Y. Sarahi García González
Reconocimiento estadístico de patrones
Tarea 1b

Problema 2.

Sea $\{X_i\}$ un conjunto de n vectores d-dimensional.

Definimos las matrices $n\times n$: $[K_{ij}]_{n\times n}$ $\quad K_{ij} = \langle X_i, X_j \rangle$

$[D^2_{ij}]_{n\times n}$ $\quad D^2_{ij} = \|X_i - X_j\|^2$

Verifica: $D^2 = c\,1^t + 1\,c^t - 2XX^t$

con $c^t = (K_{11}, K_{22}, K_{33} \dots K_{nn})$ y $1^t = (1, \dots 1)$

solución:

$$\|X_i - X_j\|^2 = \langle X_i - X_j , X_i - X_j \rangle$$
$$= (X_i - X_j)(X_i - X_j)$$
$$= \langle X_i, X_i \rangle + \langle X_j , X_j \rangle - \langle X_i , X_j \rangle - \langle X_j , X_i \rangle$$
$$= \langle X_i, X_i \rangle + \langle X_j, X_j \rangle - 2\langle X_i, X_j \rangle$$

$\Rightarrow$ La entrada i,j de la matriz $\mathbb{D}$ puede escribirse como:

$$\mathbb{D}_{ij} = \langle X_i, X_i \rangle + \langle X_j, X_j \rangle - 2\langle X_i, X_j \rangle \qquad (\alpha)$$

Por otro lado la matriz $c\,1^t$ se ve cómo:

$$\begin{pmatrix} \langle X_1, X_1 \rangle \\ \langle X_2, X_2 \rangle \\ \vdots \\ \langle X_n, X_n \rangle \end{pmatrix}_{n\times 1} (1, 1, \dots 1)_{1\times n} = \begin{pmatrix} \langle X_1, X_1 \rangle & \langle X_1, X_1 \rangle & \dots \\ \langle X_2, X_2 \rangle & \langle X_2, X_2 \rangle & \dots \\ \vdots & \vdots & \\ \langle X_2, X_2 \rangle & \langle X_2, X_2 \rangle & \dots \end{pmatrix}_{n\times n}$$

$\Rightarrow$ en general, la entrada ij : $\underbrace{[c\mathbb{1}^t]_{ij} = \langle x_i, x_i \rangle}_{(1)}$

la matriz $\mathbb{1}c^t$, cómo:

$$\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{n\times 1} \left(\langle x_1, x_1 \rangle, \langle x_2, x_2 \rangle, \dots \langle x_n, x_n \rangle \right)_{1\times n}$$

$$= \begin{pmatrix} \langle x_1, x_1 \rangle & \langle x_2, x_2 \rangle & \dots & \langle x_n, x_n \rangle \\ \langle x_1, x_1 \rangle & \langle x_2, x_2 \rangle & \dots & \langle x_n, x_n \rangle \\ \vdots & \vdots & & \vdots \\ \langle x_1, x_1 \rangle & \langle x_2, x_2 \rangle & \dots & \langle x_n, x_n \rangle \end{pmatrix}_{n\times n}$$

$\Rightarrow$ en general, la entrada ij : $\underbrace{[\mathbb{1}c^t]_{ij} = \langle x_j, x_j \rangle}_{(2)}$

y la matriz $\mathbb{X}\mathbb{X}^t$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}_{n\times 1} \left(x_1, x_2, \dots x_n \right)_{1\times n} = \begin{pmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle & \dots & \langle x_1, x_n \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle & \dots & \langle x_2, x_n \rangle \\ \\ \langle x_n, x_1 \rangle & \langle x_n, x_2 \rangle & \dots & \langle x_n, x_n \rangle \end{pmatrix}_{n\times n}$$

$\Rightarrow$ en general, la entrada ij : $\underbrace{[\mathbb{X}\mathbb{X}^t]_{ij} = \langle x_i, x_j \rangle}_{(3)}$

Además por propiedades de matrices:

$$[c\mathbb{1}^t + \mathbb{1}c^t - 2\mathbb{X}\mathbb{X}^t]_{ij} = [\underbrace{c\mathbb{1}^t}_{(1)}]_{ij} + [\underbrace{\mathbb{1}c^t}_{(2)}]_{ij} - 2[\underbrace{\mathbb{X}\mathbb{X}^t}_{(3)}]_{ij}$$

Por lo que:

$$[c\mathbb{1}^t + \mathbb{1}c^t - 2XX^t]_{ij} = \langle x_i, x_i \rangle + \langle x_i, x_j \rangle - 2\langle x_i, x_j \rangle$$

pero el lado derecho es igual a $\mathbb{D}^2_{ij}$ (α), así que por transitividad:

$$\mathbb{D}^2_{ij} = [c\mathbb{1}^t + \mathbb{1}c^t - 2XX^t]_{ij}$$

$$\Rightarrow \mathbb{D}^2 = c\mathbb{1}^t + \mathbb{1}c^t - 2XX^t$$

Problema 3.

A partir de la demostración de la maximización del cociente de Rayleigh.

demuestra que el segundo vector propio de la $\mathrm{cov}X$ es la solución del problema de maximizar el coeficiente pero con la restricción adicional de ser ortogonal al primer vector propio.

solución:

Sea V_1 el vector propio dominante de cov, buscamos l_0 que resuelve el problema:

$$l_0 = \max_l \left(\frac{l^t \mathrm{cov}X\, l}{l^t l} \right) \quad \therefore \quad l_0 \perp V_1$$

1. Recordemos que $\mathrm{cov}X$ es una matriz simétrica, lo que implica que podemos escribirla como:

$$\mathrm{cov}X = (\mathrm{cov}X)^{\frac{1}{2}}(\mathrm{cov}X)^{\frac{1}{2}}$$

$$= (U\Lambda^{\frac{1}{2}}U^t)(U\Lambda^{\frac{1}{2}}U^t)$$

donde U es la matriz con los vectores propios de $\mathrm{cov}X$ y Λ es la matriz diagonal cuyos elementos son los valores propios de $\mathrm{cov}X$: $\{\mu_i\}$

2. Reescribiendo el problema tomando en cuenta el punto anterior:

$$\text{máx}_l \left(\frac{l^t (U \Lambda^{\frac{1}{2}} U^t)(U \Lambda^{\frac{1}{2}} U^t)\, l}{l^t\, l} \right)$$

y multiplicando por la identidad en el denominador:

$$\text{máx}_l \left(\frac{l^t (U \Lambda^{\frac{1}{2}} U^t)(U \Lambda^{\frac{1}{2}} U^t)\, l}{l^t\, \mathbb{I}\, l} \right)$$

sabemos que para cualquier matriz Λ $\mathbb{I} = \Lambda \cdot \Lambda^{-1}$, y la inversa de U es U^t pues U es una matriz ortonormal (compuesta de vectores propios)

$$\text{máx}_l \left(\frac{l^t (U \Lambda^{\frac{1}{2}} U^t)(U \Lambda^{\frac{1}{2}} U^t)\, l}{l^t\, U U^t\, l} \right)$$

3. Haciendo un cambio de norma $U^t l = y$ y usando otra vez que $U^t U = \mathbb{I}$

$$\text{máx}_l \left(\frac{\overbrace{l^t U}^{y^t} \Lambda^{\frac{1}{2}} \overbrace{U^t\ U}^{\mathbb{I}} \Lambda^{\frac{1}{2}} \overbrace{U^t l}^{y}}{\underbrace{l^t U}_{y^t} \underbrace{U^t l}_{y}} \right)$$

$$\text{máx}_y \left(\frac{y^t \Lambda\, y}{y^t\, y} \right)$$

4. Maximizamos el nuevo problema:

$$\text{máx}_y \left(\frac{y^t \Lambda\, y}{y^t\, y} \right) = \text{máx}_y \frac{\sum_i \lambda_i y_i^2}{\sum_i y_i^2}$$

Pero sabemos que $y = U^t l \Rightarrow y = (V_1 \cdot l\ ,\ V_2 \cdot l\ ,\ \quad V_n \cdot l)$

y l es ortogonal a $V_1 \Rightarrow y = (0, V_2 \cdot l, \ \dots\ V_n l)$

Entonces, nuestro problema es: $\max_{y} \dfrac{\sum_{i=2} \mu_i y_i^2}{\sum_{i=2} y_i^2}$

En este caso, el valor más grande que puede tomar μ_i es μ_2, el segundo eigenvalor dominante, veamos:

$$\frac{\sum_{i=2} \mu_i y_i^2}{\sum_{i=2} y_i^2} \leq \mu_2 \frac{\sum y_i^2}{\sum y_i^2} = \mu_2$$

Entonces μ_2 es una cota superior, que de hecho se alcanza cuando

$$y = (0, 1, 0, \ldots)$$

y como $y = U^t l$, $l = Uy$

$$\therefore\ l = V_2 \quad \text{(el segundo eigenvector dominante)}$$