3(a)

We have that the loss function for k-means is as follows:

$$\sum_n r_{nk}||\boldsymbol{x}_n - \mu_k||^2$$
$$\sum_n r_{nk}(\boldsymbol{x}_n - \mu_k)(\boldsymbol{x}_n - \mu_k)$$
$$\sum_n r_{nk}[(x_n^1 - \mu_k^1)^2 + (x_n^2 - \mu_k^2)^2 + \ldots + (x_n^n - \mu_k^n)^2]$$

We perform gradient descent on this loss function by taking the partial derivative with respect to each entry in the μ vector:

(PARTIAL MU-K, ALSO HOW TO BOLD MU-K): $\sum_n r_{nk} 2(\mu_k - \boldsymbol{x}_n) = 0$

Solving for μ_k we get

$\sum_n r_{nk}\mu_k = \sum_n r_{nk}\boldsymbol{x}_n$ and

$$\mu_k = \frac{\sum_n r_{nk}\boldsymbol{x}_n}{\sum_n r_{nk}}$$

To derive the update rule for r_{nk} we notice that we have the constraint $\sum_k r_{nk} = 1$. To minimize the loss function, you set $r_{nk} = 1$ when for the argument where $||\boldsymbol{x}_n - \mu_k||^2$ is minimized. In more compact notation, this is:

$r_{nk} = 1$ for $k = \text{argmin}_{k'} \ ||\boldsymbol{x}_n - \mu_{k'}||^2$, and $r_{nk} = 0$ otherwise.