# Derivation of K-means update step

To make matters simple, we define a loss for a cluster rather than for all clusters:

$$\begin{aligned}\mathrm{Loss}(z,\mu) &= \sum_{i=1}^{n} \|\phi(x_i)-\mu\|^2 \\ &= \sum_{i=1}^{n} (\phi(x_i)-\mu)^{\mathrm{T}}(\phi(x_i)-\mu)\end{aligned}$$

where $\mu$ is the feature vector of the cluster's centroid, $x_i$ is a data example and $n$ is the number of data examples. Then, we can compute the gradient with respect to $\mu_j$, which is an element of $\mu$:

$$\begin{aligned}\frac{\partial}{\partial \mu_j}\mathrm{Loss}(z,\mu) &= \frac{\partial}{\partial \mu_j}\sum_{i=1}^{n} \|\phi(x_i)-\mu\|^2 \\ &= \sum_{i=1}^{n} \frac{\partial}{\partial \mu_j}\|\phi(x_i)-\mu\|^2 \\ &= \sum_{i=1}^{n} \frac{\partial}{\partial \mu_j}\{(\phi(x_i)-\mu)^{\mathrm{T}}(\phi(x_i)-\mu)\} \\ &= \sum_{i=1}^{n} \frac{\partial}{\partial \mu_j}(\phi(x_i)^{\mathrm{T}}\phi(x_i) - 2\phi(x_i)^{\mathrm{T}}\mu + \mu^{\mathrm{T}}\mu) \\ &= \sum_{i=1}^{n} \left\{\frac{\partial}{\partial \mu_j}(\phi(x_i)^{\mathrm{T}}\phi(x_i)) + \frac{\partial}{\partial \mu_j}(-2\phi(x_i)^{\mathrm{T}}\mu) + \frac{\partial}{\partial \mu_j}(\mu^{\mathrm{T}}\mu)\right\} \\ &= \sum_{i=1}^{n} \left\{\frac{\partial}{\partial \mu_j}(\sum_{k}^{d}\phi(x_i)_k^2) + \frac{\partial}{\partial \mu_j}(-2\sum_{k}^{d}\phi(x_i)_k\mu_k) + \frac{\partial}{\partial \mu_j}(\sum_{k}^{d}\mu_k^2)\right\} \\ &\qquad\qquad ;\ d \text{ is the dimension of } \phi(x_i) \text{ and } \mu \\ &= \sum_{i=1}^{n} \left\{0 + \frac{\partial}{\partial \mu_j}(-2\phi(x_i)_j\mu_j) + \frac{\partial}{\partial \mu_j}(\mu_j^2)\right\} \\ &= \sum_{i=1}^{n} \{-2\phi(x_i)_j + 2\mu_j\} \\ &= 2\sum_{i=1}^{n} \{\mu_j - \phi(x_i)_j\}\end{aligned}$$

Now, we can get the optimal $\mu_j$ where $\frac{\partial}{\partial \mu_j}\mathrm{Loss}(z,\mu) = 0$ (0 is a scalar):

$$
\begin{aligned}
\frac{\partial}{\partial \mu_j}\text{Loss}(z, \mu) &= 2\sum_{i=1}^{n}(\mu_j - \phi(x_i)_j) \\
&= 0 \\
\implies \sum_{i=1}^{n}\phi(x_i)_j &= \sum_{i=1}^{n}\mu_j \\
\implies \sum_{i=1}^{n}\phi(x_i)_j &= n\mu_j
\end{aligned}
$$

$$
\therefore \mu_j = \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)_j
$$