Coordinate descent

$$F(z_{1:N}, \mathbf{m}_{1:k}) = \frac{1}{2} \sum_{n=1}^{N} ||\mathbf{x}_n - \mathbf{m}_{z_n}||^2$$

- Holding the means fixed, assigning each point to its closest mean minimizes F with respect to $z_{1:N}$.
- Holding the assignments fixed, computing the centroids of each cluster minimizes F with respect to $\mathbf{m}_{1:k}$.
- Thus, k-means is a *coordinate descent* algorithm.
- It finds a *local minimum*. (Multiple restarts are often necessary.)