CS189 HW4

Hanze Yao
3033093286

2. Gaussian Features

(a) The code is shown in the appendix.
For this part, the code includes the function polynomial_augmentation in lib.py and part_a() in main.py
The plots are also shown in the appendix with the output.
From what I observe of the fitted curve, as the increase of polynomial degree, the curve fits more and more to the data points. This can also be seen from the decrease of least square error.

(b) The code is shown in the appendix
For this part, the code includes the function rbf_augmentation in lib.py and part_b() in main.py
The plots are also shown in the appendix with the output.
I choose $C_1 = -2$ $C_2 = 1$ $C_3 = 3$
Variance: 3, 2, 0.5
The least square error is 8.706.

3. Probabilistic Principal Components Analysis (PPCA)

(a) $y = Wx + \mu + Z$

mean $(y \mid x) = Wx + \mu$

Var $(y \mid x) = \Sigma = \sigma^2 I$

$\because$ This is still a gaussian distribution

$\therefore y \mid x \sim N(Wx + \mu, \sigma^2 I)$

(b) $\because x \sim N(0, I)$

$\therefore Wx \sim N(0, WW')$

$\because Z \sim N(0, \sigma^2 I)$

$\therefore y = Wx + \mu + Z \sim N(\mu, WW' + \sigma^2 I)$

4. Canonical Correlation Analysis

(a) Because this question needs to find a linear structure from paired data (X, Y). And Y is also n draws from random variable y. We care about the variation of Y too.

(b) $P = u^T X \quad Q = v^T y$

$$\rho(P, Q) = \rho(u^T x, v^T y) = \frac{Cov(u^T x, v^T y)}{\sqrt{Var(u^T x)\, Var(v^T y)}}$$

$$= \frac{E((u^T x - E(u^T x))(v^T y - E(v^T y)))}{\sqrt{Cov(u^T x, u^T x)\, Cov(v^T y, v^T y)}}$$

$$= \frac{E(u^T(x - E(x))(y - E(y))^T v)}{\sqrt{u^T E(xx^T) u v^T E(yy^T) v}}$$

$$= \frac{u^T E(xy^T) v}{\sqrt{u^T E(xx^T) u v^T E(yy^T) v}}$$

$$E(xy^T) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i^T = \frac{1}{n} X^T Y$$

$$E(xx^T) = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T = \frac{1}{n} X^T X$$

$$E(yy^T) = \frac{1}{n}\sum_{i=1}^{n} y_i y_i^T = \frac{1}{n} Y^T Y$$

$$= \frac{\frac{1}{n} u^T X^T Y v}{\sqrt{\frac{1}{n^2} u^T X^T X u v^T Y^T Y v}} = \frac{u^T X^T Y v}{\sqrt{u^T X^T X u v^T Y^T Y v}}$$

(c) After scaling the features of matrix X to have values between -1 and 1, this will not change the correlation coefficient.

Because: $\rho(u^T a x, v^T y)$

$$= \frac{a u^T E(xy^T) v}{\sqrt{a^2 u^T E(xx^T) u \cdot v^T E(yy^T) v}} = \frac{u^T E(xy^T) v}{\sqrt{u^T E(xx^T) u v^T E(yy^T) v}}$$

$$= \rho(u^T x, v^T y)$$

5. Total Least Squares

(a) $\because$ rank $(X + \mathcal{E}_x) = d$

$$\therefore X + \mathcal{E}_x = U \begin{bmatrix} s_1 & & & & \\ & \ddots & & & \\ & & s_d & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} V'$$

$$\because y + \mathcal{E}_y = (X + \mathcal{E}_x) w = U \begin{bmatrix} s_1 & & & & \\ & \ddots & & & \\ & & s_d & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} V' w = \begin{bmatrix} y_1 \\ \vdots \\ y_d \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\therefore [X + \mathcal{E}_x, y + \mathcal{E}_y] = \left(U \begin{bmatrix} s_1 & & & & \\ & \ddots & & & \\ & & s_d & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} V', \begin{bmatrix} y_1 \\ \vdots \\ y_d \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right)$$

This will be the null space of $[X + \mathcal{E}_x, y + \mathcal{E}_y]$

And the first d rows will still be linearly independent, with the addition of only one column.

$\therefore$ rank $([X + \mathcal{E}_x, y + \mathcal{E}_y]) = d$

(b)
$$\underset{n \times (d+1)}{[X, y]} = \underset{n \times (d+1)}{\begin{pmatrix} U_{xx} & U_{xy} \\ U_{yx}^T & U_{yy} \end{pmatrix}} \underset{(d+1)(d+1)}{\begin{pmatrix} \Sigma_d & \\ & \sigma_{d+1} \end{pmatrix}} \underset{(d+1)(d+1)}{\begin{pmatrix} V_{xx} & V_{xy} \\ V_{yx}^T & V_{yy} \end{pmatrix}^T}$$

$$= \begin{pmatrix} U_{xx} \Sigma_d & U_{xy} \sigma_{d+1} \\ U_{yx}^T \Sigma_d & U_{yy} \sigma_{d+1} \end{pmatrix} \begin{pmatrix} V_{xx}^T & V_{yx} \\ V_{xy}^T & V_{yy} \end{pmatrix}$$

$$[X + \mathcal{E}_x, y + \mathcal{E}_y] = \begin{pmatrix} U_{xx} & U_{xy} \\ U_{yx}^T & U_{yy} \end{pmatrix} \begin{pmatrix} \Sigma_d & \\ & 0 \end{pmatrix} \begin{pmatrix} V_{xy} & V_{xy} \\ V_{yx}^T & V_{yy} \end{pmatrix}^T$$

$$= \begin{pmatrix} U_{xx} \Sigma_d & 0 \\ U_{yx}^T \Sigma_d & 0 \end{pmatrix} \begin{pmatrix} V_{xx}^T & V_{yx} \\ V_{xy}^T & V_{yy} \end{pmatrix}$$

$$[X, Y] = \begin{pmatrix} U_{xx}\Sigma_d V_{xx}^T + U_{xy}\sigma_{d+1}V_{xy}^T & U_{xx}\Sigma_d V_{yx} + U_{xy}\sigma_{d+1}V_{yy} \\ U_{yx}^T\Sigma_d V_{xx}^T + U_{yy}\sigma_{d+1}V_{xy}^T & U_{yx}^T\Sigma_d V_{yx} + U_{yy}\sigma_{d+1}V_{yy} \end{pmatrix}$$

$$[X+E_x, Y+E_y] = \begin{bmatrix} U_{xx}\Sigma_d V_{xx}^T & U_{xx}\Sigma_d V_{yx} \\ U_{yx}^T\Sigma_d V_{xx}^T & U_{yx}^T\Sigma_d V_{yx} \end{bmatrix}$$

$$\therefore [E_x, E_y] = \begin{pmatrix} -U_{xy}\sigma_{d+1}V_{xy}^T & -U_{xy}\sigma_{d+1}V_{yy} \\ -U_{yy}\sigma_{d+1}V_{xy}^T & -U_{yy}\sigma_{d+1}V_{yy} \end{pmatrix}$$

$$= -\begin{pmatrix} U_{xy} \\ U_{yy} \end{pmatrix} \sigma_{d+1} \begin{pmatrix} V_{xy} \\ V_{yy} \end{pmatrix}^T$$

(c) $[X+E_x, Y+E_y]\begin{pmatrix} w \\ -1 \end{pmatrix} = 0$

$$[X+E_x \quad Y+E_y]\begin{pmatrix} w \\ -1 \end{pmatrix} = \begin{pmatrix} U_{xx}\Sigma_d V_{xx}^T & U_{xx}\Sigma_d V_{yx} \\ U_{yx}\Sigma_d V_{xx}^T & U_{yx}^T\Sigma_d V_{yx} \end{pmatrix}\begin{pmatrix} w \\ -1 \end{pmatrix}$$

$$= \begin{pmatrix} U_{xx}\Sigma_d V_{xx}^T w - U_{xx}\Sigma_d V_{yx} \\ U_{yx}^T\Sigma_d V_{xx}^T w - U_{yx}^T\Sigma_d V_{yx} \end{pmatrix} = 0$$

$$\therefore \begin{cases} U_{xx}\Sigma_d V_{xx}^T w = U_{xx}\Sigma_d V_{yx} \\ U_{yx}^T\Sigma_d V_{xx}^T w = U_{yx}^T\Sigma_d V_{yx} \end{cases}$$

$$w = (V_{xx}^T)^{-1} V_{yx}$$

(d) $\because \begin{pmatrix} w \\ -1 \end{pmatrix}$ is a right singular vector of $[X, Y]$ with singular value σ_{d+1}

$\therefore \begin{pmatrix} w \\ -1 \end{pmatrix}$ is an eigenvector of $[X, Y]^T[X, Y]$ with eigenvalue σ_{d+1}^2

$$\therefore \; [X, y]^T [X, y] \begin{pmatrix} w \\ -1 \end{pmatrix} = \sigma_{d+1}^2 \begin{pmatrix} w \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} X^TX & X^Ty \\ y^TX & y^Ty \end{pmatrix} \begin{pmatrix} w \\ -1 \end{pmatrix} = \sigma_{d+1}^2 \begin{pmatrix} w \\ -1 \end{pmatrix}$$

$$X^TX w - X^Ty = \sigma_{d+1}^2 w$$

$$(X^TX - \sigma_{d+1}^2 I) w = X^Ty$$

(e) The code, coefficienes and plot are shown in the appendix

(f) The code, coeffitienes and plot are shown in the appendix

(g) The code, coeffivenks and plot are shown in the appendix

The scale factor I use is 384.