Comp Sci 5480 - Deep Learning: Assignment 1

Aaron Dubale

February 2026

1 Understanding Learning as Function Approximation

1.1 Counting Model Capacity

Consider predicting house prices from square footage (x). For each hypothesis class below, write the general form and count the number of parameters.

- **Linear functions:**

  The general form of a linear function is

  $$f(x) = w_1 x + w_0$$

  This model has two parameters: the slope w_1 and the intercept w_0. Therefore, the linear hypothesis class has capacity 2.

- **Quadratic polynomials:**

  The general form of a quadratic polynomial is

  $$f(x) = w_2 x^2 + w_1 x + w_0$$

  This model has three parameters: w_2, w_1, and w_0, so it has capacity 3.

- **Cubic polynomials:**

  The general form of a cubic polynomial is

  $$f(x) = w_3 x^3 + w_2 x^2 + w_1 x + w_0$$

  This model has four parameters: w_3, w_2, w_1, and w_0. Therefore, the cubic polynomial class has capacity 4.

- **Two-layer neural network with 5 hidden units and ReLU activation:**

  The hidden layer computes

  $$h_j = ReLU(w_j x + b_j), j = 1, ..., 5$$

  While the equation representing the output layer is:

  $$y = \sum_{j=1}^{5} v_j h_j + c$$

  The first layer has 5 weights and 5 biases, while the second layer has 5 weights and 1 bias.

  Therefore, there are $10 + 6 = 16$ parameters. This gives it a model capacity of 16.

1.2 Model Capacity and Real-World Constraints

Scenario:

1.3 Why Training Error Misleads You

2 Loss Functions and Their Properties

2.1 Computing Different Losses

2.2 Outlier Sensitivity in Production Systems

2.3 Why Gradient Descent Needs Smooth Losses

3 Why Stacking Linear Layers Does Nothing

3.1 Two Linear Layers

3.2 With Nonlinearity

4 Sigmoid and Tanh Equivalence

4.1 Sigmoid in Exponential Form

4.2 Scaling the Input

4.3 Deriving the Tanh Relationship

4.4 Neural Network Implications