# Overfitting in Linear Regression

Dr. Muhammad Wasim

# Overfitting in Linear Regression


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

High Bias Underfitting High Variance Overfitting

- If we have too many features, the learned hypothesis may fit the training set very well, but fail to generalize to new examples.

# Addressing the Overfitting

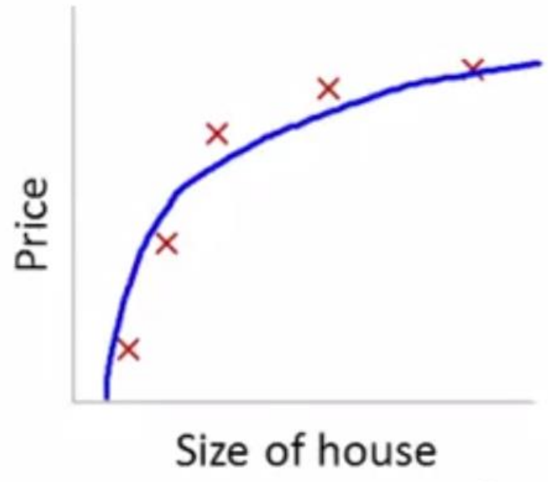
1. Reduce the number of features ✓
2. Regularization
   - Keep all the features, but reduce the magnitude of parameters $\theta_j$
   - Works well when we have  a lot of feature.

```
pf.get_feature_names_out()
```

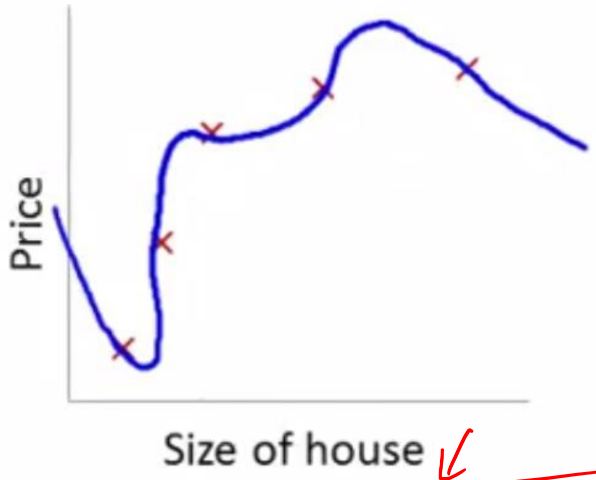
```
array(['1', 'MedInc', 'HouseAge', 'AveRooms', 'AveBedrms', 'Population',
       'AveOccup', 'Latitude', 'Longitude', 'MedInc^2', 'MedInc HouseAge',
       'MedInc AveRooms', 'MedInc AveBedrms', 'MedInc Population',
       'MedInc AveOccup', 'MedInc Latitude', 'MedInc Longitude',
       'HouseAge^2', 'HouseAge AveRooms', 'HouseAge AveBedrms',
       'HouseAge Population', 'HouseAge AveOccup', 'HouseAge Latitude',
       'HouseAge Longitude', 'AveRooms^2', 'AveRooms AveBedrms',
       'AveRooms Population', 'AveRooms AveOccup', 'AveRooms Latitude',
       'AveRooms Longitude', 'AveBedrms^2', 'AveBedrms Population',
       'AveBedrms AveOccup', 'AveBedrms Latitude', 'AveBedrms Longitude',
       'Population^2', 'Population AveOccup', 'Population Latitude',
       'Population Longitude', 'AveOccup^2', 'AveOccup Latitude',
       'AveOccup Longitude', 'Latitude^2', 'Latitude Longitude',
       'Longitude^2'], dtype=object)
```

$\theta$ is

# The idea behind Regularization




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




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

- Suppose, we penalize and make $\theta_3, \theta_4$ very small.

# Regularization

- Small values for parameters $\theta_0, \theta_1, \dots, \theta_n$
  - "Simpler" hypothesis
  - Less prone to overfitting
- Housing Example
  - Features: $x_1, x_2, \dots, x_n$
  - Parameters: $\theta_0, \theta_1, \dots, \theta_n$

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left( h\left(x^{(i)}\right) - y^{(i)} \right)^2$$

# Regularization

100

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

$\alpha$

- This special form of linear regression is called L2 regularization or Ridge Regression
- Larger the value of $\lambda$, less the variance of the model will be. It just means that we are adding some bias to the model.
- Another popular form: L1 Regularization or Lasso Regression:

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

- You can use Gradient Descent to find the best values of parameters.