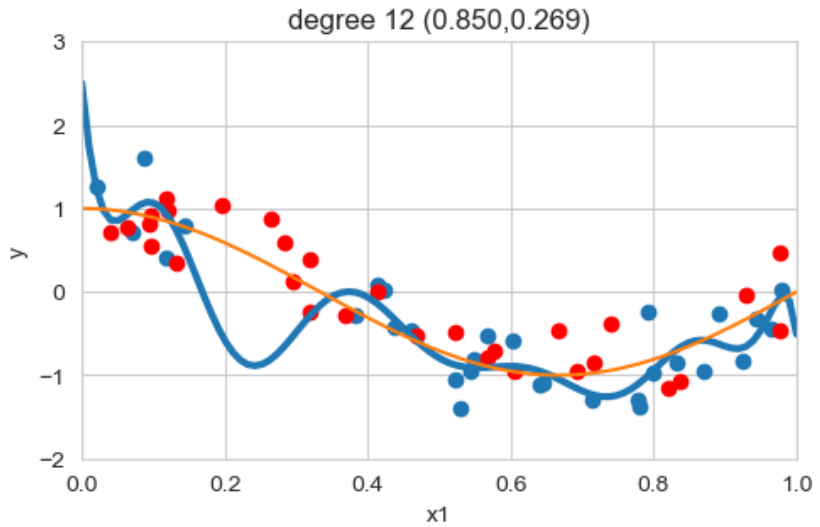
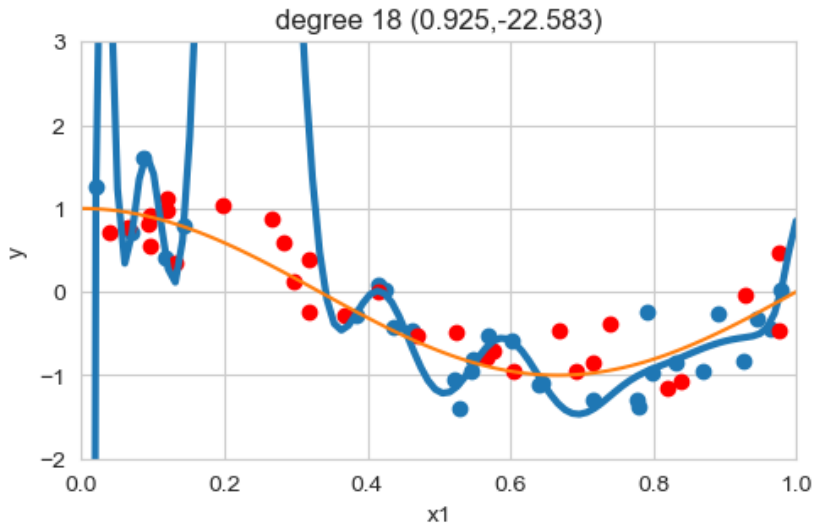
model regularization









label noise

feature noise

feature relevance

relatively small train sets

don't try to fit the data perfectly

don't try to use all features

notice how R^2 on the train set does increase with *d*

regularized linear regression

$$f(x, \theta) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_m x_m$$

cost function

$$J(\theta) = \frac{1}{2n} \sum_{i=1}^{n} (f(x^{(i)}, \theta) - y^{(i)})^2$$

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

regularized cost function

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

$$\theta_j := \theta_j - \alpha \frac{1}{n} \sum_{i=1}^{n} (f(x^{(i)}, \theta) - y^{(i)}) x_j^{(i)} - \frac{\lambda}{n} \theta_j$$

regularized logistic regression

$$f(x,\theta) = g(\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_m x_m)$$

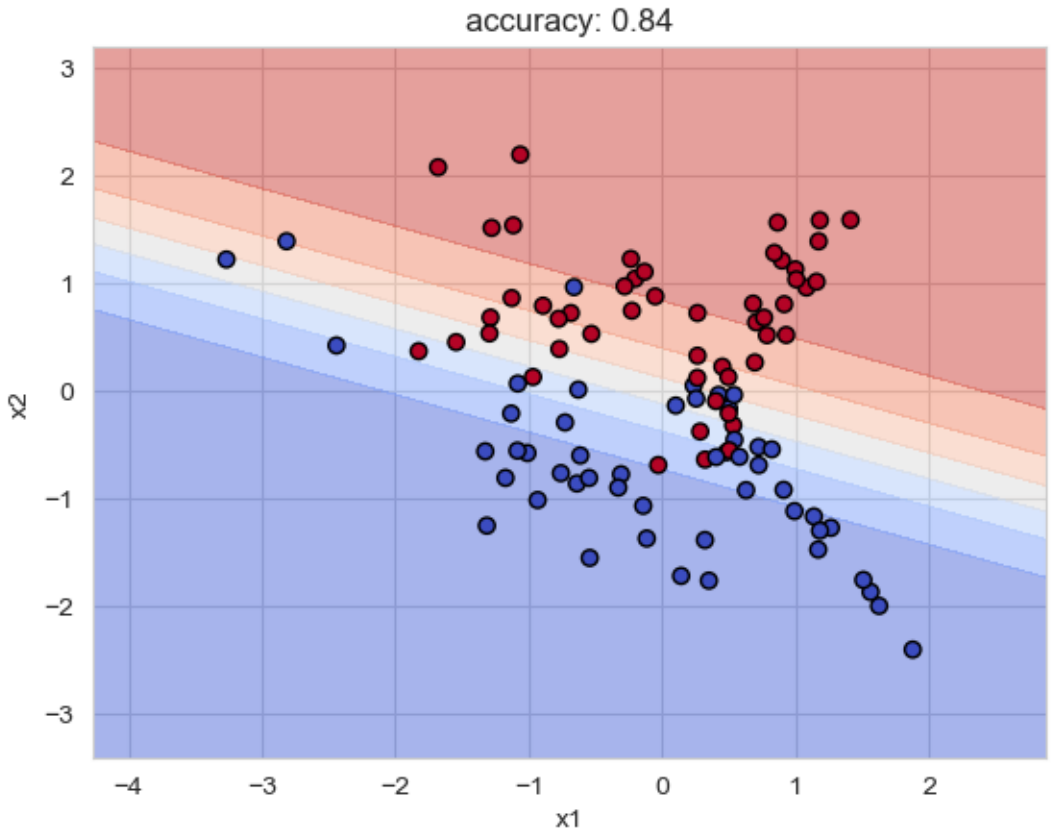
cost function

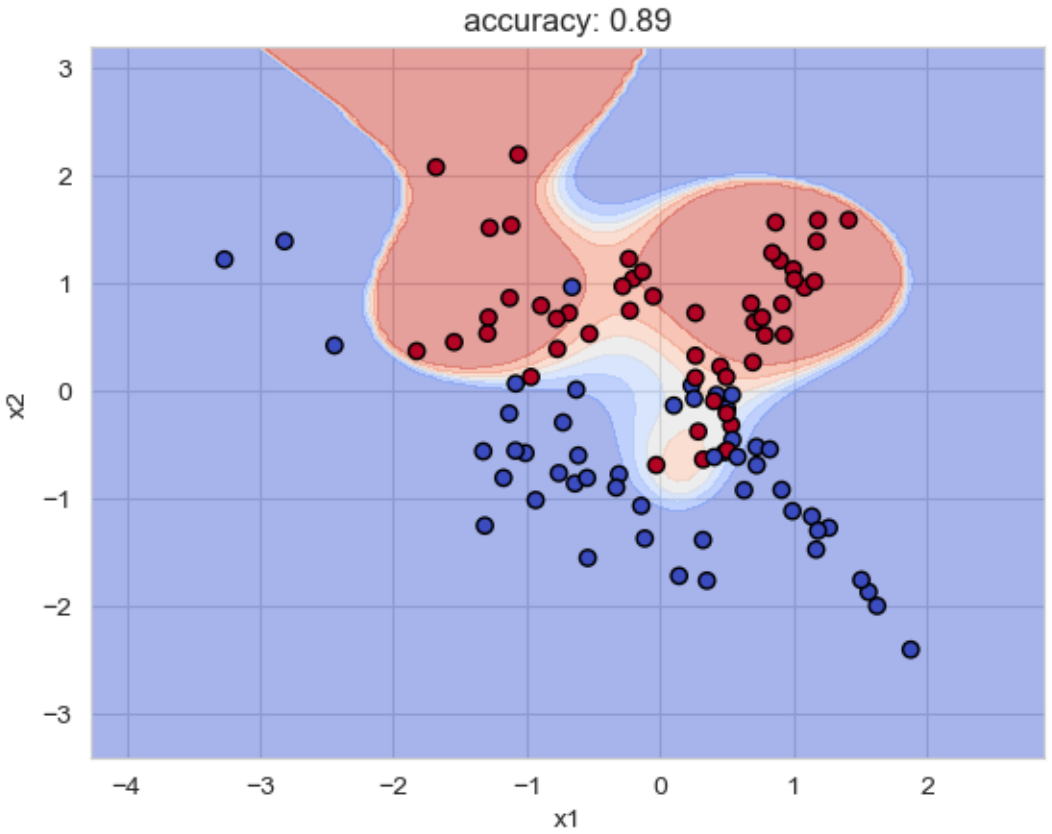
$$J(\theta) = -[\frac{1}{n}\sum_{i=1}^{n} y^{(i)} log(f(x^{(i)},\theta)) + (1 - y^{(i)}) log(1 - f(x^{(i)},\theta))]$$

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

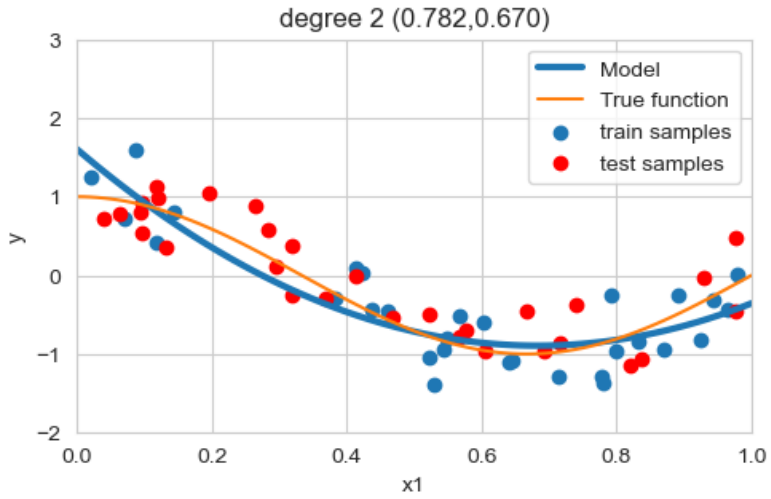
regularized cost function

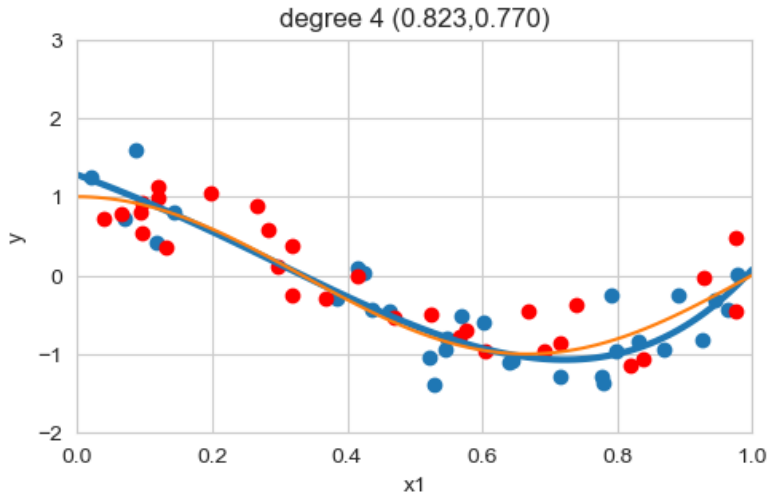
model regularization

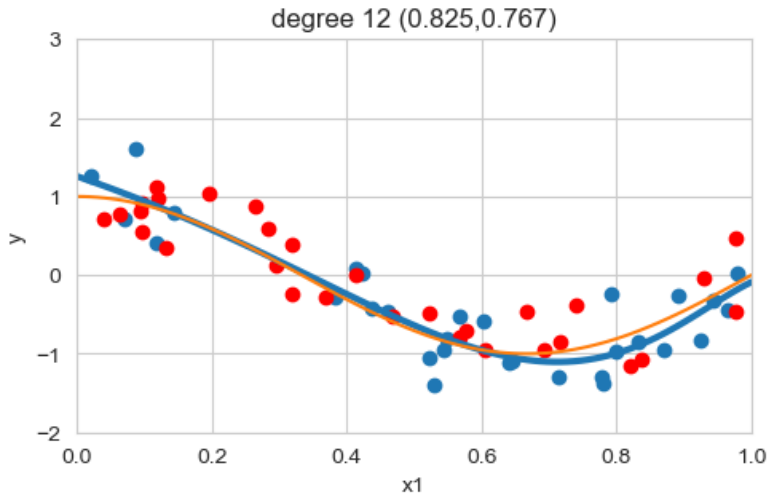


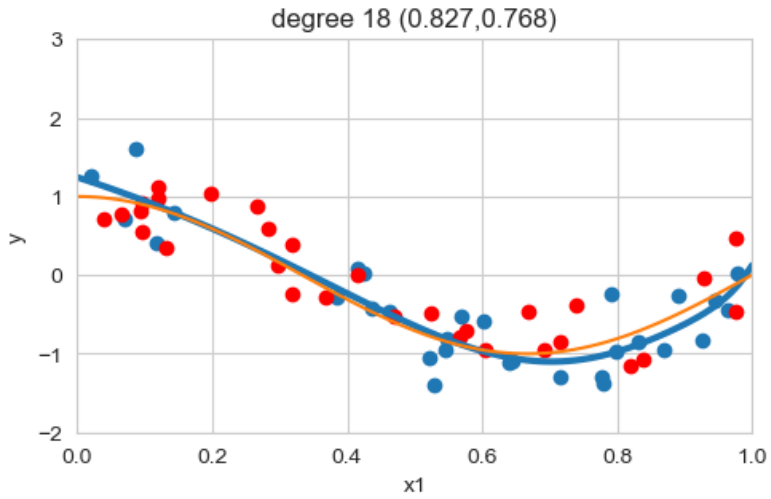


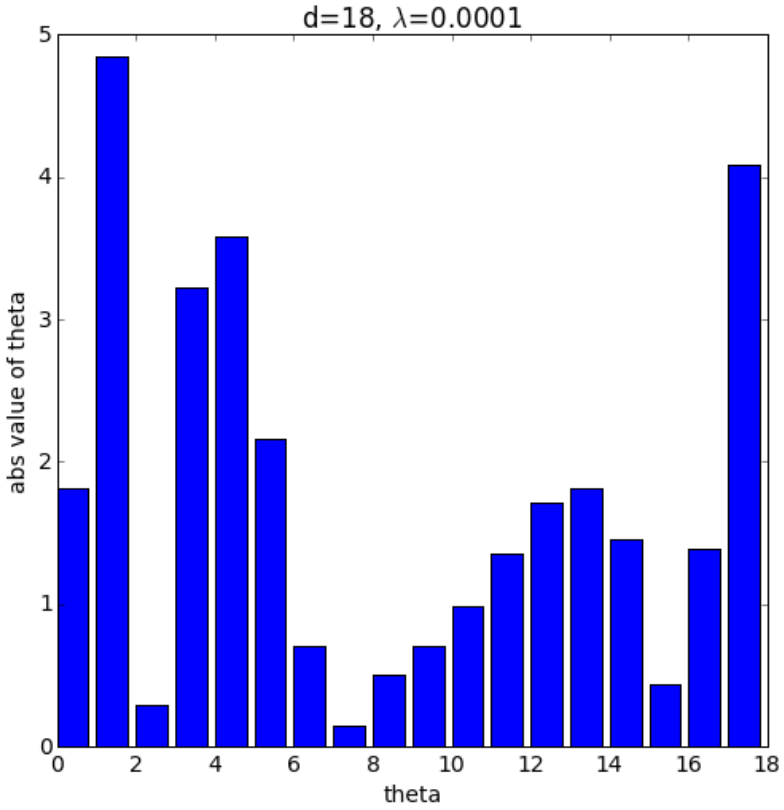
regularized linear regression



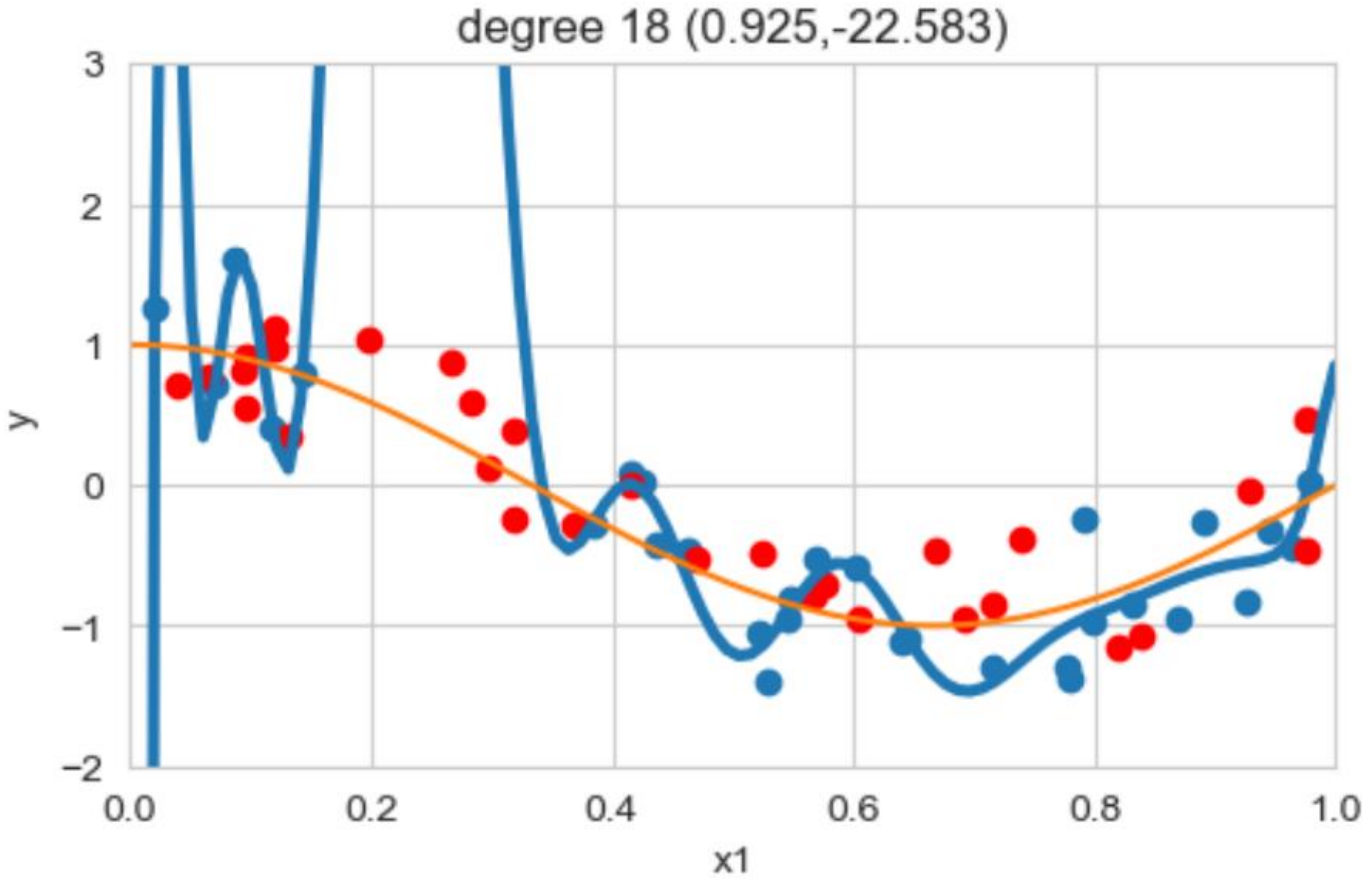


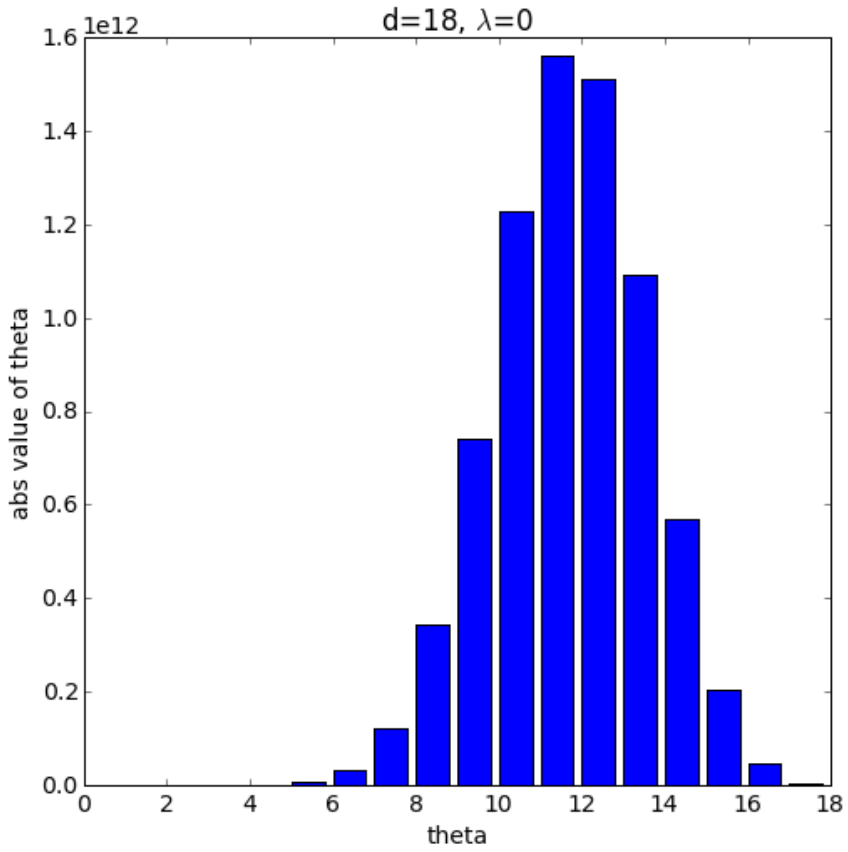






regularized linear regression





regularized linear regression