CMPUT 367

Roderick Lan

Contents

Lecture 12 - Feb 29

1 Neural Networks

Neuron takes some input $x_1 \cdots x_d$

$$\begin{aligned} z &= w_1 x_1 + \cdots + w_d x_d + b \\ &= \sum_i w_i x_i + b \\ &= w^\top x + b \end{aligned}$$

$$y = f(z) \qquad \text{(activation function)}$$

NN layers usually fully connected.
Input layer - data features
Output layer - prediction
Hidden layers (≥ 1)

Suppose the weights (params) of a neural network is known.
How can we compute the output of the neural network?

1.1 Recursion/Iteration Process (Forward Prop)

1. **Initialization**
   - first layer is simply features
2. **Recursion Step**

   Assume $y^{(l-1)}$ is known; Calculate $y^{(l)}$

$$\begin{aligned} z_i^{(l)} &= \sum_{j=1}^{d^{(l-1)}} w_{ij}^{(l)} y_j^{(l-1)} + b_i^{(l)} \\ y_i^{(l)} &= f(z_i^{(l)}) \end{aligned}$$

3. **Termination**
   - when the output layer's value is computed

Matrix Vector:

$$\begin{aligned} \mathbf{z}^{(l)} &= \mathbf{W}^{(l)} \mathbf{y}^{(l-1)} + \mathbf{b}^{(l)} \\ \mathbf{y}^{(l)} &= f(\mathbf{z}^{(l)}) \end{aligned}$$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}^{(1)} & \mathbf{y}^{(2)} & \cdots & \mathbf{y}^{(m)} \end{bmatrix}^\top$$

$$\mathbf{Z}^{(l)} = \mathbf{Y}^{(l-1)}\mathbf{W}^{(l)} + \begin{bmatrix} - & \mathbf{b}^\top & - \\ - & \mathbf{b}^\top & - \\ & \vdots & \\ - & \mathbf{b}^\top & - \end{bmatrix}$$

1.2 Train Weights

$\theta = (w^{(1)}, b^{(1)}, \cdots, w^{(L)}, b^{(L)})$

$$\theta^{(\text{new})} \leftarrow \theta^{(\text{old})} - \nabla_\theta J(\theta^{(\text{old})})$$

Basic Idea: Apply chain rule top-down

1.3 Recursive Process (Backprop)

1. **Initialization**
2. **Recursion**
3. **Termination**