

**Plan:**

1. Evaluating a learning algorithm - trying to understand is it working correctly
  - a. Things we could do
  - b. Evaluating a hypothesis
  - c. Model selection problem
2. Bias vs, Variance - underfitting vs. overfitting
  - a. Polynomial degrees
  - b. Regularization
  - c. Learning curves

**Questions:**

1. What's the difference between high bias and high variance problems?
2. What is CV (cross-validation set)?
3. If training error and test error are about the same, is the algorithm suffering from high bias, high variance, or neither?

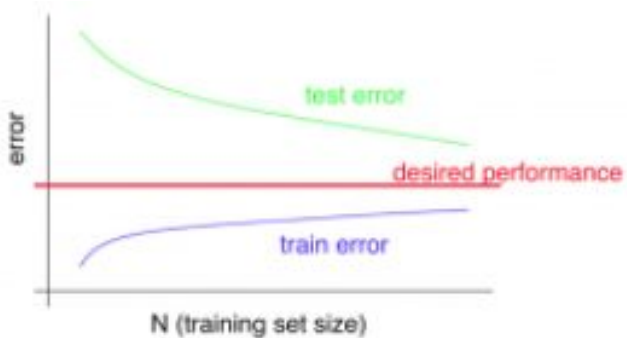
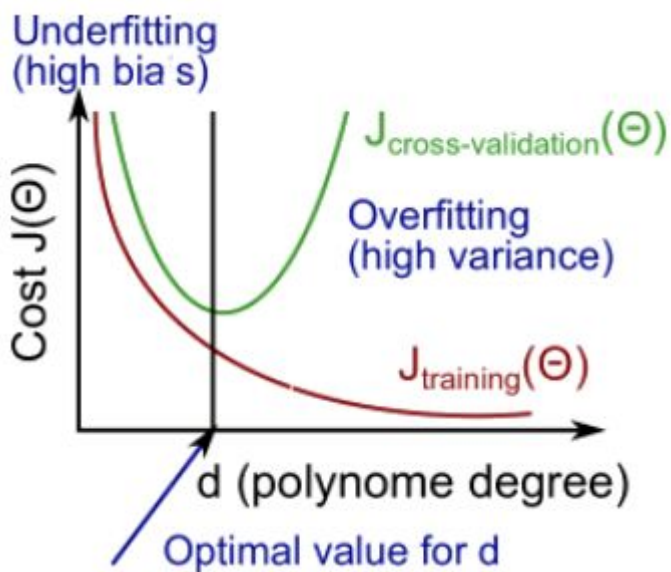
## Glossary:

High bias - underfitting - we can solve high bias problem by increasing amount of features

High variance - overfitting - we can solve high variance problem by increasing amount of features

CV - cross validation set, used to evaluate algorithms and decide which is best - typically crossvalidation set is 20% of all training examples.

- Getting more training examples
- Trying smaller sets of features
- Trying additional features
- Trying polynomial features
- Increasing or decreasing  $\lambda$



- **Getting more training examples:** Fixes high variance
- **Trying smaller sets of features:** Fixes high variance
- **Adding features:** Fixes high bias
- **Adding polynomial features:** Fixes high bias
- **Decreasing  $\lambda$ :** Fixes high bias
- **Increasing  $\lambda$ :** Fixes high variance.

More examples of output

Great presentations (did he made them himself)

Talked not only about specific for mongo stuff but for example normal forms

Questions?

Submit your weekly reflection of **about 300-400 words** no later than **2 hours after the video-conference class**. Answer the following questions:

### ***Part 1***

1. What have I learned this week?
2. What was particularly useful and how can I use it in my studies and work?
3. What was redundant? (I already knew that)
4. What difficulties have I experienced?
5. What would I change in the content/quality of the week?

### ***Part 2. Feedback on your groupmate's report (for 1 person of teacher's choice):***

1. Was the report clear, easy to listen and understand? Why/why not? What would you advise to improve the performance?
2. How relevant/interesting was the content of the report to you?

This week I learned a lot of practical tricks and methods to effectively debug learning algorithms and properly evaluate a hypothesis I use. First, we started with some ideas about what in general we can do to somehow improve our algorithm. We ended up with the list like this:

- Getting more training examples
- Trying smaller sets of features
- Trying additional features
- Trying polynomial features
- Increasing or decreasing  $\lambda$

After that, we tried to understand how exactly these operations could affect our learning rate to make it clear when we use each one of them. To understand that, we firstly, took a look at what could possibly go wrong and figured out that we could possibly have a high-bias problem (which is also called 'underfitting') or a high-variance problem (which is also called 'overfitting').

When we realized possible problems, we figured out how to identify them. For diagnosing the problem, we use learning curves. We also used these curves to understand when we should apply each of the action listed above:

- Getting more training examples: Fixes high variance
- Trying smaller sets of features: Fixes high variance
- Adding features: Fixes high bias
- Adding polynomial features: Fixes high bias
- Decreasing  $\lambda$ : Fixes high bias
- Increasing  $\lambda$ : Fixes high variance.

By building learning curves, we can understand if our learning algorithm is struggling with high-bias or high-variance and use this knowledge combined with the list above to decide how to improve our algorithm.

I think, the information I've learned this week is extremely useful and I can see a lot of scenarios of how to apply this knowledge on practice, wouldn't change anything this week.

There is some feedback on Nikita's report:

Nikita had a few presentations (don't know if he made the by himself) and that was cool, but there was a lot of redundant information and he basically skipped half of the slides. However, he talked a lot of interesting stuff on his topic and even mentioned some non-MongoDB specific stuff, which is cool in my opinion. Unfortunately, I didn't have an opportunity to ask him questions, because I got some. Either way, I think he had a great presentation and almost everything he said was clear and understandable.