



机器学习第五讲习题以及课外阅读

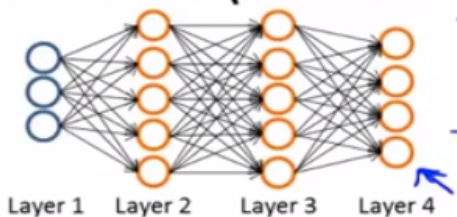
- 习题(二选一)
 - 问题1.
 - 问题2.
- 课外阅读, 请参考机器学习课程wiki.

习题(二选一)

问题1.

吴恩达视频课程中依次介绍了输出层有多个单元的神经网络及其代价函数表达式:

Neural Network (Classification)



$\rightarrow \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$
 $\rightarrow L =$ total no. of layers in network $L = 4$
 $\rightarrow s_l =$ no. of units (not counting bias unit) in layer l $s_1 = 3, s_2 = 4, s_3 = 5, s_4 = 4$

Binary classification

$y = 0$ or $1 \leftarrow$

1 output unit \leftarrow

$$h_{\Theta}(x) \in \mathbb{R}$$

$$s_L = 1, \quad K = 1 \leftarrow$$

$\Rightarrow \odot \rightarrow h_{\Theta}(x)$

Multi-class classification (K classes)

$y \in \mathbb{R}^K$ E.g. $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \leftarrow$
 pedestrian car motorcycle truck

K output units

$$h_{\Theta}(x) \in \mathbb{R}^K$$

$$s_L = K \quad (K \geq 3)$$

Andrew Ng

我们就只需要一个输出单元就可以了
we will need to use only one output unit.

Cost function

Logistic regression:

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)})) \right] + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2$$

Neural network:

$\rightarrow h_{\Theta}(x) \in \mathbb{R}^K$ $(h_{\Theta}(x))_i = i^{th}$ output

$$\rightarrow J(\Theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{k=1}^K y_k^{(i)} \log(h_{\Theta}(x^{(i)}))_k + (1 - y_k^{(i)}) \log(1 - (h_{\Theta}(x^{(i)}))_k) \right]$$

$$+ \frac{\lambda}{2m} \sum_{l=1}^{L-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (\Theta_{ji}^{(l)})^2$$

$\Theta_{ji}^{(l)} x_0 + \Theta_{ji}^{(l)} x_1 + \dots$

y_k

Andrew Ng

但这只是一个合理的约定
But this is just one possible convention

06:0

我们来看看视频讲义中对偏导数的描述,

Gradient computation

$$\rightarrow J(\Theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{k=1}^K y_k^{(i)} \log h_{\Theta}(x^{(i)})_k + (1 - y_k^{(i)}) \log(1 - h_{\Theta}(x^{(i)})_k) \right]$$

$$+ \frac{\lambda}{2m} \sum_{l=1}^{L-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (\Theta_{ji}^{(l)})^2$$

$$\rightarrow \min_{\Theta} J(\Theta)$$

Need code to compute:

$$\rightarrow -\frac{\partial}{\partial \Theta_{ij}^{(l)}} J(\Theta)$$

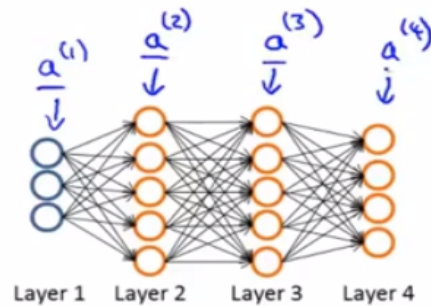
$\Theta_{ij}^{(l)} \in \mathbb{R}$

Gradient computation

Given one training example (x, y) :

Forward propagation:

$$\begin{aligned}
 & a^{(1)} = x \\
 & \rightarrow z^{(2)} = \Theta^{(1)} a^{(1)} \\
 & \rightarrow a^{(2)} = g(z^{(2)}) \quad (\text{add } a_0^{(2)}) \\
 & \rightarrow z^{(3)} = \Theta^{(2)} a^{(2)} \\
 & \rightarrow a^{(3)} = g(z^{(3)}) \quad (\text{add } a_0^{(3)}) \\
 & \rightarrow z^{(4)} = \Theta^{(3)} a^{(3)} \\
 & \rightarrow a^{(4)} = h_{\Theta}(x) = g(z^{(4)})
 \end{aligned}$$



Gradient computation: Backpropagation algorithm

Intuition: $\delta_j^{(l)}$ = "error" of node j in layer l .

For each output unit (layer $L = 4$)

$$\delta_j^{(4)} = a_j^{(4)} - y_j \quad (h_{\Theta}(x))_j \quad \delta_j^{(4)} = a_j^{(4)} - y_j$$

$$\begin{aligned}
 \rightarrow \delta^{(3)} &= (\Theta^{(3)})^T \delta^{(4)} \cdot g'(z^{(3)}) \\
 \rightarrow \delta^{(2)} &= (\Theta^{(2)})^T \delta^{(3)} \cdot g'(z^{(2)})
 \end{aligned}$$

Handwritten notes and diagrams show the backpropagation of errors. Red arrows labeled $\delta^{(1)}, \delta^{(2)}, \delta^{(3)}, \delta^{(4)}$ show the flow of errors from right to left. A handwritten formula for the derivative of the sigmoid function is shown: $\frac{\partial}{\partial z} \sigma(z) = \sigma(z)(1-\sigma(z))$. Another note says: (ignoring λ ; if $\lambda = 0$) \leftarrow .

(1) 上面的符号 $\delta_j^{(4)}$ 就是代价函数 $J(\Theta)$ 对 $y_j^{(4)}$ 的偏导数 $\frac{\partial J(\Theta)}{\partial y_j^{(4)}}$ (单个样本, $m=1$),

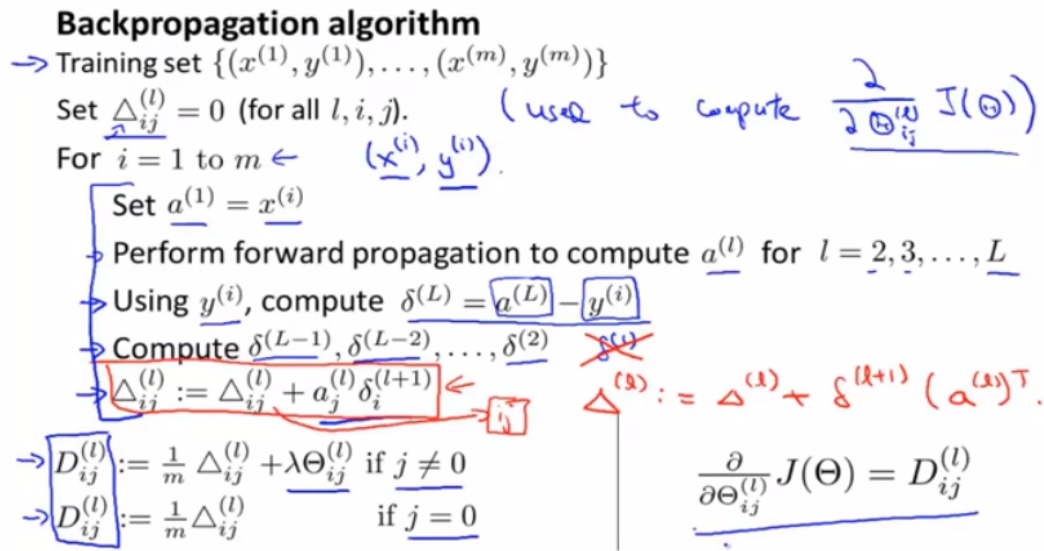
请推导这个公式.

Hint: $a_j^{(4)} = g(y_j^{(4)})$, y_j 是样本 (x, y) 中 y 的第 j 个分量.

(2) 逆推一层, 从第4层反推到第3层, 视频讲义给出了代价函数对连接第3层和第4层的参数 $\theta^{(3)}$ 在求偏导中使用的另一组偏导数 $\delta^{(3)} := \frac{\partial J(\Theta)}{\partial y^{(3)}}$, $\delta^{(2)} := \frac{\partial J(\Theta)}{\partial y^{(2)}}$ 的公式, 请推导 $\delta^{(3)}$ 的公式.

(3) 有了上面的铺垫, 我们可以继续分析吴恩达的视频讲义接下来描述的反向传播算法描述,

- 首先, 下面这段讲义中有一个记号错误, i 既用来遍历 m 个样本, 同时也用来表示参数 θ , 如 $\theta_{i,j}^{(l)}$.
- 建议保留 $\theta_{i,j}^{(l)}$ 中的记号 i , 并使用 k 来遍历 m 个样本, 请更改这段算法描述.



- $\theta_{i,j}^{(l)}$ 用来表示连接第 l 层第 j 个节点与第 $l + 1$ 层第 i 个节点的参数, 因此 i 的值从 1 开始, 而 j 的值从 0 开始.
- 这段讲义红色标注部分实际上给出了一般的代价函数 $J(\Theta)$ 对网络中任意一层参数 $\theta^{(l)}$ 的偏导数矩阵公式:
 - $\frac{\partial J(\Theta)}{\partial \theta^{(l)}} = \delta^{(l+1)} (a^{(l)})^T$
 - 请以 $l = 3, 2$ 分别为例推导这个公式, 其中 $\delta^{(3)}$ 的求解是本题的第(2)小问.

问题2.

参与这个[练习赛:图像形状分类](#), 使用神经网络模型.

(1) 尽量解释标杆模型(卷积神经网络).

(2) 修改标杆模型或使用其它模型, 提交submit, 发送排名截图.

课外阅读, 请参考[机器学习课程wiki](#).