Regularized linear . . .
Gradient descent
Regularized logistic . . .
Gradient descent

ML- MACHINE LEARNING

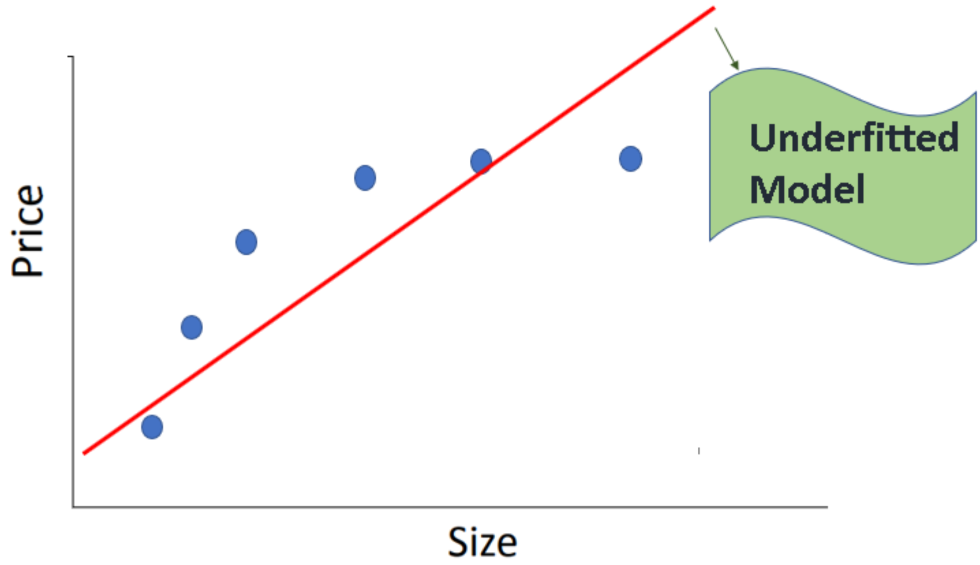
XI. Regularization

FERNEY BELTRAN
MIGUEL VILLAMIL
SANTIAGO CARVAJAL

Overfitting

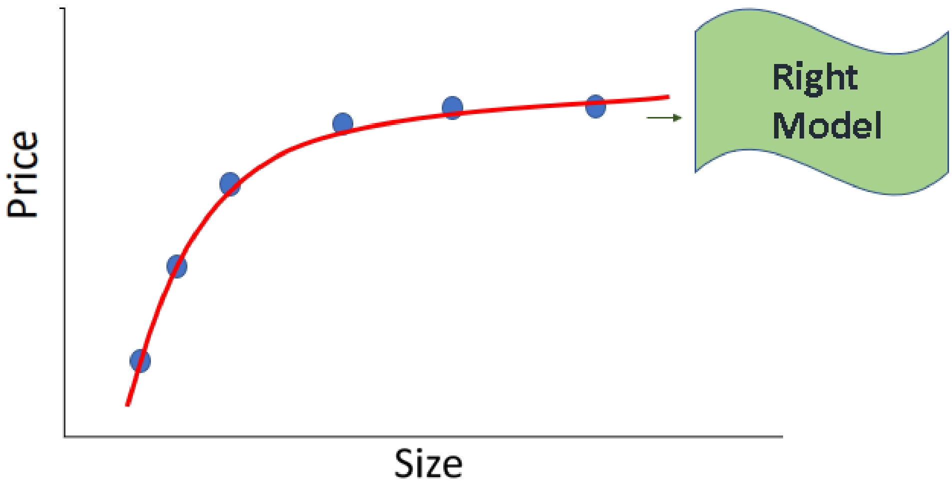
Overfitting is the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably. [OxfordDictionaries]

Housing Prices Example



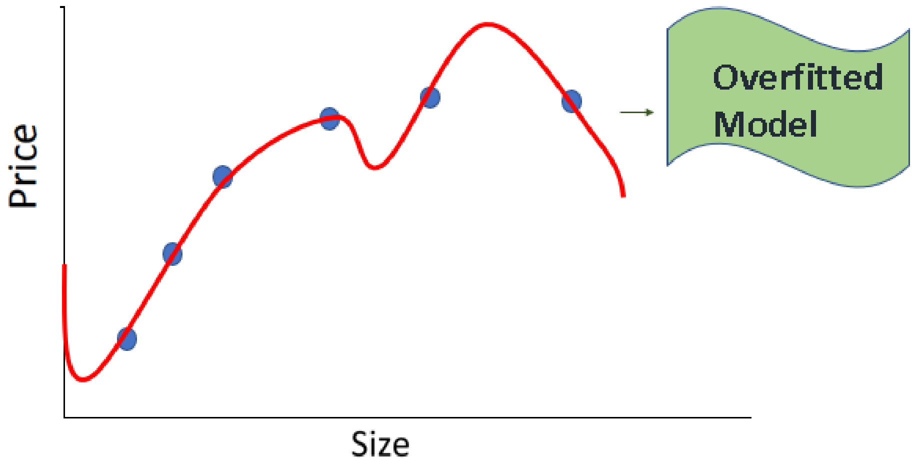
$$\theta_0 + \theta_1 x$$

Housing Prices Example



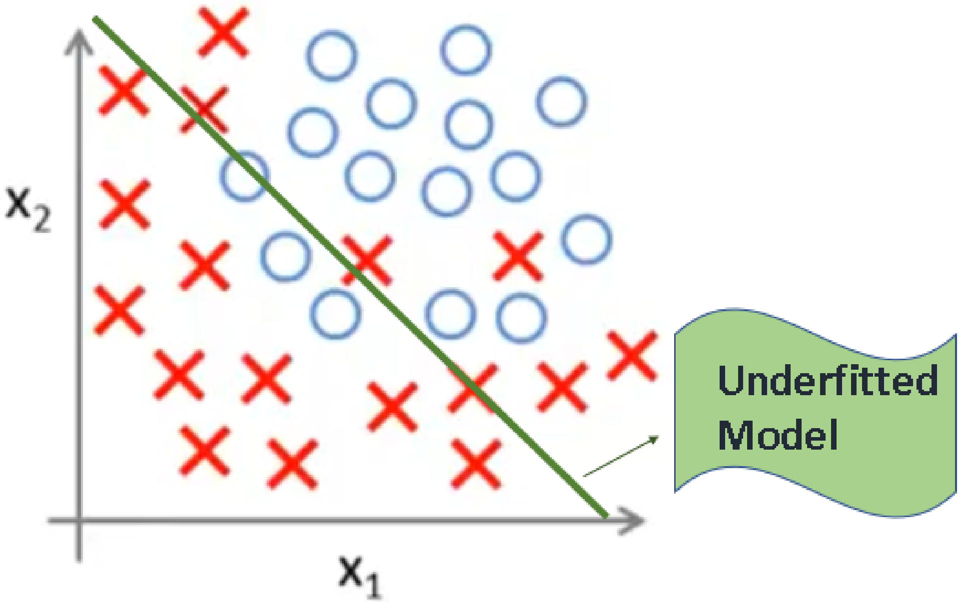
$$\theta_0 + \theta_1 x + \theta_2 x^2$$

Housing Prices Example

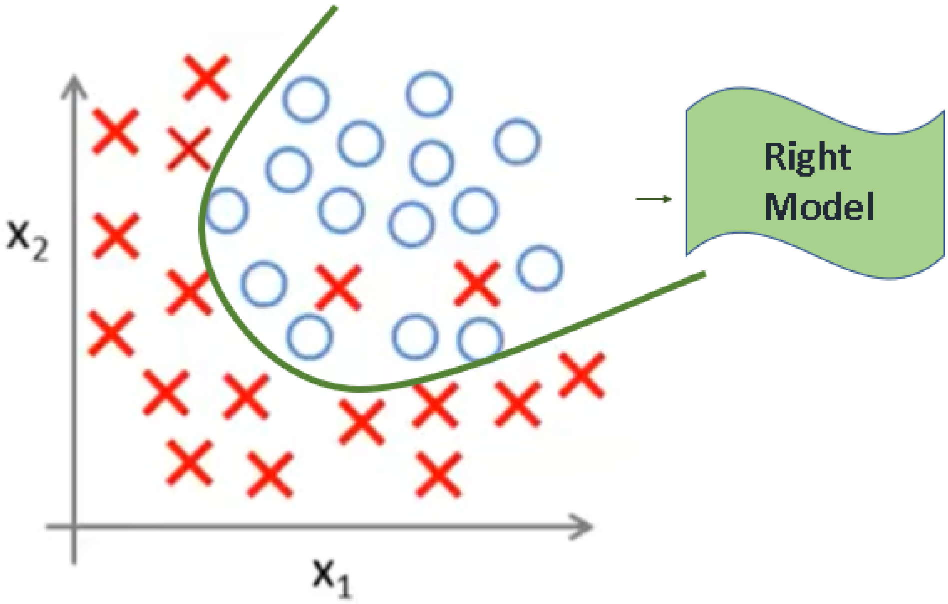


$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

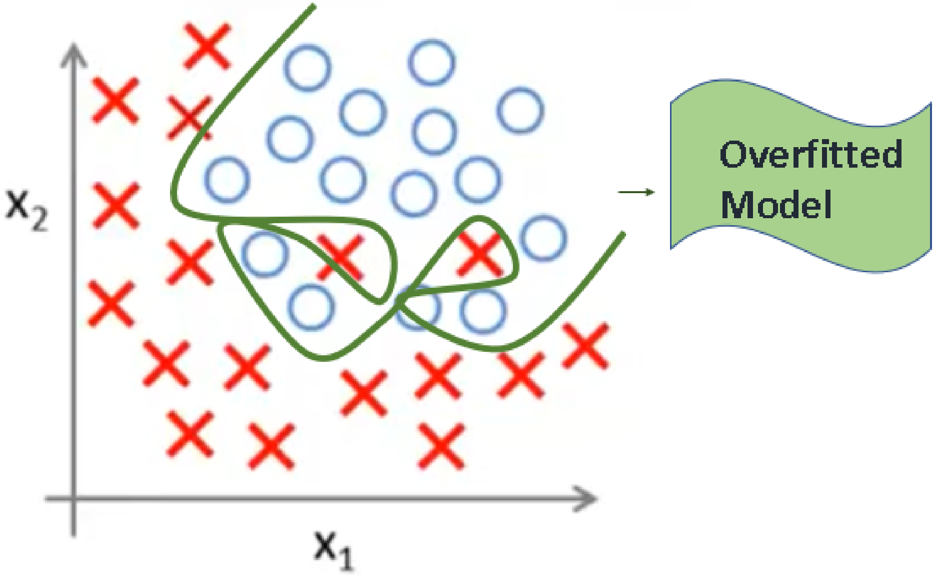
Classification Example



Classification Example



Classification Example



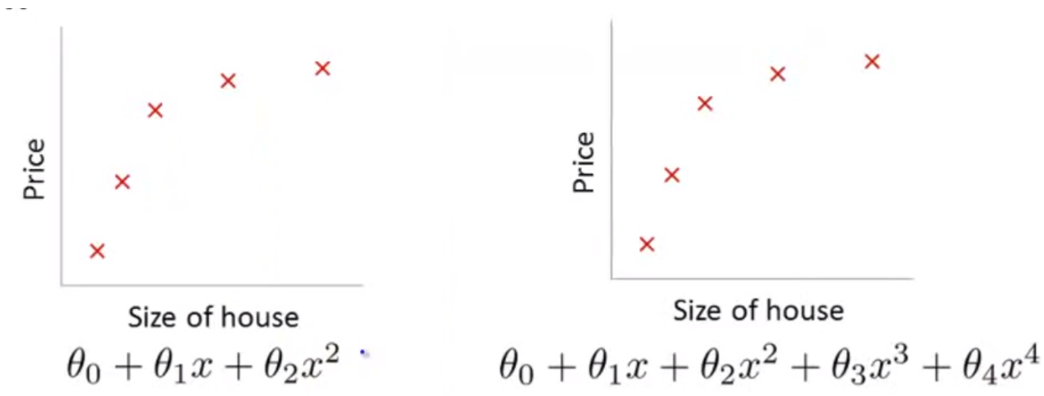
Deal with Overfitting:

Two Options:

1. Reduce number of features
2. Regularization
   -Keep all the features
   -Reduce magnitude or values of parameters

Cost Function

If we have over fitting in our hypothesis function we can reduce the weight for some terms in our function by increasing the cost.



Suppose we penalize and make θ_3, θ_4 really small.

$$\min_\theta \frac{1}{2m} \sum_{i=1}^{m} (h_\theta)(x^{(i)} - y^{(i)})$$

Cost Function

Small values for parameters $\theta_0, \theta_0, \theta_1, \theta_2, \ldots, \theta_n$

- "Simpler" hypothesis
- Less prone to overfitting

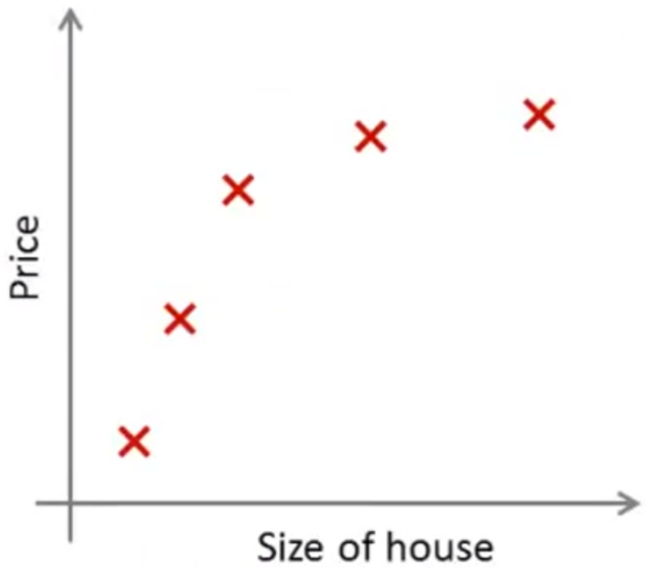
Housing:

- Features: $x_1, x_2, \ldots, x_{100}$
- Parameters: $\theta_0, \theta_0, \theta_1, \theta_2, \ldots, \theta_n$

$$\min_\theta \frac{1}{2m} \sum_{i=1}^{m} (h_\theta)(x^{(i)} - y^{(i)})$$

Cost Function

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

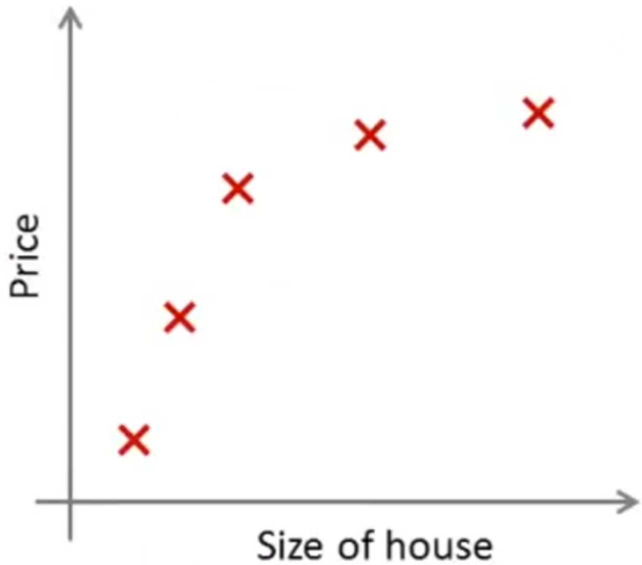


Cost Function

In regularized linear regression, we choose to minimize

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

What happen if we set λ to a large number (e.g $\lambda = 10^{10}$)?



Cost Function

What happen if we set λ to a large number (e.g $\lambda = 10^{10}$)?

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

- Algorithm works fine; setting λ to be large can't hurt it.
- Algorithm fails to eliminate overfitting.
- Algorithm results in underfitting. (Fails to fit even training data).
- Gradient descent will fail to converge.

Regularized linear regression

1. Regularized linear regression

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

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

Regularized linear . . .
Gradient descent
Regularized logistic . . .
Gradient descent

2. Gradient descent

Repeat {

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

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

$$(j = 0, 1, 2, 3, \ldots, n)$$

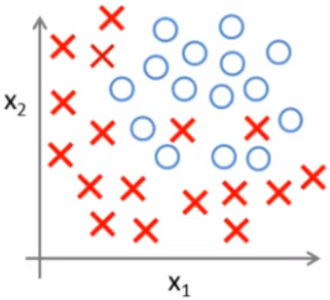
}

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

Regularized linear . . .
Gradient descent
Regularized logistic . . .
Gradient descent

Regularized logistic regression

3. Regularized logistic regression



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \dots)$$

Cost function:

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

Regularized linear . . .
Gradient descent
Regularized logistic . . .
Gradient descent

4. Gradient descent

Repeat {

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

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

$$(j = 0, 1, 2, 3, \dots, n)$$

}

References

- Machine Learning Andrew Ng Stanford University Coursera https://www.coursera.org/learn/machine-learning/home/week/3
- https://github.com/hammadshaikhha/Math-of-Machine-Learning-Course-by-Siraj/blob/master/Regularization%20in%20Linear%20Regression/Regularization%20in%20Linear%20Regression.ipynb
- https://medium.com/coinmonks/regularization-of-linear-models-with-sklearn-f88633a93a2
- https://scikit-learn.org/stable/auto_examples/model_selection/plot_underfitting_overfitting.htmlsphx-glr-auto-examples-model-selection-plot-underfitting-overfitting-py
- https://bigdata-madesimple.com/how-to-run-linear-regression-in-python-scikit-learn/
- https://towardsdatascience.com/regularization-in-machine-learning-76441ddcf99a