Comprehensive Review: Feedforward Neural Networks

DSC 255 - Machine Learning Fundamentals

Contents

1 Introduction to Feedforward Neural Networks

Feedforward neural networks (FNNs) are prediction models inspired loosely by biological neural systems. They process data in a single forward pass from input to output.

- Inputs are fed at the bottom layer.
- Computations propagate layer-by-layer to the output.
- Each internal node computes a function based on its inputs (parents).

2 Neural Network Architecture

Layer Structure

- Input layer: d nodes representing input features $x_1, \ldots, x_d$
- Hidden layers: intermediate computation nodes
- Output layer: nodes producing the final prediction

Computation at a Hidden Node

Let $z_1, \ldots, z_m$ be the parent values to a hidden node h. The value of h is computed as:

$$h = \sigma(w_1 z_1 + w_2 z_2 + \cdots + w_m z_m + b)$$

where:

- $w_1, \ldots, w_m$ are weights
- b is a bias term
- $\sigma(\cdot)$ is a non-linear activation function

3 Common Activation Functions

(a) **Threshold (Heaviside) Function**:

$$\sigma(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

(b) **Sigmoid Function**:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

(c) **Hyperbolic Tangent**:

$$\sigma(z) = \tanh(z)$$

(d) **ReLU (Rectified Linear Unit)**:

$$\sigma(z) = \max(0, z)$$

4 Why Nonlinearities Matter

If all activations are linear, the entire network collapses to a single linear transformation:

$$h_2 = W_2 W_1 x$$

This results in a network that has no greater expressive power than a single linear model. Non-linear activations enable the network to represent more complex functions.

5 Output Layer and Softmax

For classification with k labels:

- Output layer has k nodes computing $y_1, \ldots, y_k$
- Probabilities are computed using the softmax function:

$$\Pr(\text{label} = \text{j}) = \frac{e^{y_j}}{\sum_{i=1}^{k} e^{y_i}}$$

6 Model Complexity and Parameters

- Each edge in the network corresponds to a parameter (weight).
- Total parameters can grow rapidly as layers and nodes increase.
- Fully connected layers with 1000 nodes each result in 10^6 weights.

7 Universal Approximation Theorem

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a continuous function. Then, for any $\epsilon > 0$, there exists a feedforward neural net with a single hidden layer that approximates f to within ϵ.

- One hidden layer suffices in theory.
- However, it may require a very large number of nodes.
- Deeper networks can use fewer nodes per layer for the same expressiveness.

8 Conclusion

Feedforward neural networks are powerful function approximators that can model highly complex relationships through layered non-linear computations. Their effectiveness depends on appropriate architecture design, non-linear activations, and sufficient data for training.