DS Certificate Course III

Notebook:	Data Science Certificate		
Created:	4/13/2020 5:15 PM	**Updated:**	5/10/2020 4:08 PM
Author:	nataliegmoore@gmail.com		
URL:	http://www.madehow.com/Volume-3/Disposable-Diaper.html		

STEP-BY-STEP DATA ANALYSIS:

- get data and load it into workspace
- explore data visually and with tables
- clean up/prepare data
    - removing/replacing missing values or NaNs
    - fixing dtypes
    - normalizing if needed
    - classifying a multi-class feature using one-hot-encoding or risk values
    - creating subsets of the data to try out different methods/models on later
- visually explore data with target(s) in mind
- feature selection
- train and test different models to predict target(s)
- cross validate results (especially for sparse data)
- fine-tune model hyper-parameters for the best results
    - sklearn GridSearch automates this $$$$

http://www.madehow.com/Volume-3/Disposable-Diaper.html

Diaper manufacturing notes:

- absorbent pad between two sheets of nonwoven fabric
- absorbent pad first vacuum formed
- attached to a permeable top/bottom sheet
- sealed with heat or ultrasonic vibes
- elastic fibers are attached to the sheets to gather edges and fit snugly

- developed in the early 1970s
- preceded by cotton diapers
    - poor absorbtion

Lesson 2

overfitting = low train error % but high test error % large gap between their curves