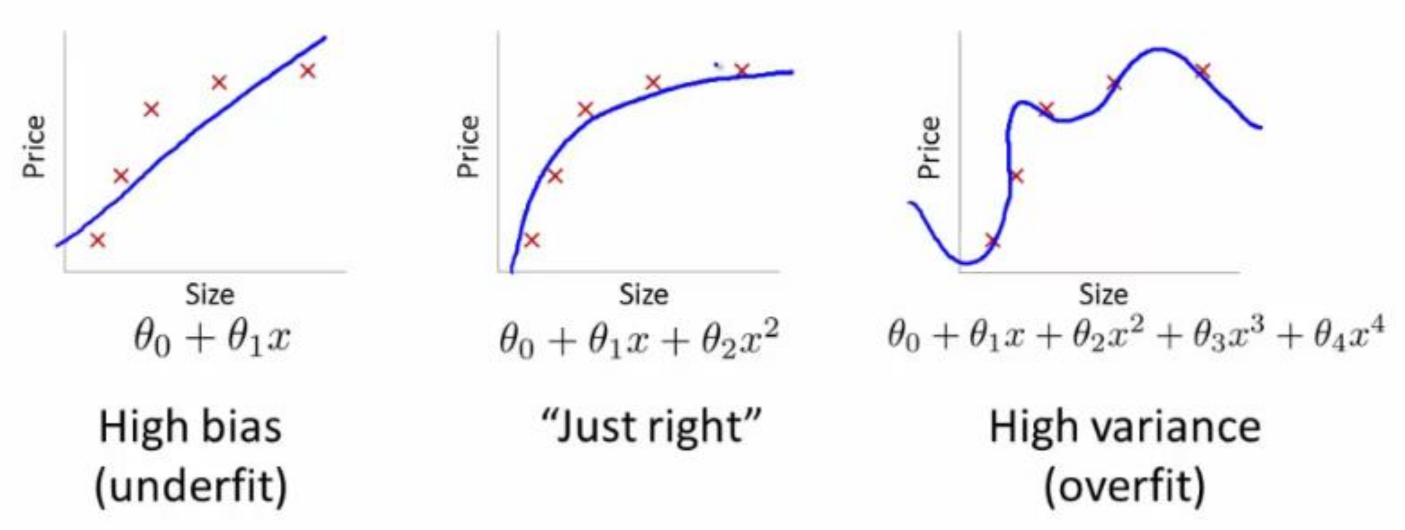
Regularized Regression Models

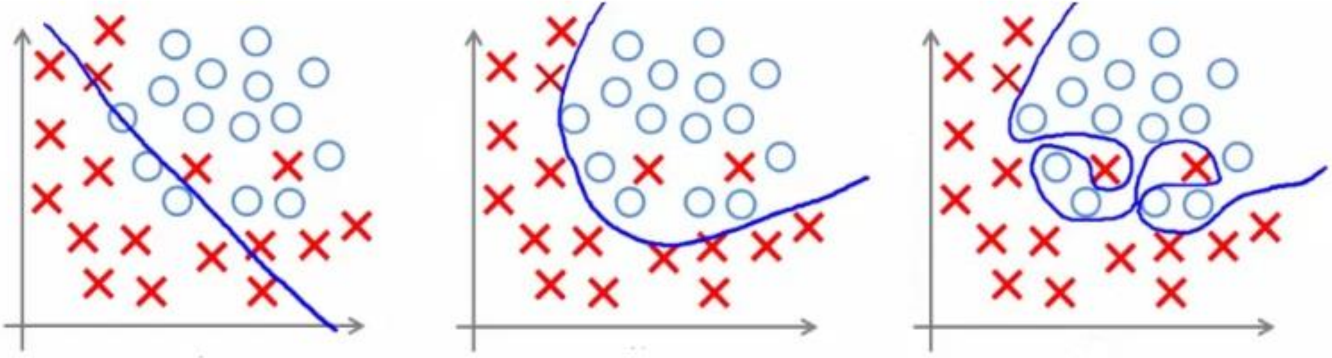
Data Science Dojo

datasciencedojo
unleash the data scientist in you

Linear Regression Fitting Example



Logistic Regression Fitting Example



Under-fitting	Appropriate-fitting	Over-fitting
(too simple to explain the variance)		(forcefitting -- too good to be true)

Overfitting

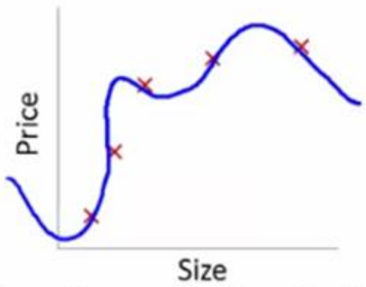
- Overfitting when:
  - Complex model, too many features, not enough training samples
- How to address overfitting
  - Go through each feature to decide which to keep
  - Use model selection algorithm to automatically choose features

datasciencedojo
unleash the data scientist in you

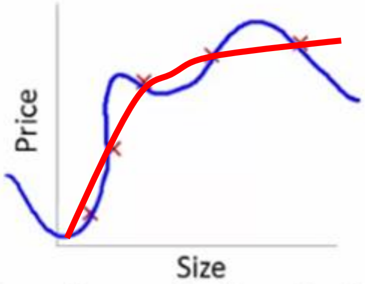
Idea of Regularization

- Keep all the features, but reduce their magnitude of parameter effects in model
- Shrink θ_j parameters

Regularized regression intuition



$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$



$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

Close to zero

- Goal: To minimize cost function θ_j

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

- Suppose we penalize and make θ_3 *and* θ_4 very small

datasciencedojo
unleash the data scientist in you

Regularization

- Two common types of regularization in linear regression
  - L2 regularization (a.k.a ridge regression)

$$\sum_{j=1}^{N}(y_j - \sum_{i=0}^{d}\theta_i \cdot x_i)^2 + \lambda\sum_{i=1}^{d}\theta_i^2$$

  - L1 regularization (a.k.a lasso regression)

$$\sum_{j=1}^{N}(y_j - \sum_{i=0}^{d}\theta_i \cdot x_i)^2 + \lambda\sum_{i=1}^{d}|\theta_i|$$

Regularized-Ridge Regression

- Regularization by shrink θ_j smaller values, as a result
  - "less complex" hypothesis function without eliminating features
  - More protection from overfitting
- L2: Ridge regression

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

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

datasciencedojo
unleash the data scientist in you

Regularized-Ridge Regression

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

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

- Goal 1: find the best fit
- Goal 2: keep parameter θj small
- λ is regularization parameter to controls a trade off

Regularized-Ridge Regression

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

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

- If λ is too large, θj become too small, as if features have no effect in predicting response.
- If λ is too small, θj are not regularized.

QUESTIONS