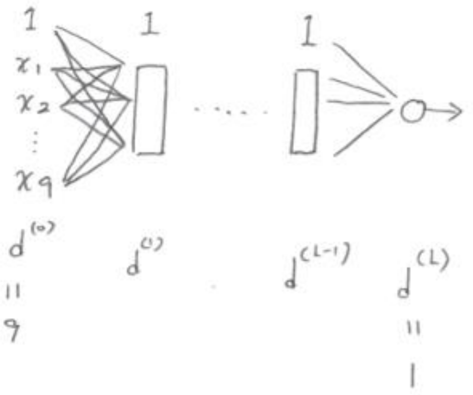
HW4 r07922003 劉濬慶

1. $d^{(0)}+1=10$, 36 hidden units, 1 output unit

$l = 1, \ldots, L-1 \in$ hidden layer

圖示：



$$\text{Weights} = \sum_{l=0}^{L-1} (d^{(l)}+1)\, d^{(l+1)}$$

$$\sum_{l=1}^{L-1} (d^{(l)}+1) = 36$$

$$\Rightarrow (d^{(1)}+1) + \cdots + (d^{(L-1)}+1) = 36$$

$$\Rightarrow d^{(1)} + d^{(2)} + \cdots + d^{(L-1)} = 37 - L$$

$$\sum_{l=0}^{L-1} (d^{(l)}+1) d^{(l+1)} = 10d^{(1)} + (d^{(1)}+1)d^{(2)} + \cdots + (d^{(L-2)}+1)d^{(L-1)} + (d^{(L-1)}+1)$$

$$= 9d^{(1)} + d^{(1)}d^{(2)} + \cdots + d^{(L-2)}d^{(L-1)} + d^{(1)} + d^{(2)} + \cdots + d^{(L-1)} + d^{(L-1)} + 1$$

$$= 9d^{(1)} + 38 - L + d^{(L-1)} + d^{(1)}d^{(2)} + \cdots + d^{(L-2)}d^{(L-1)}$$

由上述可知 L 越大越好，$d^{(1)} + d^{(2)} + \cdots + d^{(L-1)}$ 越小越好

$d^{(1)} = d^{(2)} = \cdots = d^{(L-1)} = 1$ 時 L 最大

$$\Rightarrow \text{weights} = 9 + 38 - L + 1 + L - 2$$

$$= \underline{46} \ \#$$

A: 46

2. 從 $L=1$ 開始討論

$d^{(0)}$ $d^{(1)}$ $d^{(output)}$

$9+1$ $35+1$ 1

$W = 10\times 35 + 36\times 1$

$= \underline{386}$

$L=2$ 時：

$d^{(0)}$ $d^{(1)}$ $d^{(2)}$ $d^{(output)}$

$9+1$ $x+1$ $34-x+1$ 1

$W = 10x + (x+1)(34-x) + (35-x)$

$= 69 + 42x - x^2$

$= 510 - (x-21)^2$

$x=21$ 時，最大值 $= 510$

$L=3$ 時：

$d^{(0)}$ $d^{(1)}$ $d^{(2)}$ $d^{(3)}$ $d^{(output)}$

$9+1$ $x+1$ $y+1$ $33-x-y+1$ 1

$W = 10x + (x+1)y + (y+1)(33-x-y) + (33-x-y+1)$

$= 8x + 32y - 67 - y^2$

$= 323 + 8x - (y-16)^2$

$(x,y,z) = (16,16,1) \Rightarrow 451$

$= (17,15,1) \Rightarrow 458$

$= (18,14,1) \Rightarrow 463$

$= (19,13,1) \Rightarrow 466$

$= (20,12,1) \Rightarrow 467$

$= (21,11,1) \Rightarrow 466$

L 越大，W maximum 越小

A: 最大值發生在 $L=2$，$W=510$

$d^{(0)}$ $d^{(1)}$ $d^{(2)}$ $d^{(output)}$

$9+1$ $21+1$ $13+1$ 1

#

3. $err_n(w) = |x_n - ww^T x_n|^2$

$$= x_n^T x_n - 2(w^T x_n)^2 + (w^T x_n)^2 (w^T w)$$

$$\nabla err_n(w) = -4(w^T x_n)(x_n) + 2w^T x_n (w^T w) \nabla(w^T x_n) + (w^T x_n)^2 \nabla(w^T w)$$

$$= \underline{-4 w^T x_n x_n + 2 w^T x_n w^T w x_n + 2(w^T x_n)^2 w} \ \#$$

4.

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

$$= \frac{1}{N}\sum_{n=1}^{N} (\underline{x_n - ww^T x_n} - ww^T \varepsilon_n)^T (\underline{x_n - ww^T x_n} - ww^T \varepsilon_n)$$

$$= \frac{1}{N}\sum_{n=1}^{N} |x_n - ww^T x_n|^2 + \frac{1}{N}\sum_{n=1}^{N}\left[(x_n - ww^T x_n)^T(-ww^T \varepsilon_n) - (ww^T \varepsilon_n)^T(x_n - ww^T x_n) + (ww^T \varepsilon_n)^T(ww^T \varepsilon_n)\right]$$

$$= \frac{1}{N}\sum_{n=1}^{N} |x_n - ww^T x_n|^2 + \frac{1}{N}\sum_{n=1}^{N}\left[(-x_n^T ww^T \varepsilon_n + x_n^T ww^T ww^T \varepsilon_n - \varepsilon_n^T ww^T x_n + \varepsilon_n^T ww^T ww^T x_n + \varepsilon_n^T ww^T ww^T \varepsilon_n\right]$$

$\because$ $x_n^T w w^T \varepsilon_n = \varepsilon_n^T w w^T x_n$; $x_n^T w w^T w w^T \varepsilon_n = \varepsilon_n^T w w^T w w^T x_n$

$$= \frac{1}{N}\sum_{n=1}^{N} |x_n - ww^T x_n|^2 + \frac{1}{N}\sum_{n=1}^{N}\left[\varepsilon_n^T ww^T ww^T \varepsilon_n - 2\varepsilon_n^T ww^T x_n + 2\varepsilon_n^T ww^T ww^T x_n\right]$$

expected $E_{in}(w)$ 中 $\varepsilon_n^T ww^T ww^T x_n$ 與 $\varepsilon_n^T ww^T ww^T x_n$ 其均值 $=0$

因 expected $E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N} |x_n - ww^T x_n|^2 + \frac{1}{N}\sum_{n=1}^{N} \underline{\varepsilon_n^T} ww^T w \underline{w^T \varepsilon_n}$ → 常數

$$= \frac{1}{N}\sum_{n=1}^{N} |x_n - ww^T x_n|^2 + \frac{1}{N}\sum_{n=1}^{N} w^T w w^T \underline{\varepsilon_n \varepsilon_n^T} w$$

$$= \frac{1}{N}\sum_{n=1}^{N} |x_n - ww^T x_n|^2 + \frac{1}{N}\sum_{n=1}^{N} w^T w w^T w$$

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

$$\underline{A \cdot \Omega(w) = (w^T w)^2}$$

5.

$$\underline{E = \frac{1}{d}\sum_{n=1}^{d}\left| x_n - UU^T x_n \right|^2} \quad \#$$

$U = [\, u_{ij} \,]$ of size $d \times \tilde{d}$

7.

$(x_+, 1), (x_-, -1)$, $w^T x + b$ 為通过此兩点的<u>中垂面</u>；中点 $\left(\frac{x_+, x_-}{2}, 0\right)$ 法向量 $\left(\frac{x_+ - x_-}{2}, 1\right)$

$$\left(\frac{x_+ - x_-}{2}\right)^T \left(x - \frac{x_+ + x_-}{2} \right) = 0$$

$$\left(\frac{x_+ - x_-}{2}\right)^T x - \frac{|x_+|^2 - |x_-|^2}{4} = 0 \Rightarrow (x_+ - x_-)^T x - \frac{|x_+|^2 - |x_-|^2}{2} = 0$$

$$g_{LIN}(x) = \underline{\mathrm{sign}\left((x^+ - x^-)^T x - \frac{|x_+|^2 - |x_-|^2}{2} \right)} \quad \#$$

8.

$g_{RBFNET}(x) = \mathrm{sign}\left(\beta_+ \exp(-|x - \mu_+|^2) + \beta_- \exp(-|x - \mu_-|^2) \right)$, $\beta_+ > 0 > \beta_-$

討論 $g_{RBFNET} = 1$ 之情況：

$$\beta_+ \exp(-|x - \mu_+|^2) > -\beta_- \exp(-|x - \mu_-|^2)$$

$$\ln \beta_+ - |x - \mu_+|^2 > \ln -\beta_- - |x - \mu_-|^2$$

$$\ln \frac{\beta_+}{-\beta_-} - |x - \mu_+|^2 + |x - \mu_-|^2 > 0$$

$$\ln \frac{\beta_+}{-\beta_-} - \cancel{|x|^2} + 2\mu_+^T x - |\mu_+|^2 + \cancel{|x|^2} - 2\mu_-^T x + |\mu_-|^2 > 0$$

$$\ln \frac{\beta_+}{-\beta_-} + 2(\mu_+ - \mu_-)^T x - |\mu_+|^2 + |\mu_-|^2 > 0$$

反之 $g_{RBFNET} = -1$ 之情況相似

$$g_{LIN}(x) = \mathrm{sign}\left(\underbrace{2(\mu_+ - \mu_-)^T}_{w} x + \underbrace{\ln \frac{\beta_+}{-\beta_-} - |\mu_+|^2 + |\mu_-|^2}_{b} \right) \quad \#$$

9. $\hat{d}=1$; $V = \hat{d} \times N = 1 \times N = [\overbrace{1\ 1 \cdots 1}^{N}]$ —— ①

$$\min_{W,V} E_{in}(\{W_m\},\{V_n\}) = \sum_{m=1}^{M}\Big(\sum_{(x_n,r_{nm})\in D_m}(r_{nm} - W_m^T V_n)^2\Big)$$

∵ optimal W_m 出現在 梯度 $=0$

$$\therefore\ \nabla_{W_m}\Big(\sum_{(x_n,r_{nm})\in D_m}(r_{nm} - W_m^T V_n)^2\Big) = 0$$

$$\sum_{(x_n,r_{nm})\in D_m} 2(r_{nm} - W_m^T V_n)(-V_n) = 0$$

∵ ① ∴ $\hookrightarrow \displaystyle\sum_{(x_n,r_{nm})\in D_m} 2(r_{nm} - W_m) = 0$

$$\Rightarrow \sum_{(x_n,r_{nm})\in D_m} r_{nm} = \sum_{(x_n,r_{nm})\in D_m} W_m = W_m \sum_{(x_n,r_{nm})\in D_m} 1$$

$$\Rightarrow W_m = \frac{\sum_{(x_n,r_{nm})\in D_m} r_{nm}}{\sum_{(x_n,r_{nm})\in D_{nm}} 1}$$ ← average rating of the m-th movie

#

10. $r_{nm} = V_n^T W_m$, new user $(N+1)$, $V_{N+1} = \frac{1}{N}\sum_{n=1}^{N} V_n$

推薦給新 user 电影為：

$V_{N+1}^T W_m = \frac{1}{N}\sum_{n=1}^{N} V_n^T W_m = \frac{1}{N}\sum_{n=1}^{N} r_{nm}$ = 平均前 N 了 user 的平均分數，推薦最高分的

17. $\Delta \geq 2$, if $N \geq 3\Delta\log_2\Delta$, $N^{\Delta}+1 < 2^{N}$

$3\Delta\log_2\Delta \leq N$

$\Rightarrow (3\Delta\log_2\Delta)^{\Delta}+1 \leq N^{\Delta}+1 < 2^{3\Delta\log_2\Delta} = 2^{\log_2\Delta^{3\Delta}} = \Delta^{3\Delta}$

$\Rightarrow 3^{\Delta}\cdot\Delta^{\Delta}\cdot\log_2\Delta+1 < \Delta^{\Delta}\cdot\Delta^{\Delta}\cdot\Delta^{\Delta}$

$\Rightarrow 3^{\Delta}\cdot\log_2\Delta+1 < \Delta^{\Delta}\cdot\Delta^{\Delta}$

$\Delta=2$, $3^{2}\cdot\log_2 2 = 9 < 2^{2}\cdot 2^{2} = 16$

$\Delta=3$; $3^{3}\cdot\underline{\log_2 3} < 3^{3}\cdot\underline{3}^{3}$

$\Delta=4$, $3^{4}\cdot\log_2 4 < 4^{4}\cdot 4^{4}$

$\downarrow$

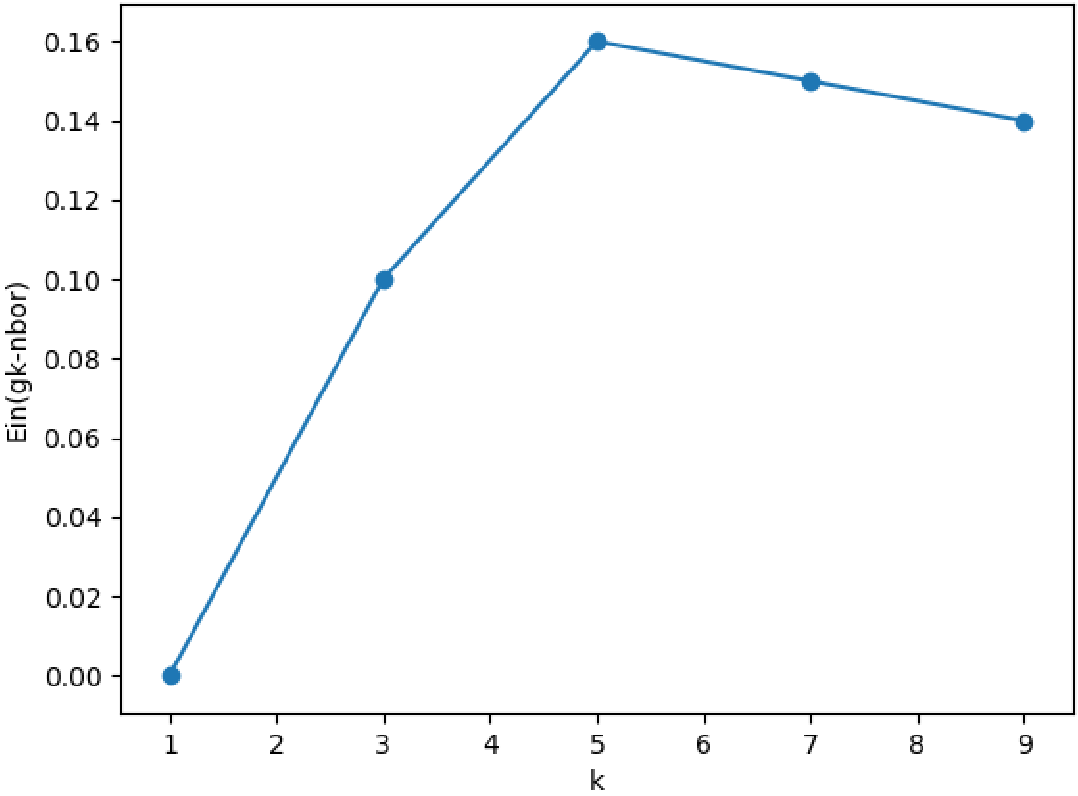
顯而易見 $3^{\Delta}\cdot\log_2\Delta+1 \ll \Delta^{2\Delta}$

$\Rightarrow 3^{\Delta}\cdot\log_2\Delta\cdot\Delta^{\Delta}+1 \ll \Delta^{2\Delta}\cdot\Delta^{\Delta}$

$\Rightarrow (3\Delta\log_2\Delta)^{\Delta}+1 \ll \Delta^{3\Delta}$ — 得證

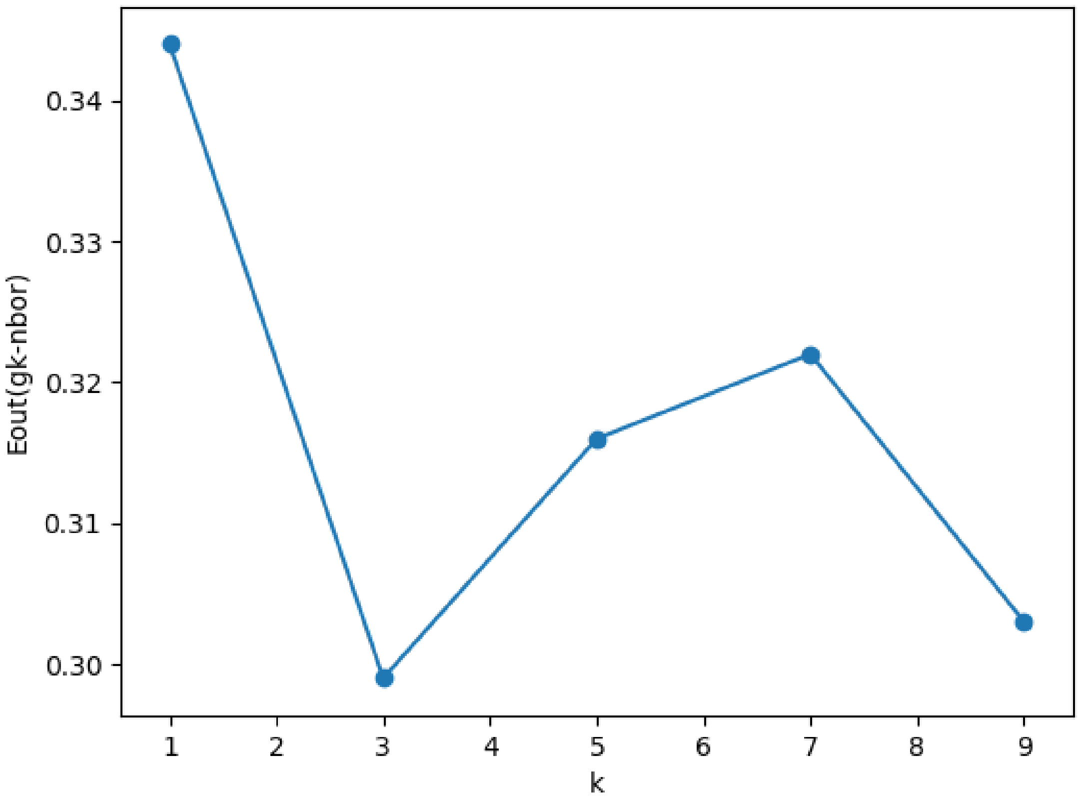
$(3\Delta\log_2\Delta)^{\Delta}+1 < 2^{N}$ 成立

Q11



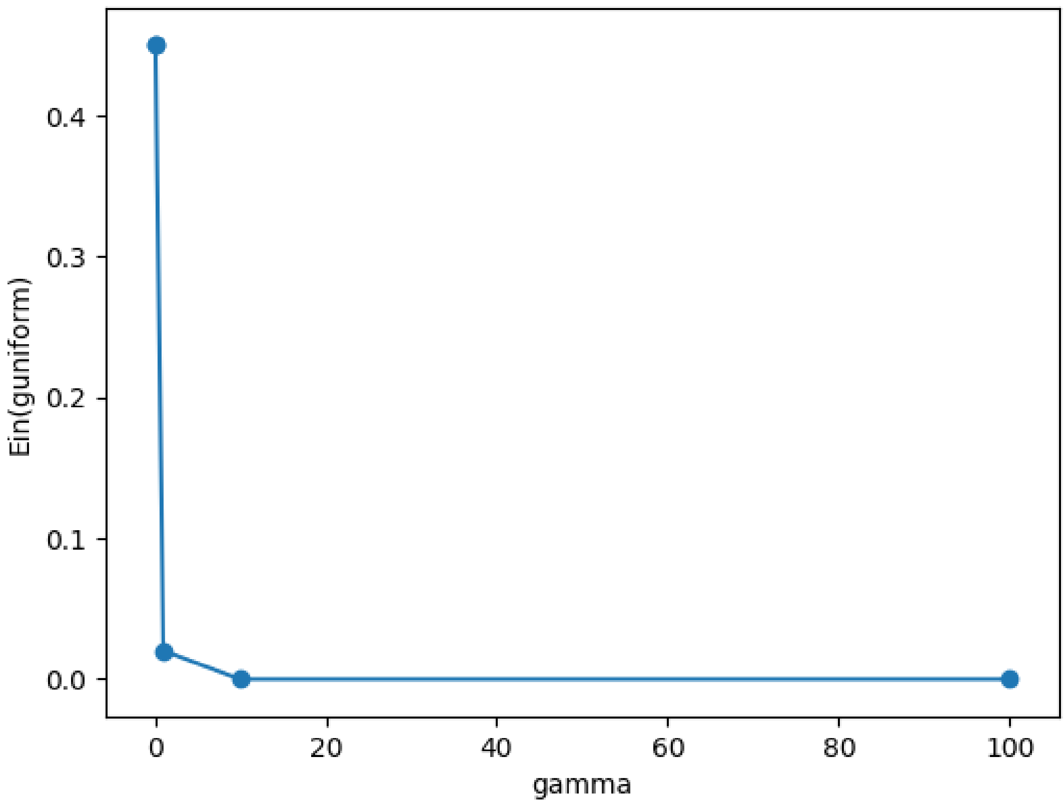
當 K=1 時，會全對因為參考自己就等於答案

Q12



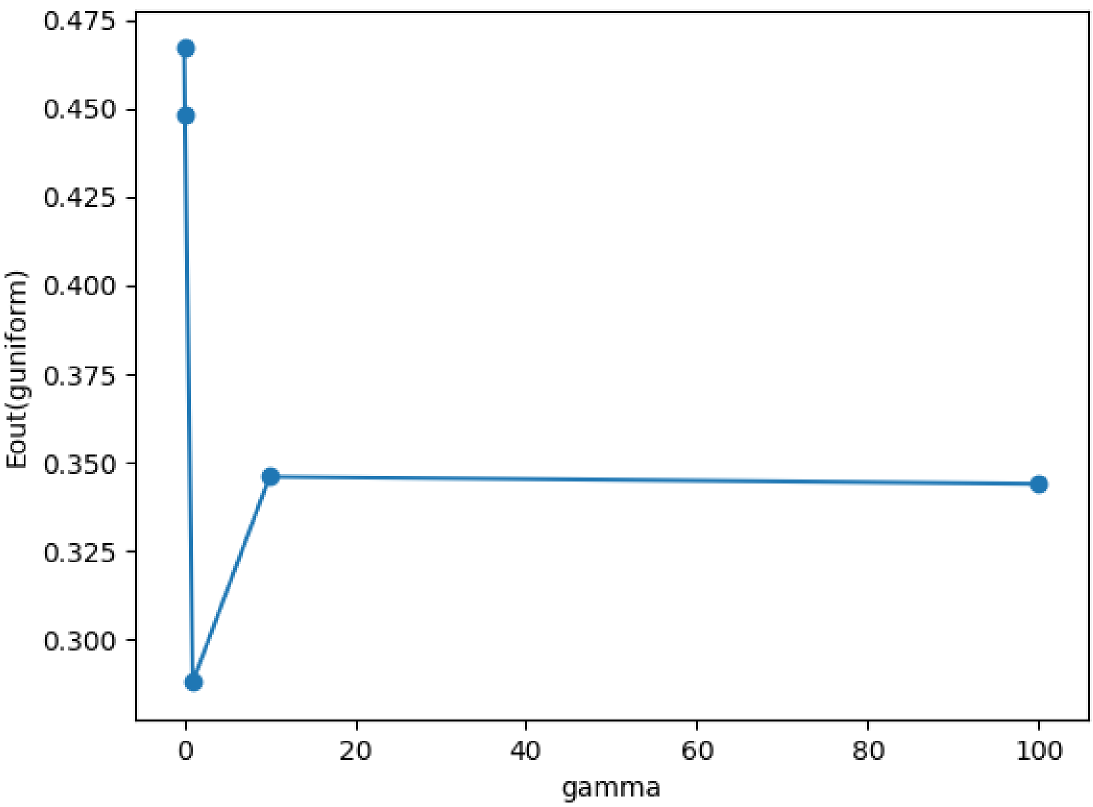
當 K=3 時最準確，也並不是參考得越多越準確

- Q13



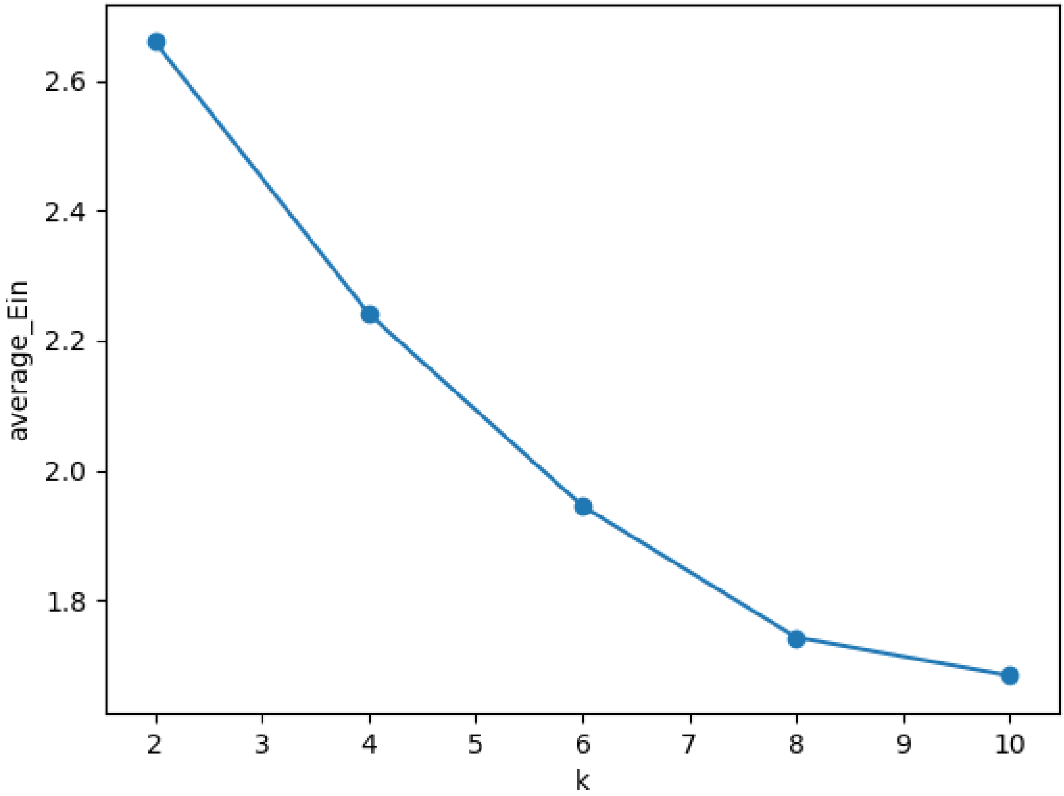
Gamma 越高越 error 越小

- Q14



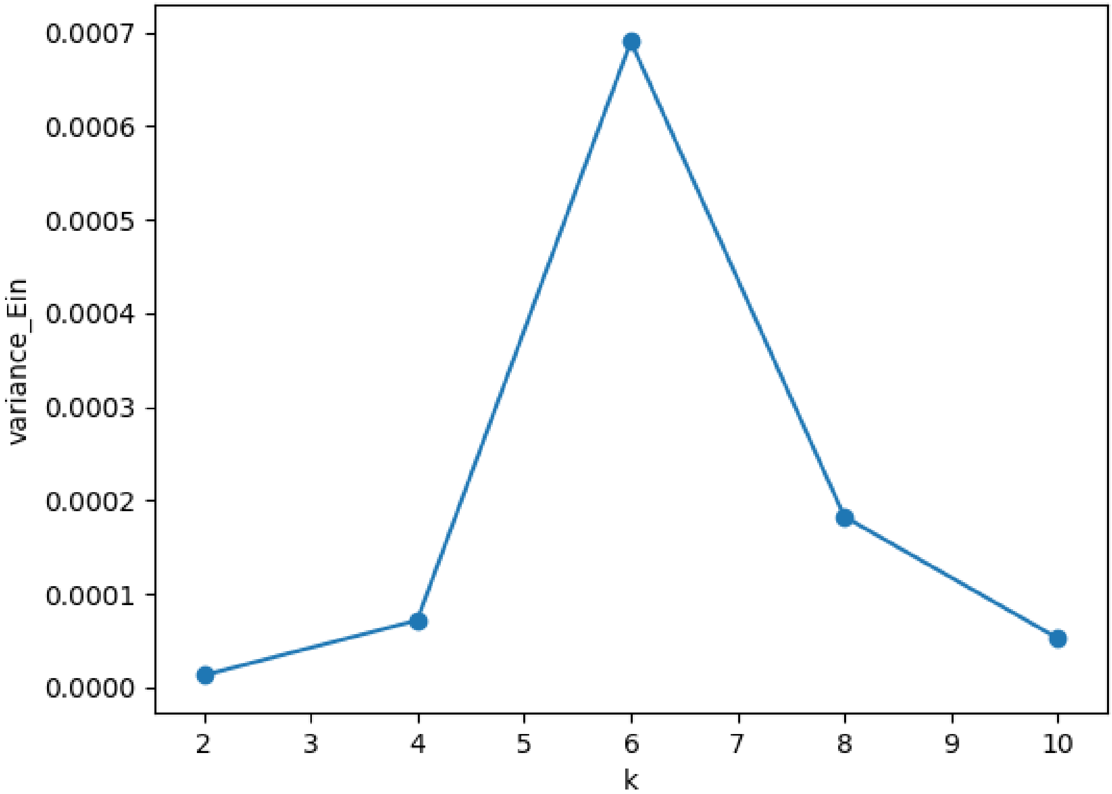
Gamma=1 時 error 最小，並非像 Ein 時 Gamma 很高 error 也小

Q15



當 K 越大時，average 越小

Q16



K 在很小跟很大時，variance 是小的