Applied Artificial Intelligence

Session 17: Last Notes on Feedforward Networks
&
Moving to Convolutional Neural Network

Fall 2018

NC State University

Instructor: Dr. Behnam Kia

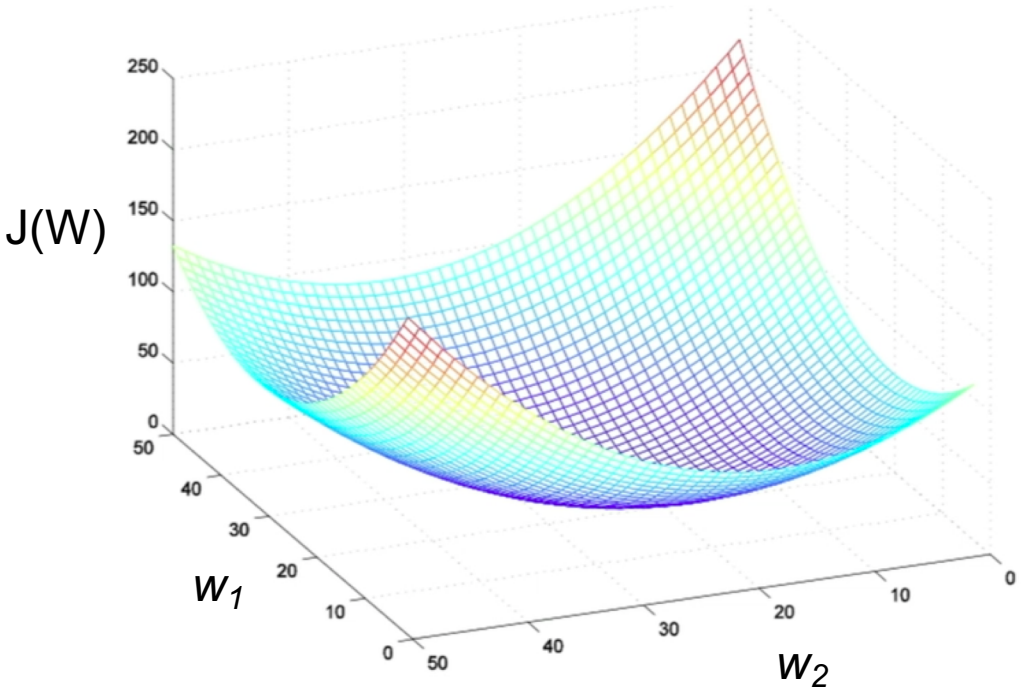
Course Website: https://appliedai.wordpress.ncsu.edu/

Training Error

$$W_{next} = W_{current} - \alpha \nabla_W J(W)$$

$$\frac{\partial J(W)}{\partial w_j} = \frac{1}{m} \sum_i x_{pj}^i (x_p^i . W - y^i)$$

$$\nabla_W J(W) = \frac{1}{m} X_p^T . (X_p . W - y)$$



(Statistical) Generalization

- The main challenge in machine learning is to make sure our algorithm perform well on new, previously unseen inputs—not just those on which our model was trained.

- The ability to perform well on previously unobserved inputs is called **Generalization**.

- What separates machine learning from optimization is that we want the generalization error (AKA test error) to be low as well. The generalization error is defined as the expected value of the error on a new, unseen input.

How can we affect performance on the test set when we can observe only the training set?

How can we affect performance on the test set when we can observe only the training set?

- Trivial (but extremely important) step:
  - Make sure training set has the same distribution as the test set.

How can we affect performance on the test set when we can observe only the training set?

- Trivial (but extremely important) step:
  - Make sure training set has the same distribution as the test set.

i.i.d assumption in statistical learning:

We typically make a set of assumptions known collectively as the i.i.d. assumptions . These assumptions are that the examples in each dataset are independent from each other, and that the training set and test set are ***identically distributed****, drawn from the same probability distribution as each other. The same distribution is then used to generate every train example and every test example. We call that shared underlying distribution the data-generating distribution.*

How can we affect performance on the test set when we can observe only the training set?

- Trivial (but extremely important) steps:
  - Make sure training set has the same distribution as the test set.

You cannot train model using one type/style of data and expect it to perform well on a different type/style of data.

- code

How can we affect performance on the test set when we can observe only the training set?

- Trivial (but extremely important) steps:
  - Make sure training set has the same distribution as the test set.
  - Make sure you have enough training data that samples and represents many different possible cases.

XOR Problem: Multilayer Neural Network Solution

XOR is not linearly separable

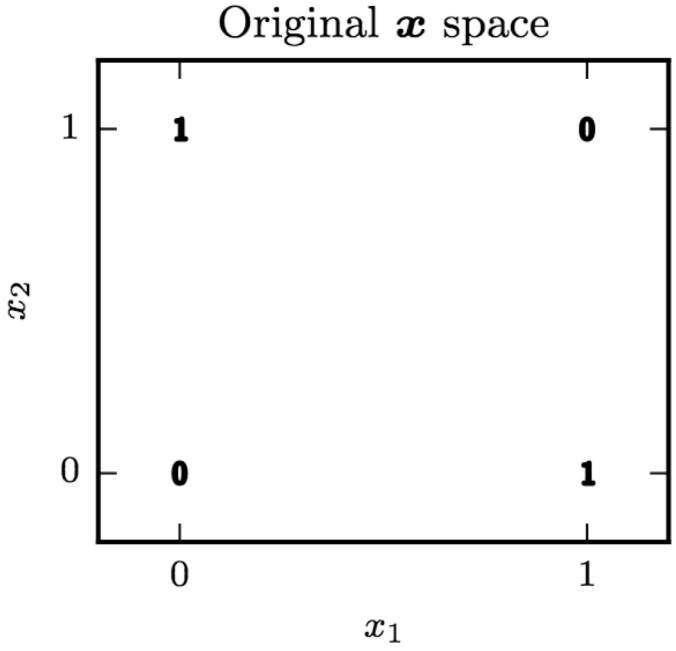


Figure 6.1, left

(Goodfellow 2017)

Slides are from Goodfellow, et.al for “Deep Learning” Book

Our Goal in Machine Learning

- 1. Make the training error small.
- 2. Make the gap between training and test error small.

Our Goal in Machine Learning

- 1. Make the training error small. —Failure to do so→ Underfitting

- 2. Make the gap between training and test error small.

- code

Our Goal in Machine Learning

- 1. Make the training error small. Failure to do so → Underfitting

- 2. Make the gap between training and test error small. Failure to do so → Overfitting

- code

Generalization and Capacity