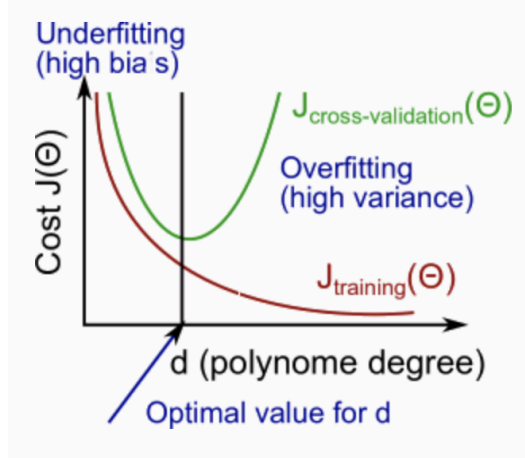


6주차 - Bias vs Variance

Diagnosing Bias vs Variance



다음과 같이 각 d 에 대하여 비용함수를 그래프로 그리면, training set에 대한 비용함수는 점점 감소하지만 cross-validation or test set에 대한 비용함수는 특정한 Optimal value를 기준으로 양쪽으로 증가한다. 왜냐하면 d 가 너무 작거나 크면 bias or variance 상황이 발생하여 비용함수가 커지기 때문이다. 이때 왼쪽(두 비용함수 모두 큰 상황)은 Underfitting(high bias)으로 볼 수 있고 오른쪽(training 비용함수는 낮으나 validation or test 비용함수는 그에 비해 엄청 크다)은 Overfitting(high variance)으로 볼 수 있다.

Regularization and Bias/Variance

Regularization을 진행할 때 좋은 λ 를 선택하는 방법을 알아본다. 과정은 다음과 같다.

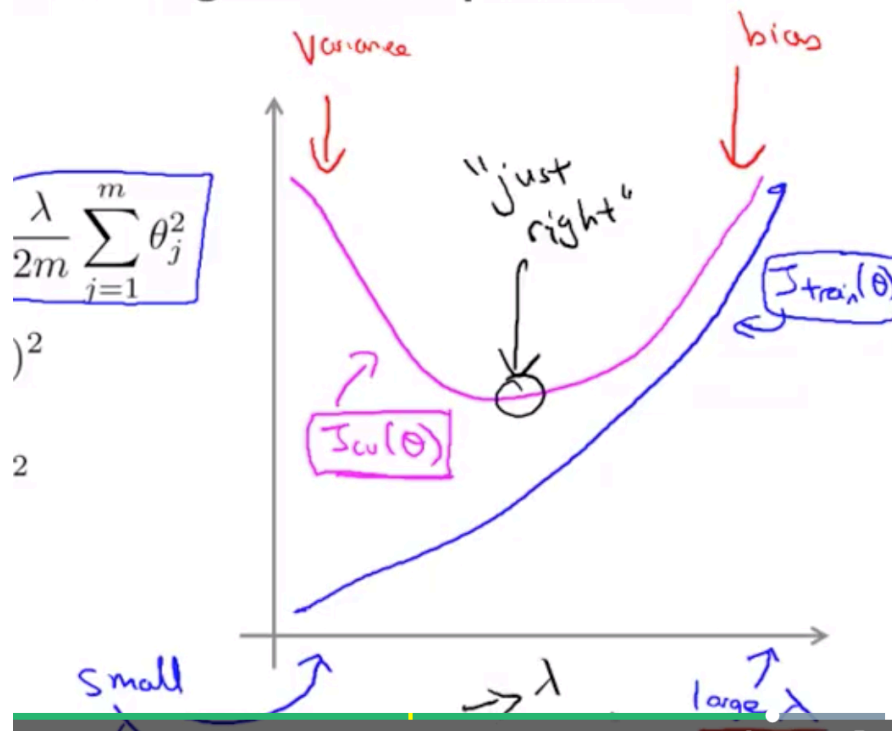
1. Create a list of λ s (i.e. $\lambda \in \{0, 0.01, 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, 1.28, 2.56, 5.12, 10.24\}$);
2. Create a set of models with different degrees or any other variants.
3. Iterate through the λ s and for each λ go through all the models to learn some θ .
4. Compute the cross validation error using the learned θ (computed with λ) on the $J_{CV}(\theta)$ without regularization or $\lambda = 0$.
5. Select the best combo that produces the lowest error on the cross validation set.
6. Using the best combo θ and λ , apply it on $J_{test}(\theta)$ to see if it has a good generalization of the problem.

Regularization을 진행할 때는 $J(\theta)$ 와 $J_{train}(\theta)$ 를 구분하여 사용한다. $J(\theta)$ 는 regularization term이 붙은 것으로 θ 를 learning 할 때 사용하고, $J_{train}(\theta)$ 는 train data set를 이용하지만 regularization term이 없는 비용함수를 의미한다. 이는 λ 에 대한 overfitting/underfitting을 진단하기 위해 그래프를 그릴 때 필요하다.

$J_{train}(\theta)$: λ 가 작을 때는 $\min(J(\theta))$ 와 거의 비슷해짐으로 역시나 최솟값을 가지고 있

다. 하지만 λ 가 커지면 직관적으로 $\text{MIN}(J(\theta))$ 와 $J_{\text{train}}(\theta)$ 값의 차이는 커지게 됨을 알 수 있다.

$J_{\text{cv}}(\theta)$: variance와 bias 상태 모두 새로운 데이터를 예측하지 못한다. 따라서 비용함수가 클 수밖에 없고 그사이 적당한 λ 값에 대해서 최저값을 갖는다.



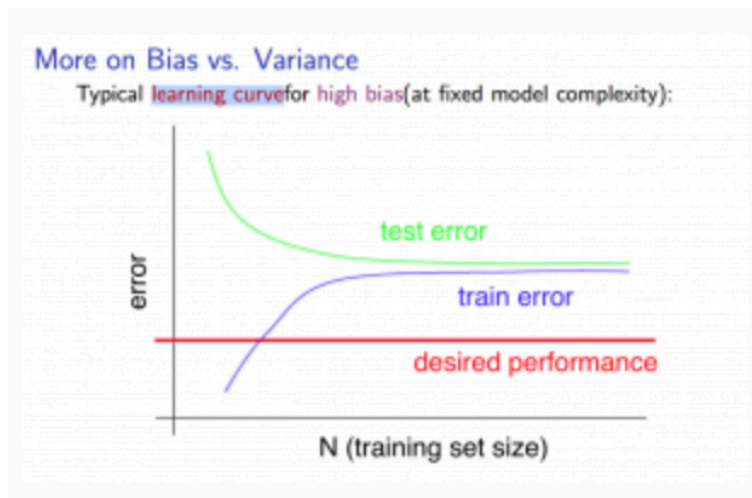
Learning Curves

learning curve는 training data set에 대한 error 값을 나타내는 직교좌표계에 $J_{\text{train}}(\theta)$ 와 $J_{\text{test}}(\theta)$ 를 그려놓은 것으로, 특정 고정된 모델에 대하여 Bias인지 Variance인지 살펴볼 수 있는데 용이하다.

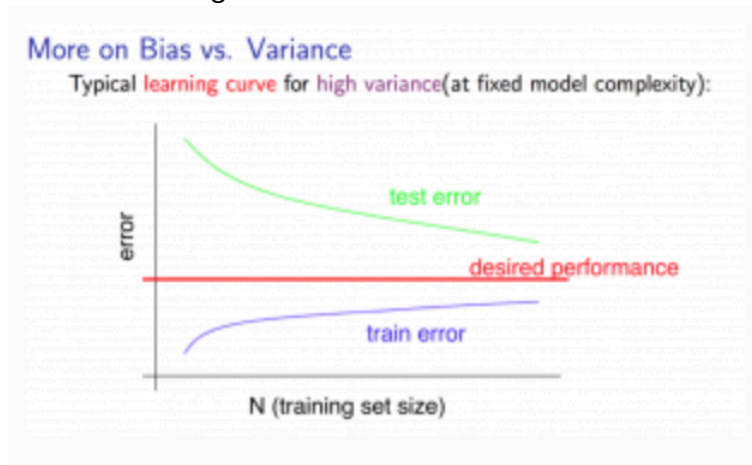
High Bias 상황의 경우

test error와 train error가 같은 값으로 수렴한다.

이미 수렴했기 때문에 여기서 데이터를 더 구하는 것은 시간낭비이다.



High Variance 상황의 경우
test error와 train error가 수렴하는 값에 차이가 생긴다.
이 경우는 training data를 더 구하는 것이 합리적이다.



Deciding What to Do Next Revisited

6주차 앞에서 배웠던 내용을 바탕으로, 우리의 예측이 틀렸을 때 할 수 있었던 process를 다시 보자.

- * 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 λ : Fixes high bias
- * Increasing λ : Fixes high variance.

learning curve나 그외 그래프 등을 통해서 결국 bias와 variance중 어느 상태인지 파악한 뒤 그에 맞는 조치를 취하면
많은 시간을 줄 일 수 있다.

- Diagnosing Nerual Network

배웠던 내용을 바탕으로 인공지능망에도 적용해 볼 수 있다.

- * A neural network with fewer parameters is prone to underfitting. It is also computationally cheaper.
- * A large neural network with more parameters is prone to overfitting. It is also computationally expensive. In this case you can use regularization (increase λ) to address the overfitting.