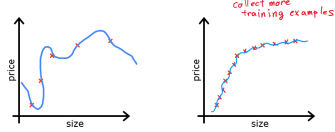


# Ungraded Lab: Overfitting

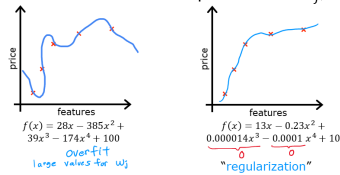
## Collect more Training Data



DeepLearning.AI

Stanford ONLINE

## Reduce the size of parameters $w_j, b$



DeepLearning.AI

Stanford ONLINE

## Select Features to Include/Exclude

size	bedrooms	floors	age	avg income	school rating	distance to coffee shop	price
$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$y$
all features							
over-fit							
selected features							
size							
bedrooms							
school ratings							
just right							
model selection							
course 2							

DeepLearning.AI

Stanford ONLINE

## Goals

In this lab, you will explore:

- the situations where overfitting can occur
- some of the solutions

In [1]:

```

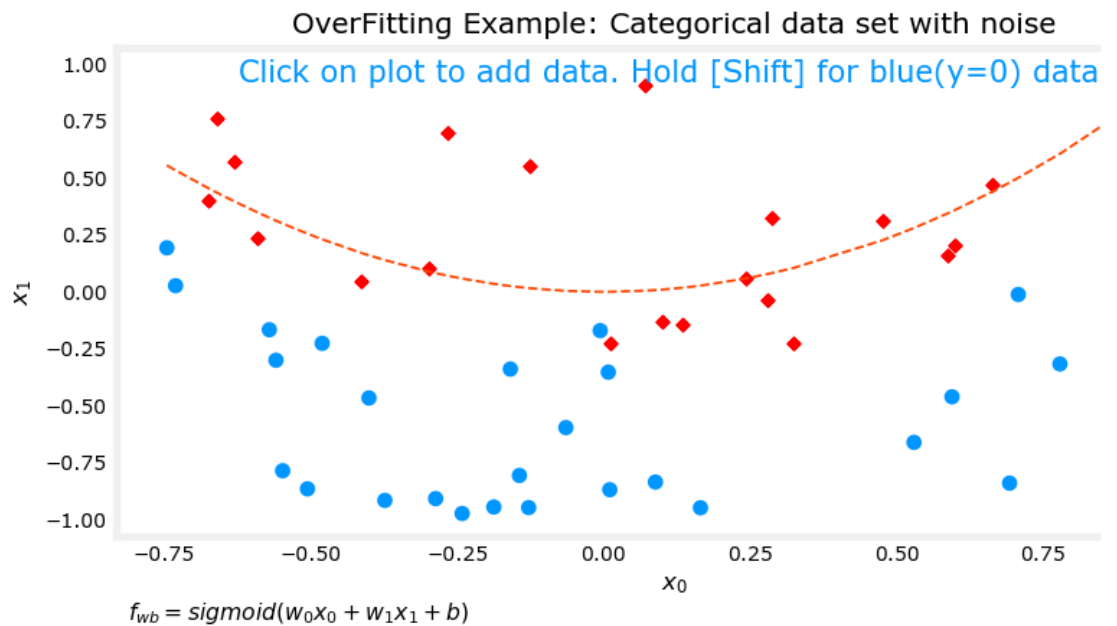
1 %matplotlib widget
2 import matplotlib.pyplot as plt
3 from ipywidgets import Output
4 from plt_overfit import overfit_example, output
5 plt.style.use('./deeplearning.mplstyle')

```

## Overfitting

The week's lecture described situations where overfitting can arise. Run the cell below to generate a plot that will allow you to explore overfitting. There are further instructions below the cell.

```
In [2]: 1 plt.close("all")
        2 display(output)
        3 ofit = overfit_example(False)
```



Degree

- 1
- 2
- 3
- 4
- 5
- 6

fit data

- Regression
- Categorical

In the plot above you can:

- switch between Regression and Categorization examples
- add data
- select the degree of the model
- fit the model to the data

Here are some things you should try:

- Fit the data with degree = 1; Note 'underfitting'.
- Fit the data with degree = 6; Note 'overfitting'
- tune degree to get the 'best fit'
- add data:
  - extreme examples can increase overfitting (assuming they are outliers).
  - nominal examples can reduce overfitting
- switch between Regression and Categorical to try both examples.

To reset the plot, re-run the cell. Click slowly to allow the plot to update before receiving the next click.

Notes on implementations:

- the 'ideal' curves represent the generator model to which noise was added to achieve the data set

- 'fit' does not use pure gradient descent to improve speed. These methods can be used on smaller data sets.

## Congratulations!

You have developed some intuition about the causes and solutions to overfitting. In the next lab, you will explore a commonly used solution, Regularization.

In [ ]: