Shucen Liu
Assn 2

1. (a) $\hat{C} = \text{argmin}_{C \in \mathcal{C}_K} R_n(C)$.
where $R_n(C) = \frac{1}{n}\sum_{i=1}^{n} \min_{1\le j\le K} \|x_i - c_j\|^2$
$C^* = \text{argmin}_{C\in\mathcal{C}_K} R(C)$
where $R(C) = E \min_{1\le j\le K} \|X - c_j\|^2$

$R(C^*)$ is the minimum of $R(C)$.
If $R(\hat{C})$ is close to $R(C^*)$,
since the distribution of X may not be equal (the probability of X may not be the same, $P(X_1) \ne P(X_2) \ne P(X_3)$ ...)
In certain case that $P(X_i) \ne \frac{1}{n}$,
$\hat{C}$ will be far away from C^*, even if $R(\hat{C})$ is close to $R(C^*)$

(b) $R^{(K)}$
$= \min R(C)$
$= \min E \min_{1\le j\le K} \|X - c_j\|^2$

As K increases, there will be more cluster centers. Thus for the best clusters, the difference between the points and the center will not increase as K increases.

(c) Taking derivative w.r.t λ_i, we'll get
$[\lambda_i]: (X_i - \mu - V_K\lambda_i)^T \lambda_i (X_i - \mu - V_K\lambda_i) = 0$.
Thus we must have $X_i = \mu + V_K\lambda_i$.
If we have $\hat{\mu} = \bar{x}_i$.
and $\lambda_i = V_K^T(x_i - \bar{x}_n)$

$\Rightarrow x_i = \bar{x}_i + V_K(V_K^T(x_i - \bar{x}_n))$
$= \bar{x}_i + x_i - \bar{x}_n$ (V_K orthognol)
$= x_i$

Thus, the formula is minimized.
$[V_K]: -\sum (X_i - \mu - V_K\lambda_i)^T V_K (X_i - \mu - V_K\lambda_i) = V_K \cdot \sum([\lambda_i])$
$[\mu]: -\sum (X_i - \mu - V_K\lambda_i) \mu (X_i - \mu - V_K\lambda_i) = \mu \cdot \sum([\lambda_i])$
As long as the derivative of $\lambda_i = 0$, the other two first order conditions are satisfied.
Thus, μ is not unique, as long as
$\mu = x_i - V_K\lambda_i$ (we can adjust λ when μ changes)