1.

$$F(A,B) = \frac{1}{N}\sum_{n=1}^{N} \ln\left(1+\exp\left(-y_n\left(A\cdot\left(w_{svm}^T\phi(x_n)+b_{svm}\right)+B\right)\right)\right)$$

$$= \frac{1}{N}\sum_{n=1}^{N} \ln\left(1+\exp(-y_n g(x))\right)$$

$$\text{let } g(x) = A\left(w_{svm}^T\phi(x_n)+b_{svm}\right)+B$$

$$\frac{\partial F}{\partial A} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{1+\exp(-y_n g(x))}\exp(-y_n g(x))\cdot\left(-y_n\left(w_{svm}^T\phi(x_n)+b_{svm}\right)\right)$$

$$= \frac{1}{N}\sum_{n=1}^{N}\theta(-y_n g(x))(-y_n z_n)$$

$$= \frac{1}{N}\sum_{n=1}^{N} -y_n p_n z_n$$

$$\frac{\partial F}{\partial B} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{1+\exp(-y_n g(x))}\exp(-y_n g(x))\cdot(-y_n)$$

$$= \frac{1}{N}\sum_{n=1}^{N}\theta(-y_n g(x))(-y_n)$$

$$= \frac{1}{N}\sum_{n=1}^{N} -y_n p_n$$

$$\Rightarrow \nabla F = \left(\frac{1}{N}\sum_{n=1}^{N} -y_n p_n z_n, \frac{1}{N}\sum_{n=1}^{N} -y_n p_n\right)$$

2. In order to calculate $H(F)$, notice that $\frac{d\theta(x)}{dx} = \theta(x)(1-\theta(x))$

$$\frac{\partial p_n}{\partial A} = p_n(1-p_n)(-y_n z_n), \quad \frac{\partial p_n}{\partial B} = p_n(1-p_n)(-y_n)$$

$$\frac{\partial^2 F}{\partial A} = \frac{1}{N}\sum_{n=1}^{N} p_n(1-p_n)(y_n z_n)^2$$

$$\frac{\partial^2 F}{\partial B} = \frac{1}{N}\sum_{n=1}^{N} p_n(1-p_n)y_n^2$$

$$\frac{\partial^2 F}{\partial B\partial A} = \frac{1}{N}\sum_{n=1}^{N} p_n(1-p_n)y_n^2 z_n$$

$$\frac{\partial^2 F}{\partial A\partial B} = \frac{1}{N}\sum_{n=1}^{N} p_n(1-p_n)y_n^2 z_n$$

$$\Downarrow$$

$$H(F) = \begin{bmatrix} \frac{1}{N}\sum_{n=1}^{N} p_n(1-p_n)(y_n z_n)^2, & \frac{1}{N}\sum_{n=1}^{N} p_n(1-p_n)y_n^2 z_n \\ \frac{1}{N}\sum_{n=1}^{N} p_n(1-p_n)y_n^2 z_n, & \frac{1}{N}\sum_{n=1}^{N} p_n(1-p_n)y_n^2 \end{bmatrix}$$

3. $H(F) = \frac{1}{N}\sum_{n=1}^{N} y_n^2 p_n(1-p_n)\begin{pmatrix} z_n^2 & z_n \\ z_n & 1 \end{pmatrix}$

$\because |1| = 1 \geq 0$ $\quad\therefore \forall n \in \{1,2,\cdots,N\}$, $\begin{bmatrix} z_n^2 & z_n \\ z_n & 1 \end{bmatrix}$ is positive semi-definite.

$|z_n^2| = z_n^2 \geq 0$

$\begin{vmatrix} z_n^2 & z_n \\ z_n & 1 \end{vmatrix} = 0 \geq 0$

又 $\because \frac{1}{N}, y_n^2, p_n = \frac{\exp(-y_n g(x))}{1+\exp(-y_n g(x))}, (1-p_n) = \frac{1}{1+\exp(-y_n g(x))}$ is non-negative

$\forall X \in R^2$, and $X \neq 0$

$X^T\left(\frac{1}{N}\sum_{n=1}^{N} y_n^2 p_n(1-p_n)\begin{pmatrix} z_n^2 & z_n \\ z_n & 1 \end{pmatrix}\right)X = \frac{1}{N}\sum_{n=1}^{N} y_n^2 p_n(1-p_n)X^T\begin{pmatrix} z_n^2 & z_n \\ z_n & 1 \end{pmatrix}X \geq 0$

Thus $H(F)$ is positive semi-definite.

4. $OR(X_1, X_2, X_3, X_4, \ X_d)$，因為不看 X_0，讓它永遠 True，+1

要全部皆為 -1 才能是 -1

$\Rightarrow W_0 = d-1, W_1 = W_2 = W_3 = \cdots = W_d$

$\Rightarrow g_A(X) = \text{sign}(X_1 + X_2 + \cdots + X_d + d-1)$

當所有 $X_i = -1, i = \{1,\cdots,d\}$

$g_A(X) = \text{sign}(-1+-1+\cdots+-1+d-1)$

$= \text{sign}(-1)$

當有一個 X_i 不是 -1，則 $X_1+X_2+\cdots X_d+d-1 \geq 1$

因此 $g_A(X) = \text{sign}(X_1+X_2+\cdots+X_d+d-1) = OR(X_1, X_2, \cdots, X_d)$

5. $\because w_{ij}^{(l)}$ is all 0 $\quad \frac{\partial e_n}{\partial w_{ij}^{(l)}} = x_i^{(l-1)}\delta_j^{(l)}$

$\therefore s_j^{(l)} = 0, (l = 1,2,\cdots L)$ $\quad \delta_j^{(L)} = -2(y_n - s_j^{(L)})$

$\delta_j^{(l)} = \sum_k \delta_k^{(l+1)} w_{jk}^{(l+1)} \tanh'(s_j^{(l)}) = 0 \ (l = 0,\cdots,L-1)$

$x_i^{(l-1)} = \tanh(s_i^{(l-1)}) = \tanh 0 = 0$

$\Rightarrow \forall l \in \{1,2,\cdots L\} \ \frac{\partial e_n}{\partial w_{ij}^{(l)}} = 0$ #

6. $\sum_{i=1}^{L}(d^{(i-1)}+1)\,d^{(i)} = \sum_{i=1}^{L} d^{(i-1)}d^{i} + \sum_{i=1}^{L} d^{(i)} = \sum_{i=1}^{L} d^{(i-1)}d^{i} + 49$

$\max \sum_{i=1}^{L} d^{i-1}d^{i} + 49$ subject $d^{(0)}=11$, $d^{(i)}\geq 1$ for $i\in[1,L]$, $\sum_{i=1}^{L-1}(d^{i}+1)=48$

經過觀察，當 hidden layer 超過 2 層時，weight 數量不會是最多

只需考慮 hidden layer 1 or 2 的情況

hidden layer $=1$

$$12\times 47+48=612$$

hidden layer $=2$.

$$\because (d^{(1)}+1)+(d^{(2)}+1)=48$$

$$\therefore d^{(2)}=46-d^{(1)}$$

$$12d^{(1)}+(d^{(1)}+1)d^{(2)}+(d^{(2)}+1)\cdot 1$$

$$=12d^{(1)}+(d^{(1)}+1)(46-d^{(1)})+47-d^{(1)}$$

$$=-(d^{(1)})^2+56d^{(1)}+93$$

$$=-(d^{(1)}-28)^2+877$$

A: 當 $d^{(1)}=28$, $d^{2}=18$ 時，會有最大值 877 #.

7. First, simplify $x_n - WW^T x_n$

$$x_n - WW^T x_n = \begin{bmatrix} x_{n,1} \\ x_{n,2} \\ \vdots \\ x_{n,d} \end{bmatrix} - \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix} [w_1\, w_2 \cdots w_d] \begin{bmatrix} x_{n,1} \\ x_{n,2} \\ \vdots \\ x_{n,d} \end{bmatrix}$$

$$= \begin{bmatrix} x_{n,1} \\ x_{n,2} \\ \vdots \\ x_{n,d} \end{bmatrix} - \begin{bmatrix} w_1w_1, w_1w_2, w_1w_3 \cdots w_1w_d \\ w_2w_1 \quad \vdots \\ \vdots \\ w_dw_1\; w_dw_2\; w_dw_3 \cdots w_dw_d \end{bmatrix} \begin{bmatrix} x_{n,1} \\ x_{n,2} \\ \vdots \\ x_{n,d} \end{bmatrix}$$

$$= \begin{bmatrix} x_{n,1} \\ x_{n,2} \\ \vdots \\ x_{n,d} \end{bmatrix} - \begin{bmatrix} w_1 \sum_{i=1}^{d} w_i x_{n,i} \\ \vdots \\ w_d \sum_{i=1}^{d} w_i x_{n,i} \end{bmatrix} \quad \text{let } \sum_{i=1}^{d} w_i x_{n,i} = s_n = \begin{bmatrix} x_{n,1} - s_n w_1 \\ x_{n,2} - s_n w_2 \\ \vdots \\ x_{n,d} - s_n w_d \end{bmatrix}$$

$$err_n(w) = |x - WW^T x_n|^2 = \sum_{i=1}^{d}(x_{n,i}^2 - 2s_n x_{n,i} + s_n^2 w_i^2)$$

$$\frac{\partial s_n}{\partial w_i} = \frac{\partial(w_1x_{n,1} + w_2x_{n,2} + \cdots w_dx_{n,d})}{\partial w_i} = x_{n,i}$$

$$\frac{\partial err_n(w)}{\partial w_i} = \sum_{j=1}^{d}(-2x_{n,i}w_jx_{n,j}) - 2s_nx_{n,i} + \sum_{j=1}^{d}(2s_nx_{n,i}w_j^2) + 2s_n^2 w_i$$

$$= -2x_{n,i}\sum_{j=1}^{d} w_j x_{n,j} - 2s_n x_{n,i} + 2x_{n,i}s_n \sum_{j=1}^{d} w_j^2 + 2s_n^2 w_i$$

$$= -2x_{n,i} W^T x_n - 2s_n x_{n,i} + 2x_{n,i}s_n W^T W + 2s_n^2 w_i$$

最後我們算 $\nabla_w err(w)$

$$\nabla_w err(w) = -2x_n w^T x_n - 2s_n x_n + 2x_n s_n w^T w + 2s_n^2 w$$
$$= -2x_n(w^T x_n) - 2(w^T x_n)x_n + 2x_n(w^T x_n)w^T w + 2(w^T x_n)^2 w$$
$$= 2(w^T x_n)^2 w - 4x_n(w^T x_n) + 2x_n(w^T x_n)w^T w \quad \#$$

8. First, simplisy $x_n - ww^T(x_n+\epsilon_n)$

$$x_n - ww^T(x_n+\epsilon_n) = \begin{bmatrix} x_{n,1} \\ x_{n,2} \\ \vdots \\ x_{n,d} \end{bmatrix} - \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix} [w_1, w_2, \dots w_d] \begin{bmatrix} x_{n,1}+\epsilon_{n,1} \\ x_{n,1}+\epsilon_{n,2} \\ \vdots \\ x_{n,d}+\epsilon_{n,d} \end{bmatrix}$$
$$= \begin{bmatrix} x_{n,1} - (s_n+t_n)w_1 \\ x_{n,2} - (s_n+t_n)w_2 \\ \vdots \\ x_{n,d} - (s_n+t_n)w_n \end{bmatrix}$$

Where $s_n = \sum_{i=1}^{d} w_i x_{n,i}$, $t_n = \sum_{i=1}^{d} w_i \epsilon_{n,i}$

$$|x - ww^T(x_n+\epsilon_n)|^2 = \sum_{i=1}^{d} (x_{n,i} - (s_n+t_n)w_i)^2$$
$$= \sum_{i=1}^{d} (x_{n,i}^2 - 2x_{n,i}(s_n+t_n)w_i + (s_n+t_n)^2 w_i^2)$$
$$= \sum_{i=1}^{d} (x_{n,i}^2 - 2s_n w_i x_{n,i} + s_n^2 w_i^2) + \sum_{i=1}^{d} (-2t_n w_n x_{n,i} + 2s_n t_n w_i^2 + t_n^2 w_i^2)$$
$$= |x_n - ww^T x_n|^2 + 2s_n t_n w^T w + t_n^2 w^T w - 2t_n w^T x_n$$
$$= |x_n - ww^T x_n|^2 + 2(w^T x_n)(w^T \epsilon_n)w^T w + (w^T \epsilon_n)^2 w^T w - 2(w^T \epsilon_n) w^T x_n$$

Thus, $E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N} |x_n - ww^T x_n|^2 + \frac{1}{N}\sum_{n=1}^{N} (2(w^T x_n)(w^T \epsilon_n)w^T w + (w^T \epsilon_n)^2 w^T w - 2(w^T \epsilon_n) w^T x_n)$

Notice that zero mean and unit variance.

$$\mathbb{E}(\epsilon_n) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0, \quad \mathbb{E}(\epsilon \epsilon_n^T) = I_n.$$

$$\Omega(w) = \mathbb{E}\left(\frac{1}{N}\sum_{n=1}^{N} (2(w^T x_n)(w^T \epsilon_n) w^T w + (w^T \epsilon_n)^2 w^T w - 2(w^T \epsilon_n) w^T x_n)\right)$$
$$= \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}(2(w^T x_n)(w^T \epsilon_n) w^T w + (w^T \epsilon_n)^2 w^T w - 2(w^T \epsilon_n) w^T x_n)$$
$$= \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}(w^T \epsilon_n \epsilon_n^T w w^T w)$$
$$= \frac{1}{N}\sum_{n=1}^{N} (w^T w)^2$$
$$= (w^T w)^2 \quad \#$$

9.

令 $U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1\tilde{d}} \\ u_{21} & u_{22} & & u_{2\tilde{d}} \\ \vdots & & \ddots & \vdots \\ u_{d1} & u_{d2} & \cdots & u_{d\tilde{d}} \end{bmatrix}$

$$E = \frac{1}{N}\sum_{n=1}^{N}(X - UU^TX)^2$$

$$UU^TX = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1\tilde{d}} \\ u_{21} & u_{22} & & u_{2\tilde{d}} \\ \vdots & & \ddots & \vdots \\ u_{d1} & u_{d2} & & u_{d\tilde{d}} \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & \cdots & u_{d1} \\ u_{12} & u_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ u_{1\tilde{d}} & u_{2\tilde{d}} & \cdots & u_{d\tilde{d}} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

$$= \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1\tilde{d}} \\ u_{21} & u_{22} & & u_{2\tilde{d}} \\ \vdots & & \ddots & \\ u_{d1} & u_{d2} & & u_{d\tilde{d}} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{d} u_{i1}x_i \\ \sum_{i=1}^{d} u_{i2}x_i \\ \vdots \\ \sum_{i=1}^{d} u_{i\tilde{d}}x_i \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{\tilde{d}}\left(u_{1i}\sum_{j=1}^{d} u_{ji}x_j\right) \\ \sum_{i=1}^{\tilde{d}}\left(u_{2i}\sum_{j=1}^{d} u_{ji}x_j\right) \\ \vdots \\ \sum_{i=1}^{\tilde{d}}\left(u_{di}\sum_{j=1}^{d} u_{ji}x_j\right) \end{bmatrix}$$

$$\Rightarrow E = \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{k=1}^{d}\left(\sum_{i=1}^{\tilde{d}} u_{ki}\sum_{j=1}^{d} u_{ji}x_{nj}\right) - x_{nk}\right)^2$$

10. First, calculate $\frac{\partial E_q(u)}{\partial u_{ab}}$

$$E_q = \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{k=1}^{d}\left(\sum_{i=1}^{\tilde{d}} u_{ki}\sum_{j=1}^{d} u_{ji}x_{n,j}\right)-x_{n,k}\right)^2$$

$$\frac{\partial E_q(u)}{\partial u_{ab}} = \frac{1}{N}\sum_{n=1}^{N}\left(2\left(\sum_{i=1}^{\tilde{d}} u_{ai}\sum_{j=1}^{d} u_{ja}x_{n,j}\right)-x_{n,a}\right)\left(\sum_{j=1}^{d} u_{jb}x_{n,j}\right)$$
$$+\sum_{k=1}^{d} 2\left(\sum_{i=1}^{\tilde{d}}\left(u_{ki}\sum_{j=1}^{d} u_{ji}x_{n,j}\right)-x_{n,k}\right)u_{kb}x_{n,a}\Bigg)$$

Rewrite E with w

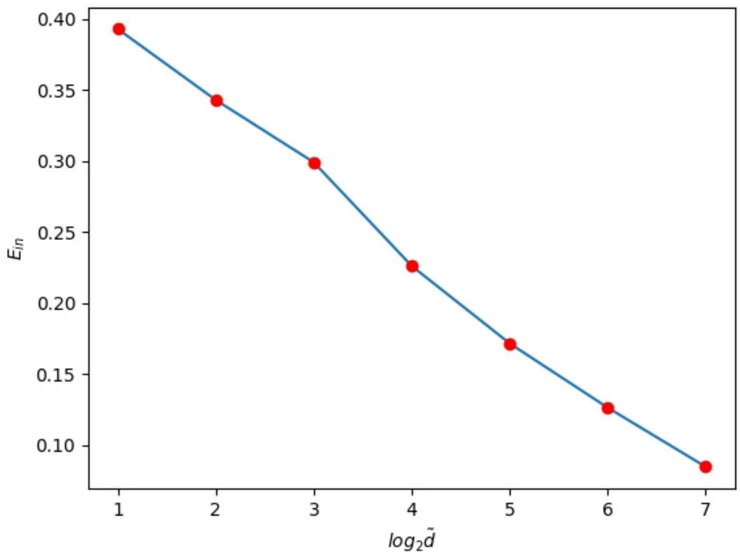
$$E_{10} = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{d}\left(\sum_{i=1}^{\tilde{d}}\left(w_{ik}^{(2)}\sum_{j=1}^{d} w_{ji}^{(1)}x_{n,j}\right)-x_{k,j}\right)^2$$

$$\frac{\partial E_{10}}{\partial w_{ab}^{(1)}} = \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{k=1}^{d} 2\left(\sum_{i=1}^{\tilde{d}}\left(w_{ik}^{(2)}\sum_{j=1}^{d} w_{ji}^{(1)}x_{n,j}\right)-x_{k,j}\right)w_{bk}^{(2)}x_{n,a}\right)$$

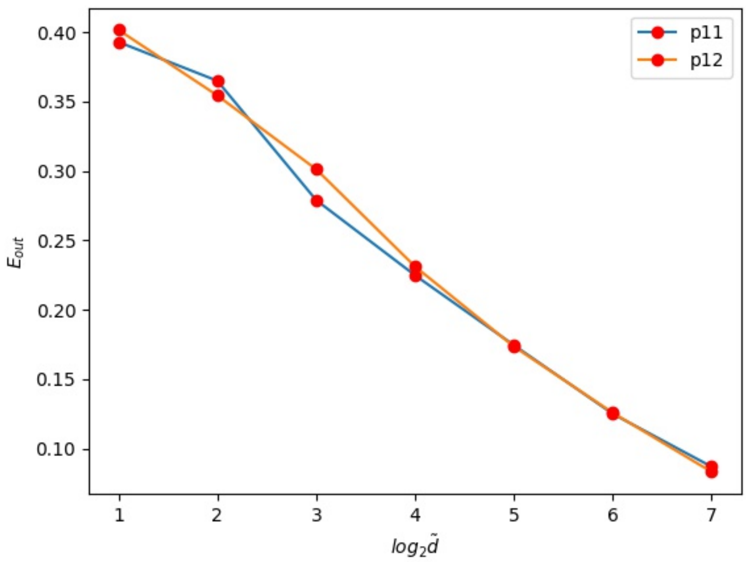
$$\frac{\partial E_{10}}{\partial w_{ba}^{(2)}} = \frac{1}{N}\sum_{n=1}^{N}\left(2\sum_{i=1}^{\tilde{d}}\left(w_{ia}^{(2)}\sum_{j=1}^{d} w_{ji}^{(1)}x_{n,j}\right)-x_{a,j}\right)\left(\sum_{j=1}^{d} w_{jb}^{(1)}x_{n,j}\right)$$

得 $\frac{\partial E_q(u)}{\partial u_{ab}} = \frac{\partial E_{10}(w)}{\partial w_{ab}^{(1)}} + \frac{\partial E_{10}(w)}{\partial w_{ba}^{(2)}}$

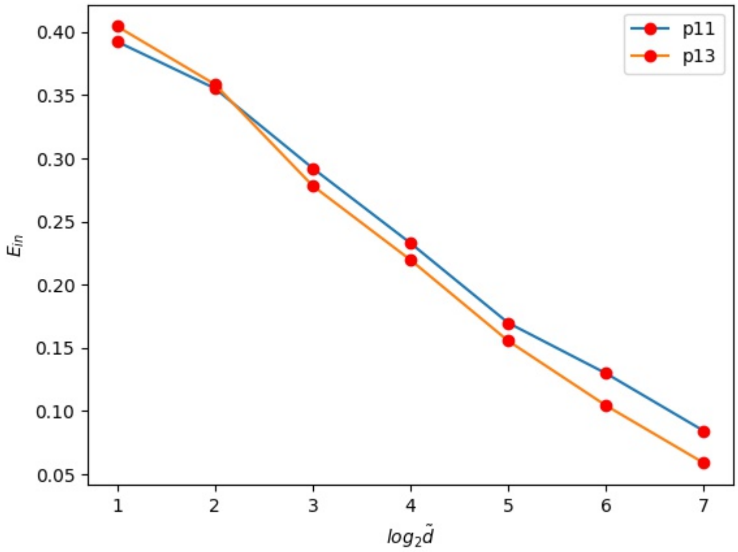
11. $\tilde{d}$越大，E_{in}會越小



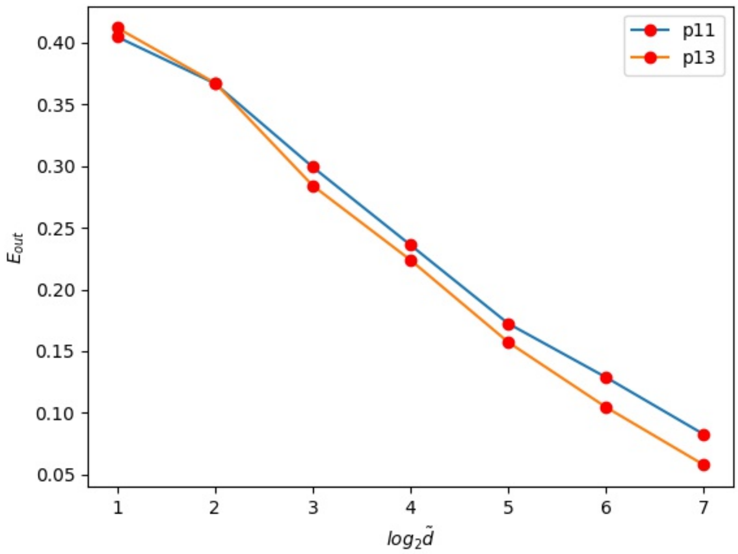
12. $\tilde{d}$越大，E_{out}會越小，但是E_{out}一開始會比E_{in}大一點點，然後會誰大誰小會互換兩次，最後 4 次的值會差不多。



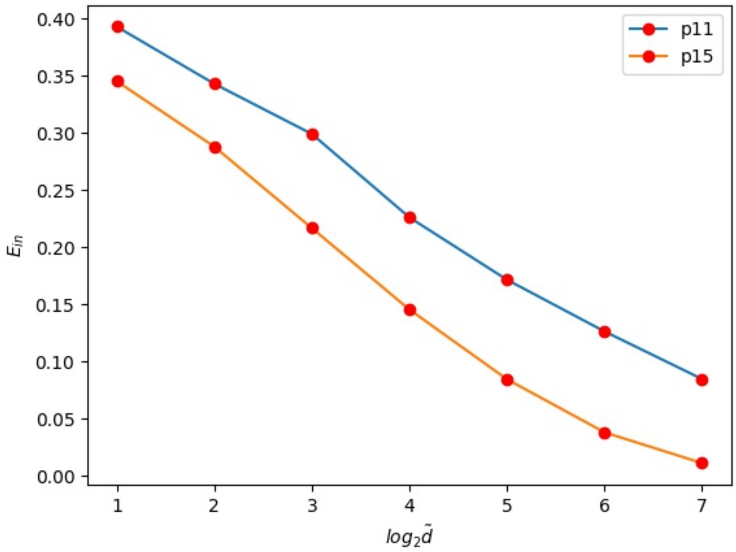
13. $\tilde{d}$越大，E_{in}會越小。Tied up weight 在$log_2\tilde{d} < 2$時，值會比 p11 稍稍高一點。$log_2\tilde{d} = 2$時，幾乎一樣。而$log_2\tilde{d} > 2$時，就會比較 p11 小一點點。



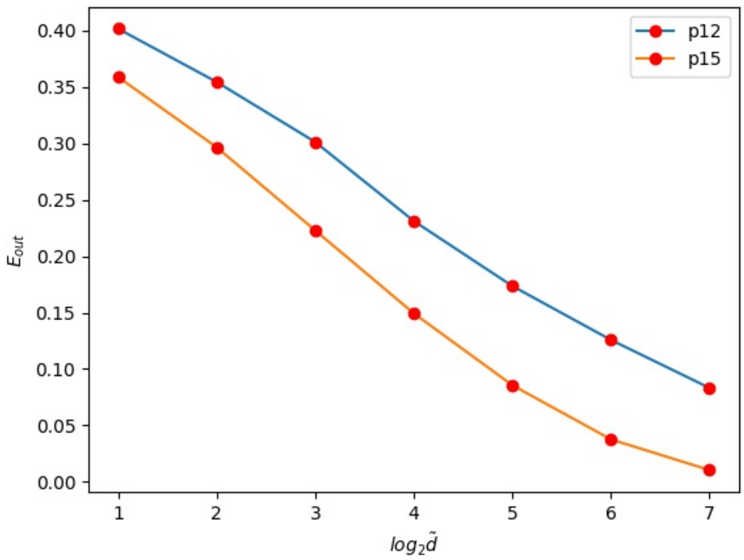
14. $\tilde{d}$越大，E_{out}會越小。Tied up weight 在$log_2\tilde{d} < 2$時，值會比 p11 稍稍高一點。$log_2\tilde{d} = 2$時，幾乎一樣。而$log_2\tilde{d} > 2$時，就會比較 p11 小一點點。



15. $\tilde{d}$越大，E_{in}會越小。也發現用 PCA 的E_{in}永遠會比 p11 的方法小



16. $\tilde{d}$越大，E_{out}會越小。也發現用 PCA 的E_{out}永遠會比 p12 的方法小



17.

18.

18. lemma 1

$\sum_{i=0}^{D}\binom{N}{i} \le N^D+1$

proof:

$D=0$, $\sum_{i=0}^{0}\binom{N}{i}=1 \le N^0+1=2$.

Assume that $\sum_{i=0}^{k}\binom{N}{i} \le N^k+1$ 成立.

當 $D=k+1$

$$\sum_{i=0}^{k+1}\binom{N}{i}=\sum_{i=0}^{k}\binom{N}{i}+\binom{N}{k+1} \le N^k+1+\binom{N}{k+1}$$

$$=N^k+1+\frac{N(N-1)\cdots(N-k)}{(k+1)!}$$

$$\le N^k+1+N(N-1)\cdots(N-k)$$

$$\le N^k+1+N^k(N-1)$$

$$=N^{k+1}+1$$

從數学歸纳法得證 #

$\because$ VC dimension of d-dimension perceptron is $d+1$,

$$m_{H3A}(N) \le (B(N,d+2))^3 \le \left(\sum_{i=1}^{d+1}\binom{N}{i}\right)^3 \le (N^{d+1}+1)^3$$

m_{H3A} is the growth function, $B(N,d+2)$ is the bounding function.

第一個不等式, bounding function 的定義

第二個不等式, MLF 的 lecture 6 p18.

第三個不等式, 上面的證明.

let $\Delta=3(d+1)+1 \ge 4$, $N \ge \Delta$

$$(N^{d+1}+1)^3=N^{3(d+1)}+3N^{2(d+1)}+3N^{d+1}+1 < N^{3(d+1)}+N^{3(d+1)}+N^{3(d+1)}+1=3N^{3(d+1)}+1 < N\cdot N^{3(d+1)}+1$$

and $N^{3(d+1)+1}+1=N^{\Delta}+1$

When $N \ge 3\Delta\log_2\Delta$, by p.17, we have.

$$m_{H3A}(N) \le (N^{(d+1)}+1)^3 < N^{\Delta}+1 < 2^N$$

代表無法 shatter 這 N 個點, 因此 H3A 的 VC dimension 小於 ?

$3\Delta\log_2\Delta \to 3(3(d+1)+1)\log_2 3(3(d+1)+1)$