You got me looking for Attention!

Monit Sharma

November 3, 2024

Table of Contents

Introduction to Logistic Regression

- Binary classification model, outputs probability between 0 and 1.
- Uses a linear combination of inputs with a sigmoid activation function:

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

- Example: $\sigma(0.5) \approx 0.622$, meaning a 62.2% probability for the positive class.

Loss, Cost, and Optimization

- **Loss Function**: Measures error for each sample.

$$\text{Loss} = -\left(y \cdot \log(\hat{y}) + (1-y) \cdot \log(1-\hat{y})\right)$$

- **Cost Function**: Average of all losses over the dataset.

$$\text{Cost} = \frac{1}{m} \sum_{i=1}^{m} \text{Loss}(y_i, \hat{y}_i)$$

- **Gradient Descent**: Minimizes cost by updating weights and bias:

$$w := w - \alpha \frac{\partial J}{\partial w}, \quad b := b - \alpha \frac{\partial J}{\partial b}$$

What is an ANN?

- ANN extends logistic regression by adding hidden layers.
- Each hidden layer increases model capacity, learning complex patterns.
- Layers:
  - **Input Layer**: Receives raw data.
  - **Hidden Layers**: Learn intermediate representations.
  - **Output Layer**: Produces final predictions.

Activation Functions

- Introduce non-linearity to learn complex relationships.
- **Sigmoid**:

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

- **Tanh**:

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

- Tanh is often preferred in hidden layers for centering outputs.

2-Layer Neural Network Structure

- Simple neural network with one hidden layer.
- Forward propagation:
  - Hidden layer:

$$z^{[1]} = W^{[1]}x + b^{[1]}, \quad a^{[1]} = \tanh(z^{[1]})$$

  - Output layer:

$$z^{[2]} = W^{[2]}a^{[1]} + b^{[2]}, \quad \hat{y} = \sigma(z^{[2]})$$

Loss, Cost, and Backward Propagation

- **Loss Function**: Same as logistic regression.

$$\text{Loss} = -\left(y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})\right)$$

- **Cost Function**: Average loss over all samples.
- **Backward Propagation**: Calculates gradients of weights and biases to minimize cost.

Parameter Updates and Prediction

- **Parameter Update**: Uses gradient descent to adjust weights and biases:

$$W := W - \alpha \frac{\partial J}{\partial W}, \quad b := b - \alpha \frac{\partial J}{\partial b}$$

- **Prediction**: Perform forward propagation with trained parameters.
- Classification is done by applying a threshold to $\hat{y}$.

Conclusion

- Logistic regression: Binary classification using sigmoid activation.
- ANN: Extends logistic regression by adding hidden layers for complex patterns.
- Key processes: Forward propagation, backward propagation, and parameter updates.
- Optimization via gradient descent with learning rate α.