模式识别作业Chap 6

数据科学与计算机学院　17大数据与人工智能

17341015　陈鸿峥

问题 1 (§6 Q3). 考虑用n个模式进行m_e次训练的一个$d-n_H-c$型网络。

(a) 此问题的空间复杂度是多少？（网络参数的存贮和模式存贮都要考虑，但不考虑程序本身）。

(b) 假设网络训练由一个随机模式来训练，时间复杂度是多少？由于它受累计乘法次数的控制所以将此作为时间复杂度的测度。

(c) 假设网络由成批模式训练，时间复杂度是多少？

解答. (a) 隐含层n_H个神经元每个对应d个输入单元，因此有$n_H d$个权重；同理，连接隐含层和输出层共$n_H c$个权重，故一共的空间复杂度为$n_H(c+d)$。而模式存储的空间复杂度为nd。

(b) 随机模式即计算

$$\mathbf{w}(t+1)=\mathbf{w}(t)+\Delta\mathbf{w}(t)$$

先考虑隐含层到输出层，由公式(16)，有

$$\Delta w_{jk}=\eta\left(t_k-z_k\right)f'\left(net_k\right)y_j$$

其中，$net_k=\sum_{j=1}^{n_H}w_{jk}y_j$由$n_H$次乘法和$n_H$次加法构成，因此$\Delta w_{jk}$时间复杂度为$c(2n_H+1)$。

再考虑输入层到隐含层，由公式(20)，有

$$\Delta w_{ji}=\eta x_i f'\left(net_j\right)\sum_{k=1}^{c}w_{kj}\delta_k$$

同上理，net_j计算量为$2d$，$\sum_{k=1}^{c}w_{kj}\delta_k$计算量为$2c$，故$\Delta w_{ji}$时间复杂度为$2n_H(d+c+1)$。

进而，在一轮迭代中$\mathbf{w}$的计算量为（$\Delta\mathbf{w}$的更新及$\Delta\mathbf{w}$加上$\mathbf{w}$的计算）

$$c(2n_H+1)+2n_H(d+c+1)+(dn_H+n_Hc)=3dn_H+5n_Hc+c+2n_H$$

得到总的时间复杂度为$(3dn_H+5n_Hc+c+2n_H)m_e$。

(c) 同上理，总的模式迭代次数为nm_e，故时间复杂度为$(3dn_H+5n_Hc+c+2n_H)nm_e$。

问题 2 (§6 Q8). 考虑具有d个输入单元、n_H个隐单元、c个输出单元以及偏置的一个标准三层反向传播网。

(a) 网络中有多少权值？

(b) 考虑权值对称。特别是，证明如果将每一个权值的符号反向，网络功能不变。

(c) 现在考虑隐单元的对称交换。隐单元上没有标记，因此它们可以相互交换（沿着对应权值）而使网络功能不受影响。证明该等价标记数（对称交换因子）为$n_H!2^{n_H}$。在$n_H = 10$的情况下估计该因子的值。

解答. (a) 同§6 Q3(a)有权重$dn_H + (n_H + 1)c$，其中多出来的+1项为偏置项。

(b) 由公式(6)

$$g_k(\mathrm{x}) \equiv z_k = f\left(\sum_{j=1}^{n_H} w_{kj} f\left(\sum_{i=1}^{d} w_{ji}x_i + w_{j0}\right) + w_{k0}\right)$$

假设激活函数f为奇函数，则内层f在权值符号反向后，值也相反；但f前面还有一个权重项w_{kj}也经过了反向，故原值不变，即

$$(-w_{kj})\left[-f\left(\sum_{j=1}^{d} w_{ji}x_i\right)\right] = (w_{kj})\left[f\left(\sum_{j=1}^{d} w_{ji}x_i\right)\right]$$

(c) 考虑n_H个隐单元构成的集合的子集，一共有2^{n_H}个。而这些子集内的隐单元都可以进行重排，因此有$n_H!$种情况。故一共有$n_H!2^{n_H}$个对称交换因子。对于$n = 10$，该值为3715891200。

问题 3 (§6 Q10)**.** 在如下两种情况下，将*sigmoid*的导数用*sigmoid*本身来表示（$a, b > 0$）：

(a) 完全为正的*sigmoid*：$f(net) = \frac{1}{1+\mathrm{e}^{a \cdot net}}$。

(b) 反对称的*sigmoid*：$f(net) = a\tanh(b \cdot net)$

解答. (a) 变换得$1 + \mathrm{e}^{a \cdot net} = 1/f(net)$，进而

$$f'(net) = \frac{-a\mathrm{e}^{a \cdot net}}{(1 + \mathrm{e}^{a \cdot net})^2} = -a(f(net) - f^2(net))$$

(b) 变换得$\tanh(b \cdot net) = f(net)/a$，进而

$$f'(net) = ab\mathrm{sech}^2(b \cdot net) = ab(1 - \tanh^2(b \cdot net)) = ab(1 - f^2(net)/a^2)$$

问题 4 (§6 Q12)**.** 解释为什么输入层到隐含层的权值必须相互不等（即是随机的），否则学习不能顺利进行。更明确地说，如果权值初始化为相同的值，将出现什么现象？

解答. 若输入层到隐层的权值均为w_o，则

$$net_j = f\left(net_j\right) = \sum_{i=1}^{d} w_{ji}x_i = w_o \sum_i x_i = w_o\mathbf{x}$$

为常数，因此梯度不会产生变化，进而无法训练。

问题 5 (§6 Q17). *完成导出式(26)的推导步骤*

解答. 由式(25)有

$$J(\mathbf{w})=n\left[\frac{n_k}{n}\frac{1}{n_k}\sum_{\mathbf{x}\in\omega_k}\left[g_k(\mathbf{x},\mathbf{w})-1\right]^2+\frac{n-n_k}{n}\frac{1}{n-n_k}\sum_{\mathbf{x}\notin\omega_k}g_k(\mathbf{x},\mathbf{w})^2\right]$$

当$n\to\infty$时，ω_k中的样本比例趋于$P(\omega_k)$，由大数定律

$$\frac{1}{n_k}\sum_{\mathbf{x}\in\omega_k}\left[g_k(\mathbf{x},\mathbf{w})-1\right]^2$$

趋于

$$\mathbb{E}\left(\left[g_k(\mathbf{x},\mathbf{w})-1\right]^2|\mathbf{x}\in\omega_k\right)=\int\left[g_k(\mathbf{x},\mathbf{w})-1\right]^2p\left(\mathbf{x}|\omega_k\right)\,\mathrm{d}\mathbf{x}$$

类似的有

$$\frac{1}{n-n_k}\sum_{x\notin\omega_k}\left[F_k(\mathbf{x},\omega)\right]^2$$

趋于

$$\mathbb{E}\left([g_k(\mathbf{x},\mathbf{w})]^2\mid\mathbf{x}\in\omega_{i\neq k}\right)=\int[g_k(\mathbf{x},\mathbf{w})]^2p(\mathbf{x}\mid\omega_{i\neq k})\,\mathrm{d}\mathbf{x}$$

进而得到

$$\begin{aligned}
J(\mathbf{w})&=P\left(\omega_k\right)\int\left[g_k(\mathbf{x},\mathbf{w})-1\right]^2p\left(\mathbf{x}|\omega_k\right)d\mathbf{x}+P\left(\omega_{i\neq k}\right)\int\left[g_k(\mathbf{x},\mathbf{w})\right]^2p\left(\mathbf{x}|\omega_{i\neq k}\right)d\mathbf{x}\\
&=\int\left[g_k(\mathbf{x},\mathbf{w})-1\right]^2p\left(\mathbf{x},\omega_k\right)d\mathbf{x}+\int\left[g_k(\mathbf{x},\mathbf{w})\right]^2p\left(\mathbf{x}|\omega_{i\neq k}\right)d\mathbf{x}\\
&=\int\left[g_k^2(\mathbf{x},\mathbf{w})+1-2g_k(\mathbf{x},\mathbf{w})\right]p\left(\mathbf{x},\mathbf{w}_k\right)d\mathbf{x}+\int\left[g_k(\mathbf{x},\mathbf{w})\right]^2p\left(\mathbf{x}|\omega_{i\neq k}\right)d\mathbf{x}\\
&=\int\left[g_k(\mathbf{x},\mathbf{w})-P\left(\omega_k|\mathbf{x}\right)\right]^2p(\mathbf{x})d\mathbf{x}+\int P\left(\omega_k|\mathbf{x}\right)\left[1-P\left(\omega_k|\mathbf{x}\right)\right]p(\mathbf{x})d\mathbf{x}\\
&=\int\left[g_k(\mathbf{x},\mathbf{w})-P\left(\omega_k|\mathbf{x}\right)\right]^2p(\mathbf{x})dx+\int P\left(\omega_k|\mathbf{x}\right)P\left(\omega_{i\neq k}|\mathbf{x}\right)p(\mathbf{x})d\mathbf{x}
\end{aligned}$$

问题 6 (§6 Q26). *考虑S型激活函数*

$$f(net)=a\tanh(b\cdot net)=a\left[\frac{1-\mathrm{e}^{-b\cdot net}}{1+\mathrm{e}^{-b\cdot net}}\right]=\frac{2a}{1+\mathrm{e}^{-b\cdot net}}-a$$

(a) 证明它的导数$f'(net)$可简单写成$f(net)$的形式

(b) 在$net=-\infty$、0、$+\infty$时$f(net)$、$f'(net)$、$f''(net)$分别是多少？

解答. 注意原题的等式就是有问题的，$\tanh(b\cdot net)$展开的那条等式并不成立！因此下面是按照最右侧的等式，即$f(net)=\frac{2a}{1+\mathrm{e}^{-b\cdot net}}-a$进行计算。

(a) 同题§6 Q10(b)，有

$$f'(net) = \frac{2ab\mathrm{e}^{b\cdot net}}{(1+\mathrm{e}^{b\cdot net})^2} = \frac{b}{2a}[a^2 - (f(net))^2]$$

(b) 注意到$f''(net) = -\frac{b}{a}f(net)f'(net)$，故

– 当$net = \infty$时，有

$$\begin{aligned} f(\infty) &= \frac{2a}{1+\mathrm{e}^{-b\cdot net}} - a = 2a - a = a \\ f'(\infty) &= \frac{b}{2a}\left[a^2 - (f(net))^2\right] = \frac{b}{2a}\left(a^2 - a^2\right) = 0 \\ f''(\infty) &= -\frac{b}{2a}f(net)f'(net) = 0 \end{aligned}$$

– 当$net = 0$时，有

$$\begin{aligned} f(0) &= \frac{2a}{1+\mathrm{e}^{-b\cdot net}} - a = 2a/2 - a = a \\ f'(0) &= \frac{b}{2a}\left[a^2 - (f(net))^2\right] = \frac{b}{2a}a^2 = \frac{ab}{2} \\ f''(0) &= 0 \end{aligned}$$

– 当$net = -\infty$时，有

$$\begin{aligned} f(-\infty) &= \frac{2a}{1+\mathrm{e}^{-b\cdot net}} - a = 0 - a = -a \\ f'(-\infty) &= \frac{b}{2a}\left[a^2 - (f(net))^2\right] = \frac{b}{2a}\left(a^2 - a^2\right) = 0 \\ f''(-\infty) &= 0 \end{aligned}$$