4.1 (k-means).

$$\sum_{j=1}^{k} \sum_{x\in C_j} \|x-\mu_j\|^2 = \sum_{j=1}^{k}\sum_{x\in C_j} \langle x-\mu_j ,\ x-\mu_j\rangle$$

$$= \sum_{j=1}^{k}\sum_{x\in C_j} \left(\langle x, x\rangle - 2\langle x, \mu_j\rangle + \langle \mu_j, \mu_j\rangle \right)$$

$$= \underbrace{\sum_{j=1}^{k}\sum_{x\in C_j} \langle x, x\rangle} - 2\sum_{j=1}^{k}\sum_{x\in C_j} \langle x, \mu_j\rangle + \sum_{j=1}^{k}\sum_{x\in C_j} \langle \mu_j, \mu_j\rangle$$

$$= \sum_{x\in X} \|x\|^2 - 2\sum_{j=1}^{k} \langle \sum_{x\in C_j} x ,\ \mu_j\rangle + \sum_{j=1}^{k}\sum_{x\in C_j} \langle \mu_j, \mu_j\rangle$$

$$= \sum_{x\in X} \|x\|^2 - 2\cdot\sum_{j=1}^{k} \langle |C_j|\cdot\mu_j , \mu_j\rangle + \sum_{j=1}^{k}\sum_{x\in C_j} \langle \mu_j, \mu_j\rangle$$

$$= \sum_{x\in X} \|x\|^2 - \sum_{j=1}^{k}\sum_{x\in C_j} \langle \mu_j, \mu_j\rangle$$

$$= \underline{\sum_{x\in X} \|x\|^2} - \sum_{j=1}^{k} |C_j| \langle \mu_j, \mu_j\rangle .$$

$$\arg\min_{C} \sum_{j=1}^{k}\sum_{x\in C_j} \|x-\mu_j\|^2 = \arg\max_{C} \sum_{j=1}^{k} |C_j| \langle \mu_j, \mu_j\rangle .$$

(a). $$\sum_{x, x'\in C_j} \|x-x'\|^2 = \sum_{x\in C_j}\sum_{x'\in C_j} \langle x-x',\ x-x'\rangle$$

$$= \sum_{x\in C_j}\sum_{x'\in C_j} \left(\langle x, x\rangle - 2\langle x, x'\rangle + \langle x', x'\rangle \right)$$

$$= |C_j|\cdot\sum_{x\in C_j} \langle x, x\rangle + |C_j|\cdot\sum_{x'\in C_j}\langle x', x'\rangle - 2\langle \sum_{x\in C_j} x, \sum_{x'\in C_j} x'\rangle$$

$$= 2|C_j|\cdot\sum_{x\in C_j}\langle x, x\rangle - 2|C_j|^2\cdot\langle \mu_j, \mu_j\rangle .$$

$$\sum_{j=1}^{k} \frac{1}{2|C_j|}\cdot \sum_{\substack{x\in C_j\\ x'\in C_j}} \|x-x'\|^2 = \cdots = \sum_{x\in X}\|x\|^2 - \sum_{j=1}^{k} |C_j|\cdot\|\mu_j\|^2 .$$

4.2. (PCA).

$$\| x^{(i)} - (x^{(i)T}u)u\|^2 = \langle x^{(i)} - (x^{(i)T}u)\cdot u ,\ x^{(i)} - (x^{(i)T}u)\cdot u \rangle$$

$$= \langle x^{(i)}, x^{(i)}\rangle - 2\langle x^{(i)}, (x^{(i)T}u)\cdot u\rangle + (x^{(i)T}u)^2 \underbrace{\langle u, u\rangle}_{1}$$

$\langle u, u\rangle = 1.$

$$= \langle x^{(i)}, x^{(i)}\rangle + (x^{(i)T}u)^2 - 2(x^{(i)T}u)\cdot\langle x^{(i)}, u\rangle .$$

$$= \underbrace{\langle x^{(i)}, x^{(i)}\rangle}_{\text{does not change with } u.} - \underline{(x^{(i)T}u)^2} .$$

(b). $\frac{1}{n}\cdot\sum_{i=1}^{n}(x^{(i)T}u)^2 = u^T\Sigma u$

$A = B$, $A = C$, $B = C$

$\frac{1}{n}\cdot\sum_{i=1}^{n} \underbrace{x^{(i)T}u}_{= u^T\cdot x^{(i)}} x^{(i)T}u$

$= \frac{1}{n}\cdot\sum_{i=1}^{n} u^T\cdot \underbrace{x^{(i)}}_{m\times 1}\cdot \underbrace{x^{(i)T}}_{1\times m} u = u^T\cdot\underline{\left(\frac{1}{n}\cdot\sum_{i=1}^{n} x^{(i)}\cdot x^{(i)T}\right)}\cdot u.$

$u^T v = tr(u^T v) = tr(v\cdot u^T).$

4.3. If $\Sigma = \sigma_1 u_1 u_1^T + \sigma_2 u_2 u_2^T + \ldots + \sigma_n\cdot u_n u_n^T.$

$u_2 = \arg\max\ u^T\Sigma u.$

s.t. $\langle u,u\rangle = 1$

$u_1^T u = 0.$

$u^T\cdot\Sigma u = \beta_1^2 = \sigma_2^2.$

$\beta_2 = 0.$

$\mathcal{L}(u,\lambda_1,\lambda_2) = u^T\Sigma u + \beta_1\cdot(u^Tu - 1) + \beta_2\cdot u_1^T u$

$\Sigma u = -\beta_1\cdot u$

$\frac{\partial\mathcal{L}}{\partial u} = 2\Sigma u + 2\beta_1 u + \beta_2 u_1 = 0.$ $\quad(*)$

$\frac{\partial\mathcal{L}}{\partial\lambda_1} = u^Tu - 1 = 0$ $\quad(1)$

$\frac{\partial\mathcal{L}}{\partial\lambda_2} = u_1^T u = 0.$ $\quad(2)$

$u_1^T\cdot(*)$.

$2\cdot \underbrace{u_1^T\cdot\Sigma\cdot u}_{=0} + \underbrace{2\beta_1\cdot u_1^T\cdot u}_{=0} + \beta_2\cdot \underbrace{u_1^T u_1}_{=1} = 0. \Rightarrow 0 + 0 + \beta_2 = 0$

$\Downarrow$

$\beta_2 = 0.$

$\Sigma = \sigma_1 u_1 u_1^T + \sigma_2 u_2 u_2^T + \ldots + \sigma_n\cdot u_n u_n^T.$

① $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$

② $\langle u_i, u_j \rangle = \delta_{ij} = \begin{cases} 0 & , i \neq j \\ 1 & , i = j. \end{cases}$

$\langle u_1, u_1 \rangle = 1$, $\langle u_1, u_2 \rangle = 0$.

$$u_1^T \cdot \Sigma = u_1^T (\sigma_1 u_1 u_1^T + \sigma_2 u_2 u_2^T + \ldots + \sigma_n \cdot u_n u_n^T)$$

$$= \sigma_1 \cdot \underline{u_1^T \cdot u_1} \cdot u_1^T + \sigma_2 \cdot u_1^T \cdot u_2 \cdot u_2^T + \ldots + \sigma_n \cdot u_1^T \cdot u_n \cdot u_n^T.$$

$$= \sigma_1 \cdot 1 \cdot u_1^T + \sigma_2 \cdot 0 \cdot u_2^T + \ldots + \sigma_n \cdot 0 \cdot u_n^T$$

$$= \sigma_1 \cdot u_1^T.$$