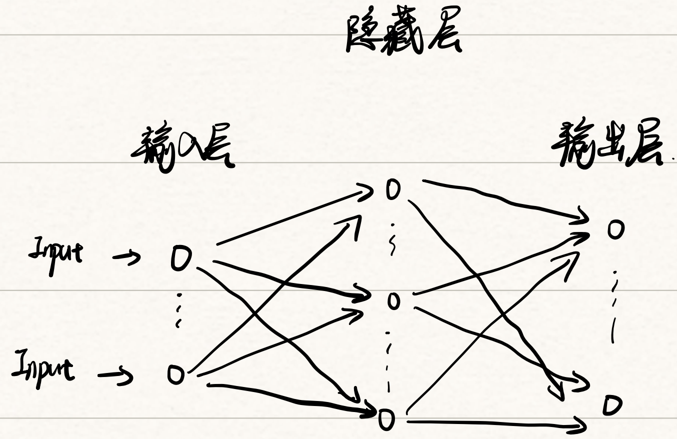
计算机视觉 作业 4

20110625 杨炽烽

(1) 对于两层神经网络.



变量声明:

w_{ij}: 表示第 i 和第 j 个结点之间的权值

d_i: 表示预期输出. y_i 表示实际输出.

θ: 表示节点的阈值

激活函数: sigmoid 函数.

首先, 定义误差 $E = \frac{1}{2}\sum_{k=1}^{L}(d_k - y_k)^2$

权值调整公式为: $W_{jk}^{(t+1)} = W_{jk}^{(t)} + \Delta W_{jk}$

由梯度下降原理可得, $\Delta W_{jk} = -\eta \cdot \frac{\partial E}{\partial W_{jk}}$, 其中 $\eta \in [0,1]$, 称为增益因子.

由链式求导法则可得, $\frac{\partial E}{\partial W_{jk}} = \frac{\partial E}{\partial I_k} \cdot \frac{\partial I_k}{\partial W_{jk}}$

我们可设 $\delta_k = -\frac{\partial E}{\partial I_k}$

由于 $\frac{\partial I_k}{\partial w_{jk}} = \frac{\partial}{\partial w_{jk}}\left(\sum_{j=1}^{m} w_{jk}\cdot O_j\right) = O_j$

则 $\Delta w_{jk} = \eta\cdot\delta_k\cdot O_j$，因此我们只需求出 δ_k 的表达式

① 当 k 为输出层，$\delta_k = -\frac{\partial E}{\partial I_k}$

$$= -\frac{\partial E}{\partial O_k}\cdot\frac{\partial O_k}{\partial I_k}$$

$$= -\frac{\partial E}{\partial y_k}\cdot\frac{\partial y_k}{\partial I_k}$$

$$= -\frac{\partial}{\partial y_k}\cdot\left[\frac{1}{2}\cdot\sum_{k=1}^{l}(d_k-y_k)^2\right]\cdot\frac{\partial y_k}{\partial I_k}$$

$$= (d_k-y_k)\times f'(I_k)$$

$$= (d_k-y_k)\,y_k\,(1-y_k)$$

则有 $w_{jk}^{(t+1)} = w_{jk}^{(t)} + \eta\cdot(d_k-y_k)\cdot y_k\cdot(1-y_k)\cdot O_j$

② 当 k 为隐藏层时，有 $\delta_j = -\frac{\partial E}{\partial I_j}$

$$= -\frac{\partial E}{\partial O_j}\cdot\frac{\partial O_j}{\partial I_j}$$

$$= -\frac{\partial E}{\partial O_j}\times f'(I_j)$$

由于 $-\frac{\partial E}{\partial O_j} = -\sum_{k=1}^{l}\frac{\partial E}{\partial I_k}\cdot\frac{\partial I_k}{\partial O_j} = \sum_{k=1}^{l}\left[\left(-\frac{\partial E}{\partial I_k}\right)\times\frac{\partial}{\partial O_j}\left(\sum_{j=1}^{m} w_{jk}O_j-\theta_k\right)\right]$

$$= \sum_{k=1}^{l}\left(-\frac{\partial E}{\partial I_k}\cdot w_{jk}\right)$$

代入可得，$\delta_j = f'(I_j) \cdot \sum_{k=1}^{L} \delta_k \cdot w_{jk}$

$$= f(I_j) \cdot [1 - f(I_j)] \cdot \sum_{k=1}^{L} \delta_k \cdot w_{jk}$$

故 $w_{jk}^{(t+1)} = w_{jk}^{(t)} + \eta\, O_j (1 - O_j) (\sum_{k=1}^{L} \delta_k w_{jk}) \cdot x_i$

(2) 多层神经网络.

定义损失函数，$\min J(w, b, x, y) = \frac{1}{2} \cdot \| a^L - y \|^2$

对于第L层，前向传播输出为：

$$a^L = g(w^L a^{L-1} + b^L) = g(z^L)$$

其中 g为激活函数，w为权值矩阵，b为偏置矩阵.

因此 $J(w, b, x, y) = \frac{1}{2} \cdot \| g(w^L a^{L-1} + b^L) - y \|_2^2$

因此我们有 W的梯度 $\frac{\partial J}{\partial w^L} = [(a^L - y) \odot g'(z^L)] \cdot (a^{L-1})^T$

b的梯度 $\frac{\partial J}{\partial b^L} = (a^L - y) \odot g'(z^L)$

我们设 $\delta^L = (a^L - y) \odot g'(z^L)$

由数学归纳法可得，$\delta^l = (\frac{\partial z^{l+1}}{\partial z^l})^T \frac{\partial J}{\partial z^{l+1}} = (\frac{\partial z^{l+1}}{\partial z^l})^T \cdot \delta^{l+1}$

且有 $z^{l+1} = w^{l+1} \cdot a^l + b^{l+1} = w^{l+1} \cdot g(z^l) + b^{l+1}$

故 $\frac{\partial z^{l+1}}{\partial z^l} = W^{l+1} \cdot diag[g'(z^l)]$

因此有递推式：$\delta^l = \mathrm{diag}[g'(z^l)](W^{l+1})^T \cdot \delta^{l+1}$

$$= (W^{l+1})^T \delta^{l+1} \odot g'(z^l)$$

故第L层的梯度 $\begin{cases} \dfrac{\partial J}{\partial W^l} = \delta^l (a^{l-1})^T \\ \dfrac{\partial J}{\partial b^l} = \delta^l \end{cases}$

迭代中可计算得 $\begin{cases} a^{i,L} = g(z^{i,L}) = g(W^L a^{i,L-1} + b^L) \\ \delta^{i,L} = (W^{L+1})^T \delta^{i,L-1} \odot g'(z^{i,L-1}) \end{cases}$

因此可有 $W^L = W^L - \alpha \cdot \sum_{i=1}^{m} \delta^{i,L} (a^{i,L})^T$

$b^L = b^L - \alpha \cdot \sum_{i=1}^{m} \delta^{i,L}$，其中$\alpha$为学习率.

用以更新权值