## Deciding What to Try Next

We now know many machine learning techniques. But, there is a big difference between someone who knows an algorithm vs. someone less familiar and doesn't understand how to apply them. Here we focus deciding what avenues to try

## Debugging a learning algorithm:

Suppose you have implemented regularized linear regression to predict housing prices.

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

However, when we test our hypothesis on a new set of houses, we find that it makes unacceptably large errors in its predictions. What should we try next?

There are many things we can do:

- **Get more training examples:** One thing we could try, is to get more training examples. And concretely, we can setting up phone surveys, going door to door, to try to get more data on how much different houses sell for (but sometimes getting more training data doesn't actually help).
- **Try smaller sets of features:** Other things we might do are to try a smaller set of features. Maybe we want to spend time carefully selecting some small subset of them to prevent overfitting.
- **Try getting additional features:** Maybe we need to get additional features. Maybe the current set of features aren't informative enough and we want to collect more data in the sense of getting more features.
- **Try adding polynomial features ($x_1^2$, $x_2^2$, $x_1x_2$, etc.)**
- **Try decreasing $\lambda$ (The regularization parameter)**
- **Try increasing $\lambda$**

**Video Question:** Which of the following statements about diagnostics are true? Check all that apply.

- It’s hard to tell what will work to improve a learning algorithm, so the best approach is to go with gut feeling and just see what works.

> Diagnostics can give guidance as to what might be more fruitful things to try to improve a learning algorithm.

> Diagnostics can be time-consuming to implement and try, but they can still be a very good use of your time.

> A diagnostic can sometimes rule out certain courses of action (changes to your learning algorithm) as being unlikely to improve its performance significantly.

## Machine learning diagnostic:

Diagnostic: A test that you can run to gain insight what is/isn't working with a learning algorithm, and gain guidance as to how best to improve its performance.

Diagnostics can take time to implement, but doing so can be a very good use of our time