Briana Zhao
U93987617
EC414 HW 7

1

ⓐ when $\vec{x}$ is 2-D: $\phi(\vec{x}) = \begin{pmatrix} 1 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ \sqrt{2}x_1x_2 \\ x_1^2 \\ x_2^2 \end{pmatrix}$

when $\vec{x}$ is 3-D: $\phi(\vec{x}) = \begin{pmatrix} 1 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ \sqrt{2}x_3 \\ \sqrt{2}x_1x_2 \\ \sqrt{2}x_1x_3 \\ \sqrt{2}x_2x_1 \\ \sqrt{2}x_2x_3 \\ \sqrt{2}x_3x_1 \\ \sqrt{2}x_3x_2 \\ x_1^2 \\ x_2^2 \\ x_3^2 \end{pmatrix}$

Pattern is:
→ first entry is 1
→ followed by d entries of form $\sqrt{2}x_i$
→ followed by $\frac{d(d-1)}{2}$ entries of form $\sqrt{2}x_ix_j$
→ followed by d entries of x_i^2

So $\phi(\vec{x})$ has $1 + d + \frac{d(d-1)}{2} + d$ dimensions when $\vec{x}$ is d-dimensional.

$$\phi(\vec{x})^T\phi(\vec{y}) = \underset{\text{first entry}}{1} + \underbrace{\sum_{i=1}^{d}(\sqrt{2}x_i)(\sqrt{2}y_i)}_{\text{next } d \text{ entries}} + \underbrace{\sum_{i=1,\, j=1,\, i\neq j}^{d}(\sqrt{2}x_ix_j)(\sqrt{2}y_iy_j)}_{\text{next } \frac{d(d-1)}{2} \text{ entries}} + \underbrace{x_1^2y_1^2 + x_2^2y_2^2 + x_3^2y_3^2 + \dots + x_d^2y_d^2}_{\text{last } d \text{ entries}}$$

$$= 1 + \underbrace{\sum_{i=1}^{d} 2x_iy_i}_{2x^Ty} + \underbrace{\sum_{i=1,\, j=1,\, i\neq j}^{d} 2x_ix_jy_iy_j + x_1^2x_1^2 + x_2^2y_2^2 + \dots + x_d^2y_d^2}_{(x^Ty)^2}$$

Therefore, $\boxed{\phi(\vec{x})^T\phi(\vec{y}) = (\vec{x}^T\vec{y})^2 + 2\vec{x}^T\vec{y} + 1 = (\vec{x}^T\vec{y}+1)^2}$

ⓑ for K-Means: $\mu_j = \frac{1}{n_j}\sum_{\vec{x}_n \in C_j} \vec{x}_n$ ← for points in cluster j

with mapping: $\mu_j = \frac{1}{n_j}\sum_{\vec{x}_n \in C_j} \phi(\vec{x}_n)$

So $\|\phi(\vec{x}_i) - \vec{\mu}_j\|_2^2 = (\phi(\vec{x}_i) - \vec{\mu}_j)^T(\phi(\vec{x}_i) - \vec{\mu}_j)^T$

$$= \left(\phi(\vec{x}_i) - \frac{1}{n_j}\sum_{\vec{x}_n \in C_j}\phi(\vec{x}_n)\right)^T\left(\phi(\vec{x}_i) - \frac{1}{n_j}\sum_{\vec{x}_n \in C_j}\phi(\vec{x}_n)\right)$$

$$= \phi(\vec{x}_i)^T\phi(\vec{x}_i) - \frac{2}{n_j}\sum_{\vec{x}_n \in C_j}\phi(\vec{x}_i)^T\phi(\vec{x}_n) + \frac{1}{n_j^2}\sum_{\vec{x}_n \in C_j}\sum_{\vec{x}_m \in C_j}\phi(\vec{x}_n)^T\phi(\vec{x}_m)$$

Using our formula from part ⓐ, this becomes:

$$(\vec{x}_i^T\vec{x}_i+1)^2 - \frac{2}{n_j}\sum_{\vec{x}_n\in C_j}(\vec{x}_i^T\vec{x}_n+1)^2 + \frac{1}{n_j^2}\sum_{\vec{x}_n\in C_j}\sum_{\vec{x}_m\in C_j}(\vec{x}_n^T\vec{x}_m+1)^2$$

Therefore, $\boxed{\text{distance } \|\phi(\vec{x}_i)-\mu_j\|_2^2 = (\vec{x}_i^T\vec{x}_i+1)^2 - \frac{2}{n_j}\sum_{\vec{x}_n\in C_j}(\vec{x}_i^T\vec{x}_n+1)^2 + \frac{1}{n_j^2}\sum_{\vec{x}_n\in C_j}\sum_{\vec{x}_m\in C_j}(\vec{x}_n^T\vec{x}_m+1)^2}$

ⓒ 1. Initialize k clusters of the data: $C_1^{(0)}, C_2^{(0)}, C_3^{(0)}, \dots C_k^{(0)}$ and set $t=0$

2. For each point x_i find its new cluster index as

$$i^*(x) = \underset{i}{\operatorname{argmin}}\|\phi(\vec{x}_i)-\mu_j\|_2^2$$ ← use equation found in part ⓑ

3. Update the new clusters:

$$C_i^{(t+1)} = \{\vec{x} \mid i^*(\vec{x}) = i\}$$

4. If not converged, $t=t+1$ and return to step 2. Otherwise, stop.

ⓓ For regular k-means (in original space), the boundary between C_1 & C_2 is given by the set of points $\vec{x}$ such that $\|\vec{x}-\vec{m}_1\|^2 = \|\vec{x}-\vec{m}_2\|^2$, which is the same as $2(\vec{m}_2-\vec{m}_1)^T\vec{x} + (\vec{m}_1^T\vec{m}_1 - \vec{m}_2^T\vec{m}_2) = 0$ which has a form like $\vec{w}^T\vec{x}+b$. This is why k-means has linear decision boundaries.

In the mapped space, we have $\|\phi(\vec{x})-\vec{m}_1\|^2 = \|\phi(\vec{x})-\vec{m}_2\|^2$ as the decision boundary. The decision boundary has form $\vec{w}^T\phi(\vec{x})+b=0$. From part ⓐ, we found that $\phi(\vec{x})$ is $1+d+\frac{1}{2}d(d-1)+d$ dimensions when d-dimensional. Therefore, $\vec{w}^T\phi(\vec{x})+b$ is a polynomial and the decision boundaries will be a polynomial decision surface. In the case where $d=2$, the decision surface in mapped space is quadratic.