Normal Equation

Normal equation

Formula

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

Normal equation

vs GD

The following is a comparison of gradient descent and the normal equation:

Gradient Descent	Normal Equation
Need to choose alpha	No need to choose alpha
Needs many iterations	No need to iterate
O (kn^2)	O (n^3), need to calculate inverse of $X^T X$
Works well when n is large	Slow if n is very large

With the normal equation, computing the inversion has complexity $\mathcal{O}(n^3)$. So if we have a very large number of features, the normal equation will be slow.

Over Fitting & Regularization

OF & Regularization

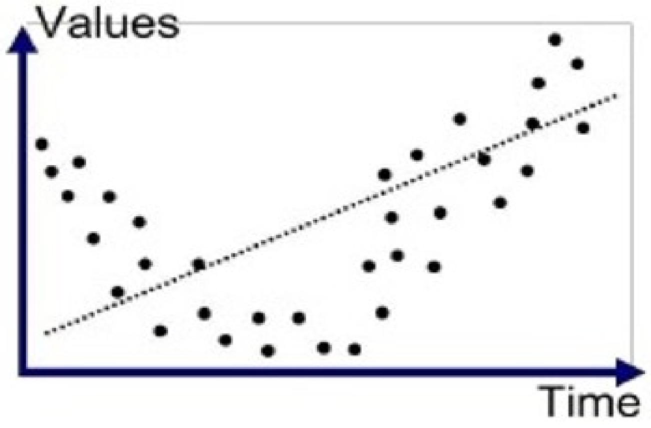
Outline

- Underfitting
- Overfitting
- Regularization

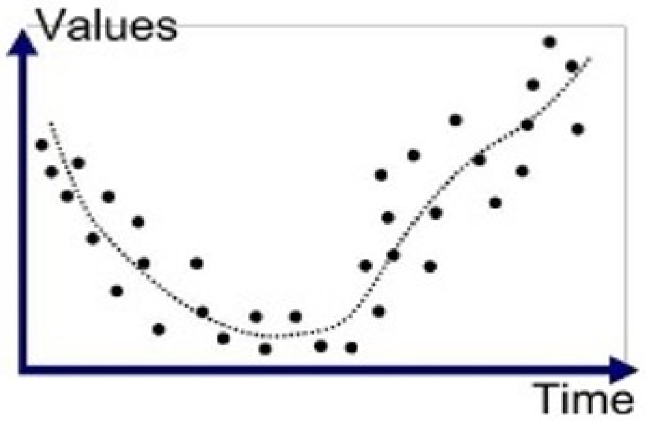
Underfitting

A statistical model or a machine learning algorithm is said to have underfitting when it cannot capture the underlying trend of the data.
It is also called a high bias problem.

An example of an underfitting problem.



Underfitted

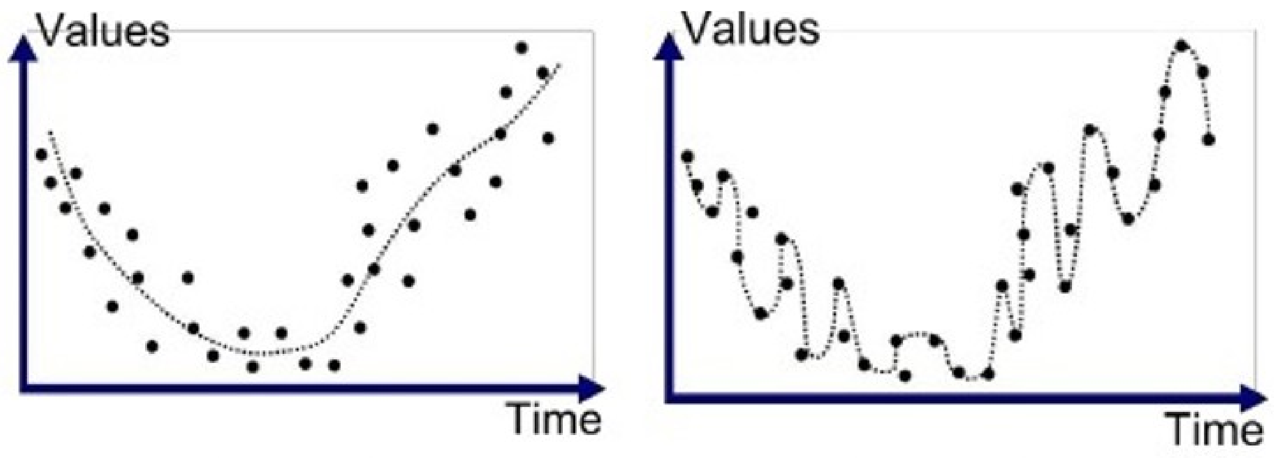


Good Fit/Robust

Overfitting

Overfitting happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data. This means that the noise or random fluctuations in the training data is picked up and learned as concepts by the model.

The model in this case fits the training data so well but fails fitting the testing data.

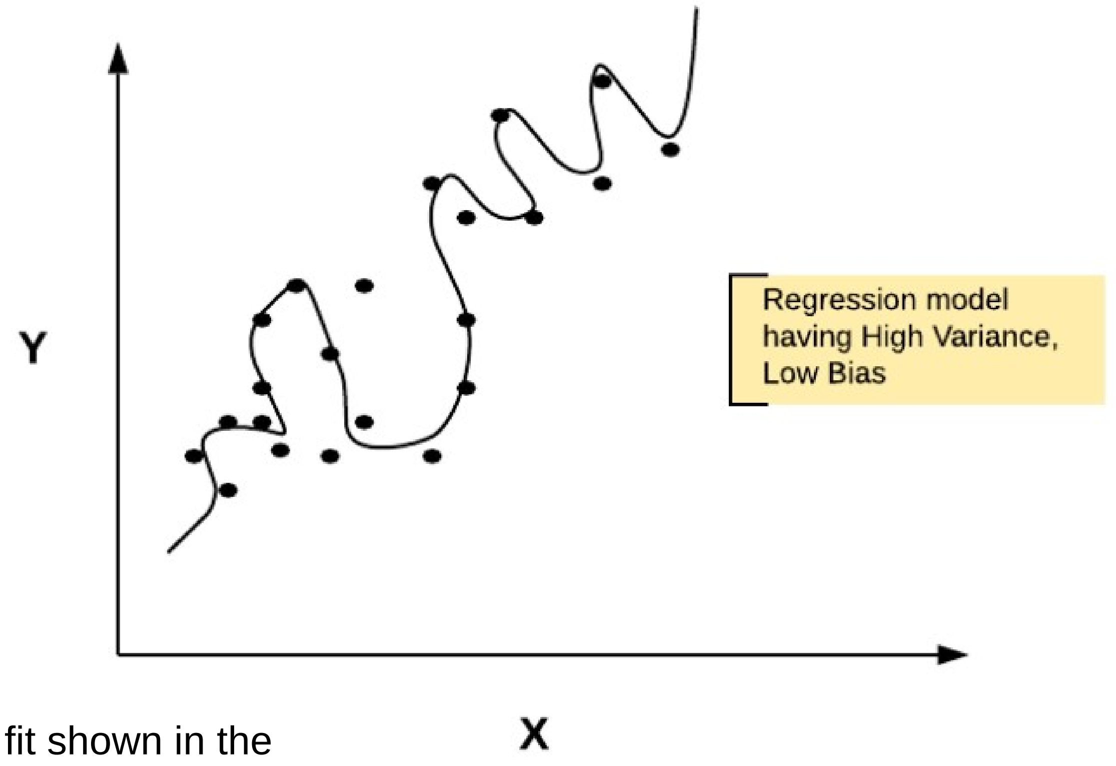


Good Fit/Robust

Overfitted

Overfitting

An overfitted model is so called to have high variance.

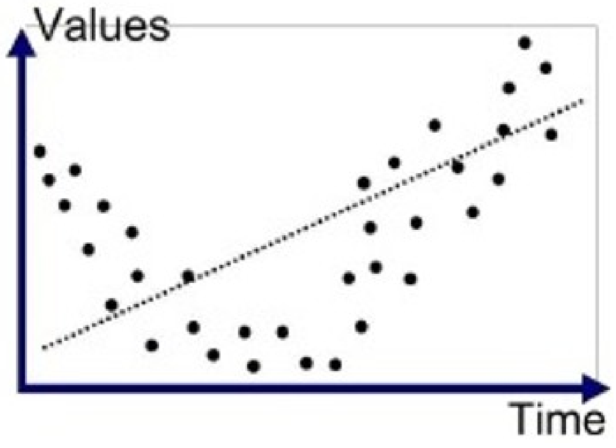


In such a model , a linear fit would perform finer than the fit shown in the graph using high order polynomial terms with such a noise or variance in the training set generates a high variance overfitted model.

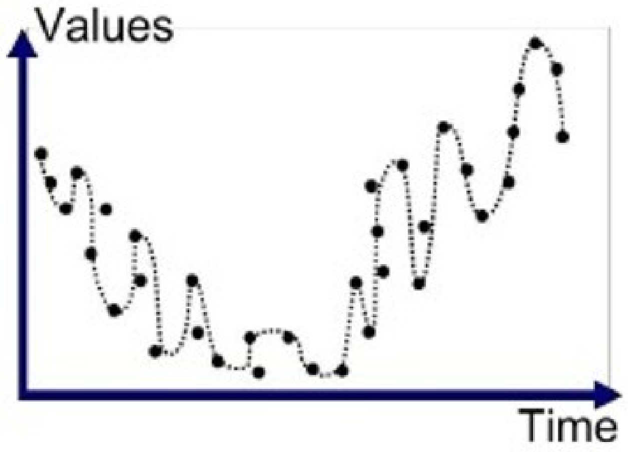
Regularization needed for reducing overfitting in the regression model

Fitting

Overfitting: Good performance on the training data, poor generliazation to other data.
Underfitting: Poor performance on the training data and poor generalization to other data



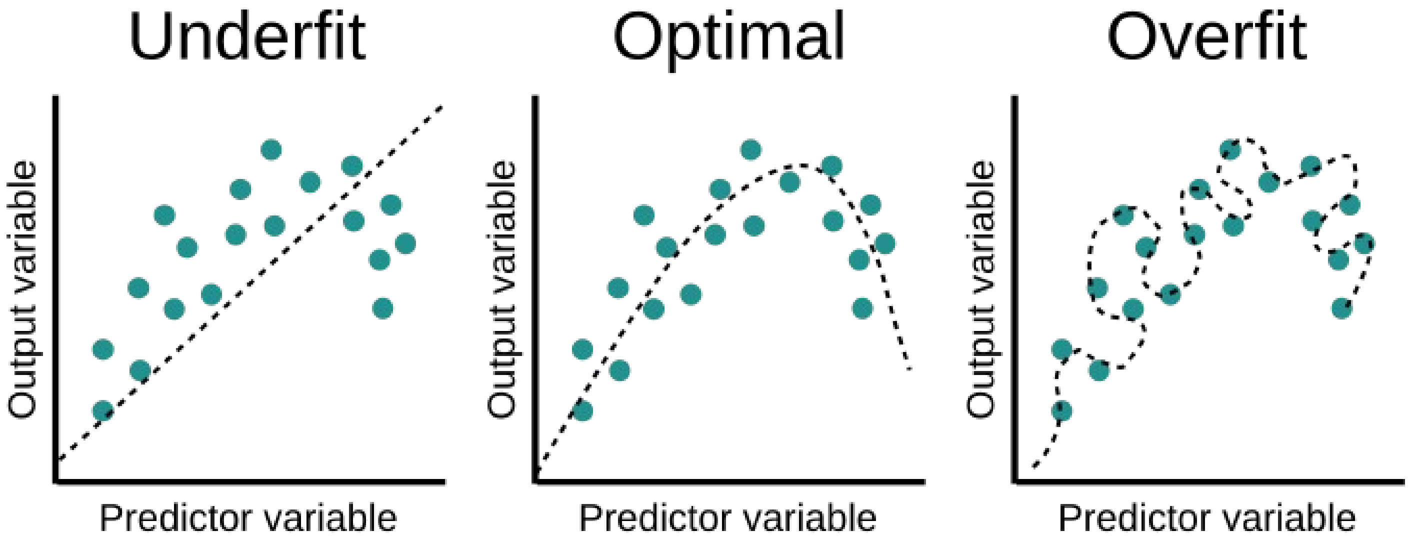
Underfitted



Overfitted

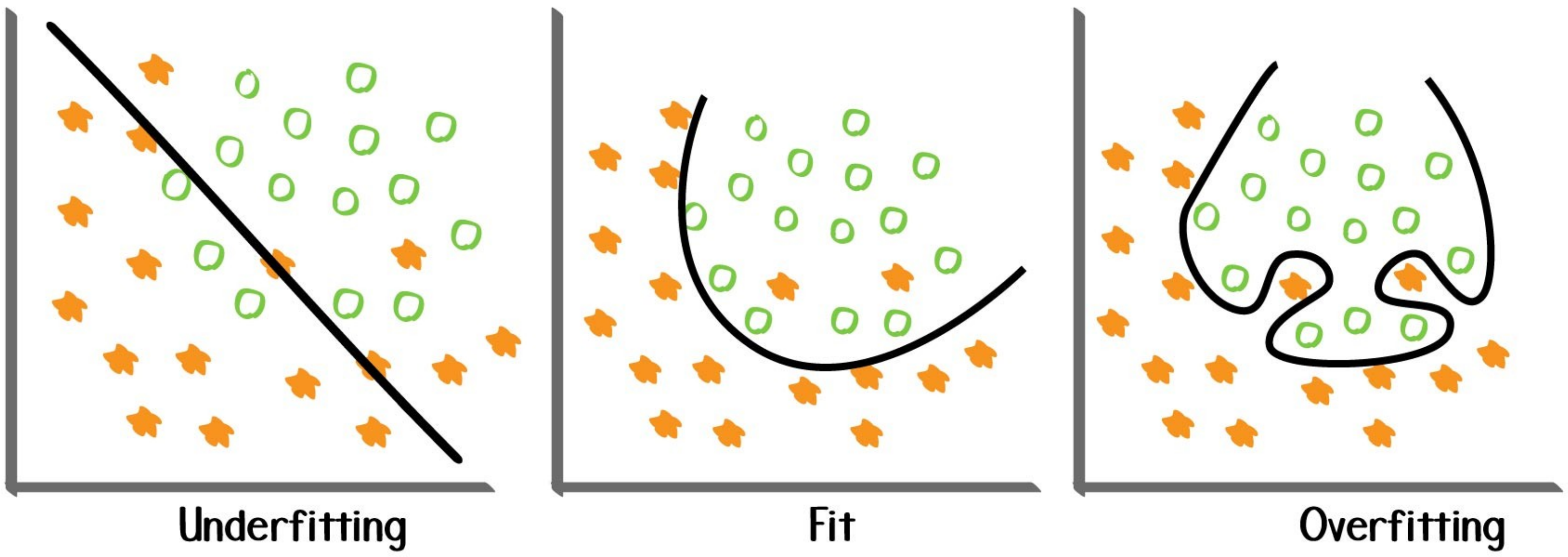
Optimal model

In the example shown below, a linear regression model might underfit the data and a higher order model might get severely affected by the noise and variance in the dataset, an optimal fit would be in between.



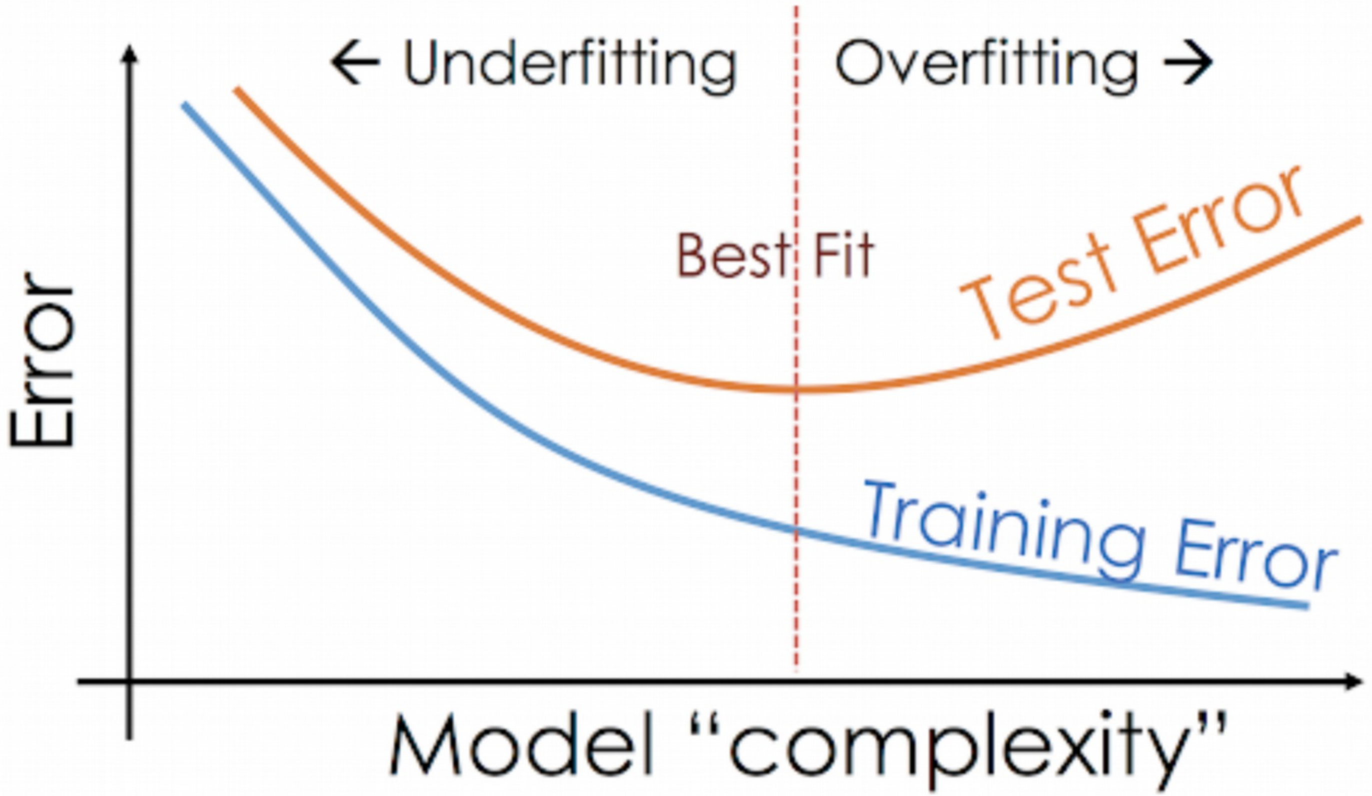
Optimal model

The same applies for logistic regression.



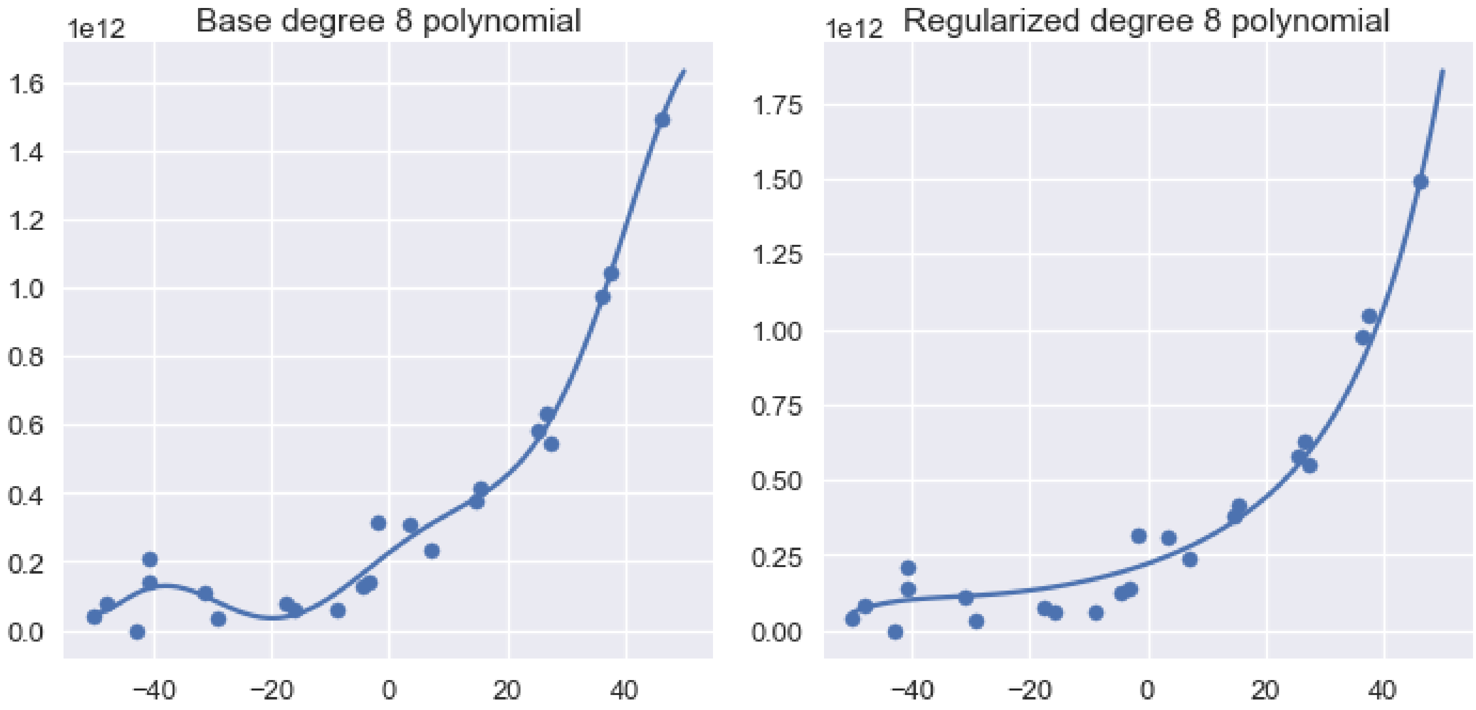
Optimal model

Model complexity
vs
Test & Training Errors.



Regularization

Regularization is a technique used to reduce the error (testing error) by fitting a function appropriately on the given training set and avoid overfitting.



Regularization

In regularization, a simple term added in the cost function whether it was a linear regression or a logistic regression, it helps minimizing the values of the parameters while minimizing the error value, thus penalizing the parameters that try to fit the model as neatly as possible and reduce the overfitting problem.

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

Regularization

Thus, the updated function changes to the following.

Regularization Term

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

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

Regularization

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

Goodbye