HW4 r12631001 許喬淇

Problem 1

a) Why did they obtain different results?

原因為 GDP 數值單位不同。Joe 以百萬美元為單位，他老闆以十億美元為單位，而十億美元的 GDP 數值比百萬美元的數值大了 1000 倍，對於變異數的影響也較大。

b) How should Joe and his Boss have treated the data before the PCA?

對資料進行標準化或統一單位。

Problem 2

(a) given: $[(\alpha,\alpha),(-\alpha,-\alpha),(-\beta,\beta),(\beta,-\beta)]$

$$\mathrm{cov}(\vec{X}) = \frac{1}{N-1}\sum_{i=1}^{N}(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)$$

$\bar{x}_1 = 0$, $\bar{x}_2 = 0$

$$\mathrm{cov}(x_1,x_2) = \frac{1}{3}\left[\alpha\cdot\alpha + (-\alpha)\cdot(-\alpha) + (-\beta)\cdot\beta + \beta\cdot(-\beta)\right]$$

$$= \frac{1}{3}(2\alpha^2 - 2\beta^2) \Rightarrow \mathrm{cov}(x_1,x_1)、\mathrm{cov}(x_2,x_2) \text{ 同理}$$

$$\Rightarrow \mathrm{cov}(\vec{x}) = \frac{1}{3}\begin{bmatrix} 2(\alpha^2+\beta^2) & 2(\alpha^2-\beta^2) \\ 2(\alpha^2-\beta^2) & 2(\alpha^2+\beta^2)\end{bmatrix}$$

(b)

$$\left|\mathrm{cov}(\vec{x}) - \lambda I\right| = \left|\mathrm{cov}(\vec{x}) - \lambda\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right|$$

$$= \begin{vmatrix} \frac{2}{3}(\alpha^2+\beta^2)-\lambda & \frac{2}{3}(\alpha^2-\beta^2) \\ \frac{2}{3}(\alpha^2-\beta^2) & \frac{2}{3}(\alpha^2+\beta^2)-\lambda \end{vmatrix}$$

$$= \cancel{\frac{4}{9}\alpha^4} + \cancel{\frac{4}{9}\beta^4} + \lambda^2 + \frac{8}{9}\alpha^2\beta^2 - \frac{4}{3}\alpha^2\lambda - \frac{4}{3}\beta^2\lambda - \left(\cancel{\frac{4}{9}\alpha^4} - \frac{8}{9}\alpha^2\beta^2 + \cancel{\frac{4}{9}\beta^4}\right)$$

$$= \lambda^2 + \frac{16}{9}\alpha^2\beta^2 - \frac{4}{3}\alpha^2\lambda - \frac{4}{3}\beta^2\lambda = 0$$

$$\Rightarrow 9\lambda^2 + 16\alpha^2\beta^2 - 12\alpha^2\lambda - 12\beta^2\lambda = 0$$

$$\lambda = \frac{(12\alpha^2+12\beta^2) \pm \sqrt{(12\alpha^2+12\beta^2)^2 - 36\cdot16\alpha^2\beta^2}}{18}$$

$$= \frac{2}{3}(\alpha^2+\beta^2) \pm \frac{\sqrt{144(\alpha^2+2\alpha\beta+\beta^2) - 576\alpha^2\beta^2}}{18}$$

$$= \frac{2}{3}(\alpha^2+\beta^2) \pm \frac{2}{3}\sqrt{\alpha^4-2\alpha^2\beta^2+\beta^4}$$

$$= \frac{2}{3}(\alpha^2+\beta^2) \pm \frac{2}{3}(\alpha^2-\beta^2) \Rightarrow \begin{cases} \lambda_1 = \frac{4}{3}\alpha^2 \\ \lambda_2 = \frac{4}{3}\beta^2 \end{cases}$$

① $\lambda_1 = \frac{4}{3}\alpha^2$

$$\begin{vmatrix} \frac{2}{3}(\alpha^2+\beta^2)-\frac{4}{3}\alpha^2 & \frac{2}{3}(\alpha^2-\beta^2) \\ \frac{2}{3}(\alpha^2-\beta^2) & \frac{2}{3}(\alpha^2+\beta^2)-\frac{4}{3}\alpha^2 \end{vmatrix} \vec{V}_1 = 0$$

$$\begin{vmatrix} -\frac{2}{3}\alpha^2+\frac{2}{3}\beta^2 & \frac{2}{3}(\alpha^2-\beta^2) \\ \frac{2}{3}(\alpha^2-\beta^2) & -\frac{2}{3}\alpha^2+\frac{2}{3}\beta^2 \end{vmatrix} \rightarrow \begin{vmatrix} -\frac{2}{3}\alpha^2+\frac{2}{3}\beta^2 & \frac{2}{3}\alpha^2-\frac{2}{3}\beta^2 \\ 0 & 0 \end{vmatrix}$$

$$\begin{pmatrix} -\frac{2}{3}\alpha^2+\frac{2}{3}\beta^2 & \frac{2}{3}\alpha^2-\frac{2}{3}\beta^2 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0$$

$$\vec{V}_1 = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

② $\lambda_2 = \frac{4}{3}b^2$

$$\begin{vmatrix} \frac{2}{3}(\alpha^2+\beta^2)-\frac{4}{3}\beta^2 & \frac{2}{3}(\alpha^2-\beta^2) \\ \frac{2}{3}(\alpha^2-\beta^2) & \frac{2}{3}(\alpha^2+\beta^2)-\frac{4}{3}\beta^2 \end{vmatrix} \vec{V}_2 = 0$$

$$\begin{vmatrix} \frac{2}{3}\alpha^2-\frac{2}{3}\beta^2 & \frac{2}{3}\alpha^2-\frac{2}{3}\beta^2 \\ \frac{2}{3}\alpha^2-\frac{2}{3}\beta^2 & \frac{2}{3}\alpha^2-\frac{2}{3}\beta^2 \end{vmatrix} \rightarrow \begin{vmatrix} \frac{2}{3}\alpha^2-\frac{2}{3}\beta^2 & \frac{2}{3}\alpha^2-\frac{2}{3}\beta^2 \\ 0 & 0 \end{vmatrix}$$

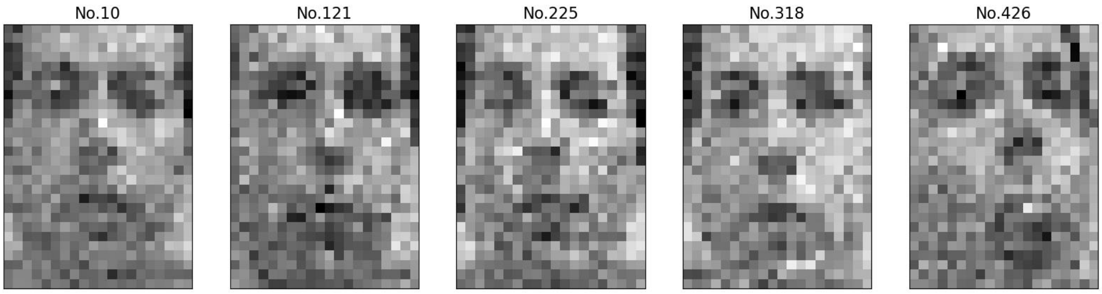
$$\begin{pmatrix} \frac{2}{3}\alpha^2-\frac{2}{3}\beta^2 & \frac{2}{3}\alpha^2-\frac{2}{3}\beta^2 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0$$

$$\vec{V}_2 = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

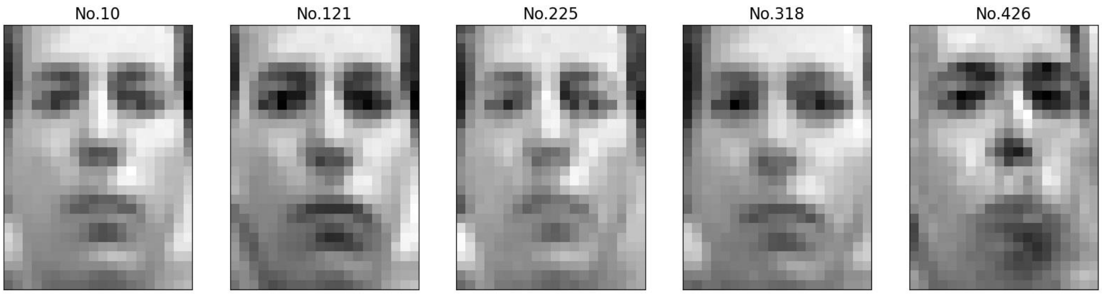
Problem 3

a) Apply PCA to the noisy data. Suppose the correct intrinsic dimensionality of the data is 10 (i.e., only the first 10 eigenvalues and eigenvectors contain information; the rest eigenvalues and eigenvectors contain noise). Compute reconstructed images using the top 10 eigenvectors and plot the 10th, 121st, 225th, 318th, and 426th original and reconstructed images.

根據下圖二可以看到，取 10 個 component 做訓練可以得到較清晰的臉部五官。



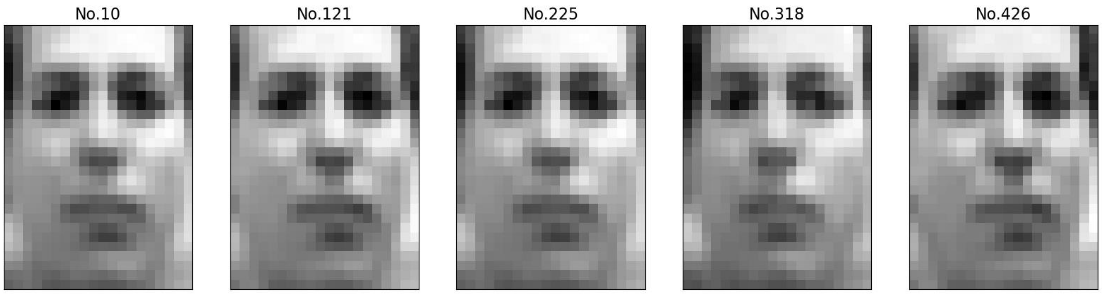
圖一 原始影像



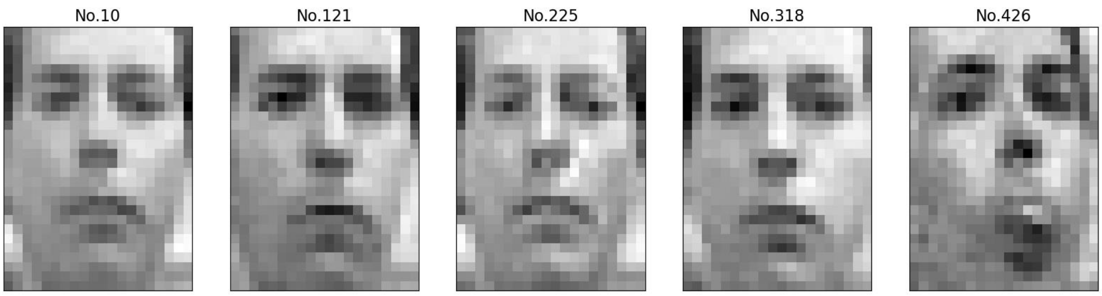
圖二 component = 10

b) Repeat part a), assuming the correct intrinsic dimensionality of the data is 2 and 30.

比較圖三和圖四，發現圖三的臉部眼窩較暗且唇部、眉毛線條較不清楚，而圖四的臉部肌膚顏色較不均勻、五官線條則較明顯。



圖三 component = 2



圖四 component = 30

c) **Determine the best intrinsic dimensionality of the dataset using the techniques learned from the class. You will need to check all the possible dimensionality. Explain the approaches you use and the reasons of choosing the best intrinsic dimensionality in details.**

藍線為每個 eigenvalue 所佔的比例，bar 為累積的數值，可以發現 PC1~PC5 為影像的主要 PC，而 PC5 後之 PC 較可能為 noise。