作业7

2017年5月1日　9:13

1. $1-u_+^2-u_-^2 = 1-u_+^2-(1-u_+)^2 = -2u_+^2+2u_+$
$= -2(u_+-\frac{1}{2})^2+\frac{1}{2}$
$u_+=u_-=\frac{1}{2}$ 时，$1-u_+^2-u_-^2$ 有最大值 $\frac{1}{2}$

2. $u_+(1-(u_+-u_-))^2+u_-(-1-(u_+-u_-))^2$
$= u_+(2u_-)^2+u_-(-2u_+)^2 = 4u_+u_-(u_++u_-)$
$= 4u_+u_- = 4u_+(1-u_+) = -4(u_+-\frac{1}{2})^2+1$
$u_+=u_-=\frac{1}{2}$ 时，有最大值 1，令1比较易知结论成立

3. $\lim_{N\to\infty}(1-\frac{1}{N})^{pN} = [\lim_{N\to\infty}(1-\frac{1}{N})^N]^p = e^p$

故 $e^{-p}N$ 个样例不会被选择

4. g_1, g_2, g_3 中2个或3个出错时，G出错
$E_{out}(g_1)+E_{out}(g_2)+E_{out}(g_3)<1$

故可取到 $E_{out}(G)\geq 0$
g_1, g_2, g_3 的重叠部分最多 0.35　故 $0\leq E_{out}(G)\leq 0.35$

5. 最多可能有错误 $\Sigma_{k=1}^{K} e_k$，这些错误至少平均分配给 $\frac{K+1}{2}$ 个点，才能使G犯错．故 $E_{out}(G)\leq \Sigma_{k=1}^{K} e_k \Big/ \frac{K+1}{2}$
$= \frac{2}{K+1}\Sigma_{k=1}^{K} e_k$

6. $U_{t+1} = \Sigma_{n=1}^{N} u_n^{(t+1)} = \Sigma_{n=1}^{N} u_n^{(t)}\cdot ◆t\cdot [\![y_n\neq g_t(x_n)]\!] + \sum_{n=1}^{N} u_n^{(t)}/◆t\cdot [\![y_n = g_t(x_n)]\!]$

$= \epsilon_t\cdot◆t\cdot\Sigma_{n=1}^{N} u_n^{(t)} + (1-\epsilon_t)/◆t\cdot\Sigma_{n=1}^{N} u_n^{t}$
$= U_t(\epsilon_t\cdot◆t+\frac{1-\epsilon_t}{◆t}) = 2\sqrt{\epsilon_t(1-\epsilon_t)}\,U_t$

故 $U_3 = 2\sqrt{\epsilon_2(1-\epsilon_2)}\,U_2 = 4\sqrt{\epsilon_2(1-\epsilon_2)}\cdot\sqrt{\epsilon_1(1-\epsilon_1)}\,U_1 = 4\sqrt{\epsilon_2(1-\epsilon_2)\epsilon_1(1-\epsilon_1)}$

7. $g_1(x)=2 \Rightarrow g_1(x_n)=2$

$\min_\eta \frac{1}{N}\sum_{n=1}^{N}((y_n-s_n)-\eta g_1(x_n))^2 = \min_\eta \frac{1}{N}\sum_{n=1}^{N}(y_n-2\eta)^2$

$$\frac{\partial E}{\partial \eta}=0 \Rightarrow \frac{1}{N}\sum_{n=1}^{N} 2(y_n-2\eta)\cdot(-2)=0 \Rightarrow \eta=\frac{1}{2N}\sum_{n=1}^{N} y_n$$

$$\alpha_1=\eta=\frac{1}{2N}\sum_{n=1}^{N} y_n \qquad s_n=\alpha_1 g_1(x_n)=\frac{1}{N}\sum_{n=1}^{N} y_n$$

8. $E=\frac{1}{N}\sum_{n=1}^{N}((y_n-s_n)-\eta g_t(x_n))^2$

$$\frac{\partial E}{\partial \eta}=0 \Rightarrow \frac{1}{N}\sum_{n=1}^{N} 2(y_n-s_n^{(t-1)}-\eta g_t(x_n))\cdot(-g_t(x_n))=0$$

$$\Rightarrow \alpha_t=\eta=\frac{\sum_{n=1}^{N} g_t(x_n)(y_n-s_n^{(t-1)})}{\sum_{n=1}^{N} g_t^2(x_n)}$$

$$\Rightarrow \alpha_t\sum_{n=1}^{N} g_t^2(x_n)+\sum_{n=1}^{N} g_t(x_n)s_n^{(t-1)}=\sum_{n=1}^{N} g_t(x_n)y_n$$

故 $\sum_{n=1}^{N} s_n^{(t)} g_t(x_n)=\sum_{n=1}^{N}(s_n^{(t-1)}+\alpha_t g_t(x_n))g_t(x_n)=\sum_{n=1}^{N} g_t(x_n)y_n$

9. $w_0=d-1 \qquad w_1=w_2=\cdots w_d=1$

当 $x_1\ x_2\cdots x_d$ 均为 -1 时 $\quad g_A(x)=sign(-1)=-1$

当 $x_1\ x_2\cdots x_d$ 至少有1个+1时 $\quad \sum_{i=0}^{d} w_ix_i \geq d-1+1-(d-1)\times 1=1$

故 $g_A(x)=+1$

10. n个运算元的 n 维异或的值为真 当且仅当其中值为真的运算元有奇数个

中间层构造如下 5个神经元

x_1' 存在5个+1 $\quad x_2'$ 至少4个+1 $\quad x_3'$ 至少3个+1 $\quad x_4'$ 至少2个+1 $\quad x_5'$ 至少1个+1

$w_{i1}^{(1)}=(-5,1,1,1,1,1)$

$w_{i2}^{(1)}=(-4,1\ 1\ 1\ 1\ 1)$

$w_{i3}^{(1)}=(-3,1\ 1\ 1\ 1\ 1)$ $\qquad\qquad w_{i1}^{(2)}=(-1,1,-1,1,-1,1)$

$w_{i4}^{(1)}=(-2,1\ 1\ 1\ 1\ 1)$

$w_{i5}^{(1)}=(-1,1\ 1\ 1\ 1\ 1)$

11. $e_n=(y_n-NNet(x_n))^2=(y_n-s_1^{(L)})^2$

$$=(y_n-\tanh(x_1^{(L)}))^2=(y_n-\tanh\sum_{i=1}^{d^{(L-1)}} w_{i1}^{(L)}x_i^{(L-1)})^2$$

$$\frac{\partial e_n}{\partial w_{i1}^{(L)}}=-2(y_n-\tanh(x_1^{(L)}))\tanh'(x_1^{(L)})(x_i^{(L-1)})$$

$$\frac{\partial e_n}{\partial w_{ij}^{(l)}} = \frac{\partial e_n}{\partial s_j^{(l)}} \cdot \frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}} = \delta_j^{(l)} (x_i^{(l-1)}) \qquad (1 \le l < L)$$

$$\delta_j^{(l)} = \sum_k \delta_k^{(l+1)} (w_{jk}^{(l+1)}) (\tanh'(s_j^{(l)})) \qquad (1 \le l < L)$$

$w_{ij}^{(l)} = 0$ 由前向传播规则知 $x_i^{(l)} = 0$ $(l \ge 1 \quad i > 0)$

$1 \le l < L$ 时 $\delta_j^{(l)} = 0$ 梯度 $\frac{\partial e_n}{\partial w_{ij}^{(l)}} = 0$

$l = L$ 时 若 $i > 0$ 则 梯度 $\frac{\partial e_n}{\partial w_{i1}^{(L)}} = 0$

$i = 0$ 注意到 $x_1^{(L)} = 0$

$$\frac{\partial e_n}{\partial w_{01}^{(L)}} = -2 y_n x_0^{(L-1)}$$

故当且仅当 $y_n \ne 0$ 且 $x_0^{(L-1)} \ne 0$ 时 梯度 $\frac{\partial e_n}{\partial w_{01}^{(L)}} \ne 0$

12. $w_{ij}^{(l)} = 1$ 时 $w_{ij}^{(1)}$ 的初始值相同,

由11知 $l = 1$ 时 $\delta_j^{(1)} = \sum_k \delta_k^{(2)} \cdot 1 \cdot \tanh'(\sum_{i=0}^{d^{(0)}} x_i^0)$

故 $\delta_j^{(1)}$ 与 j 无关、均相同, 则 $w_{ij}^{(1)}$ 根据梯度下降的更新值也相同,

故始终有 $w_{ij}^{(1)} = w_{i(j+1)}^{(1)}$ $1 \le j < d^{(1)}$

13～20 见代码

13, 内部结点10个

14, E_{in} 0.00

15, E_{out} 0.13

16, $\overline{E_{in}(g_t)}$ 0.05

17, $E_{in}(G)$ 0.00

18, $E_{out}(G)$ 0.07

19, $E_{in}(G)$ 0.11

20, $E_{out}(G)$ 0.15