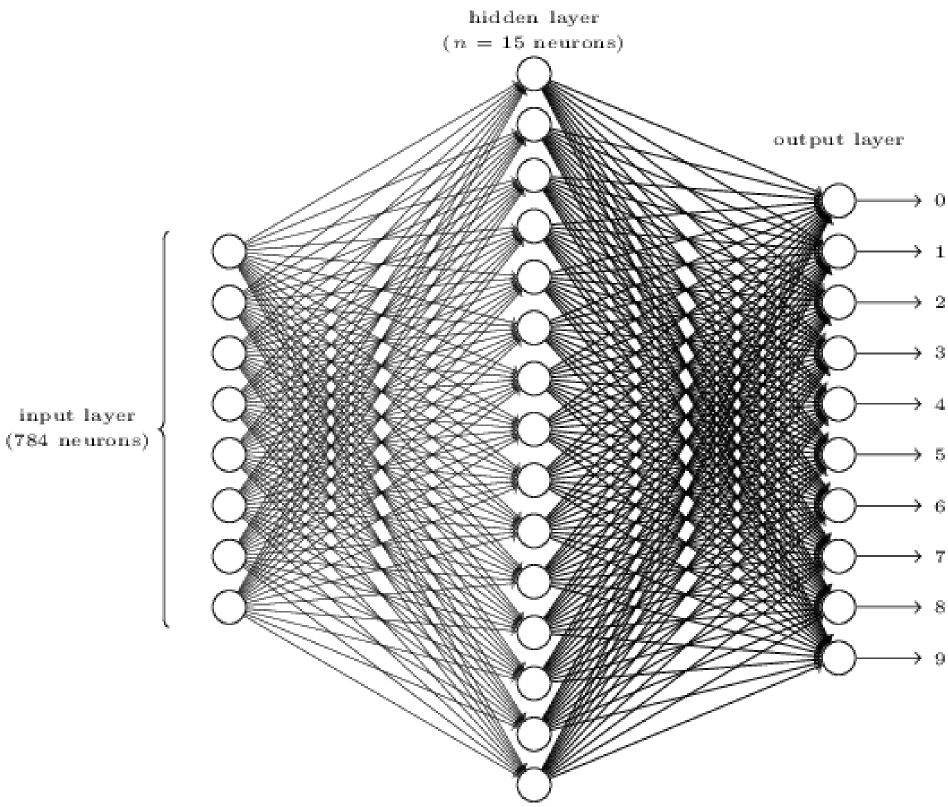
Machine Learning

Neural Networks



1. Basics

sigmoid function:

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

$$\sigma'(x) = \sigma(x) \cdot [1 - \sigma(x)]$$

hyperbolic function:

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

$$\tanh'(x) = 1 - \tanh^2(x)$$

softmax function:

$$\mathbf{y} = \mathrm{softmax}(\mathbf{x})$$

$$y_i = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$$

$$\frac{\partial y_i}{\partial x_j} = \begin{cases} -y_i \cdot y_j, & i \neq j \\ y_i \cdot (1 - y_i), & i = j \end{cases}$$

2. Model

input:

$$x \in \mathbb{R}^n$$

layer 1**:**

$$a^1 = x$$

layer l**:**

$$a^l = \sigma(w^l a^{l-1} + b^l) \quad (l = 2, \dots, L)$$

layer L**:**

$$\hat{y} = a^L$$

output:

$$\hat{y} \in \mathbb{R}^m$$

3. Backpropagation

cost function:

$$C = C(\hat{y})$$

definition:

$$z^l = w^l a^{l-1} + b^l \quad (l = 2, \ldots, L)$$

$$\delta^l = \frac{\partial C}{\partial z^l} \quad (l = 2, \ldots, L)$$

output error δ^L:

$$\begin{aligned}\delta^L = \frac{\partial C}{\partial z^L} &= \frac{\partial C}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z^L} \\ &= \frac{\partial C}{\partial a^L} \cdot \frac{\partial a^L}{\partial z^L} \\ &= \frac{\partial C}{\partial a^L} \odot \sigma'(z^L) \quad (\text{need } a^L, z^L)\end{aligned}$$

backpropagate the error:

$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l) \quad (\text{need } z^l; l = L-1, L-2, \ldots, 2)$$

output:

$$\begin{aligned}\frac{\partial C}{\partial b^l} &= \delta^l \quad (l = L, L-1, \ldots, 2) \\ \frac{\partial C}{\partial w^l} &= \delta^l \cdot (a^{l-1})^T \quad (\text{need } a^{l-1}; l = L, L-1, \ldots, 2)\end{aligned}$$