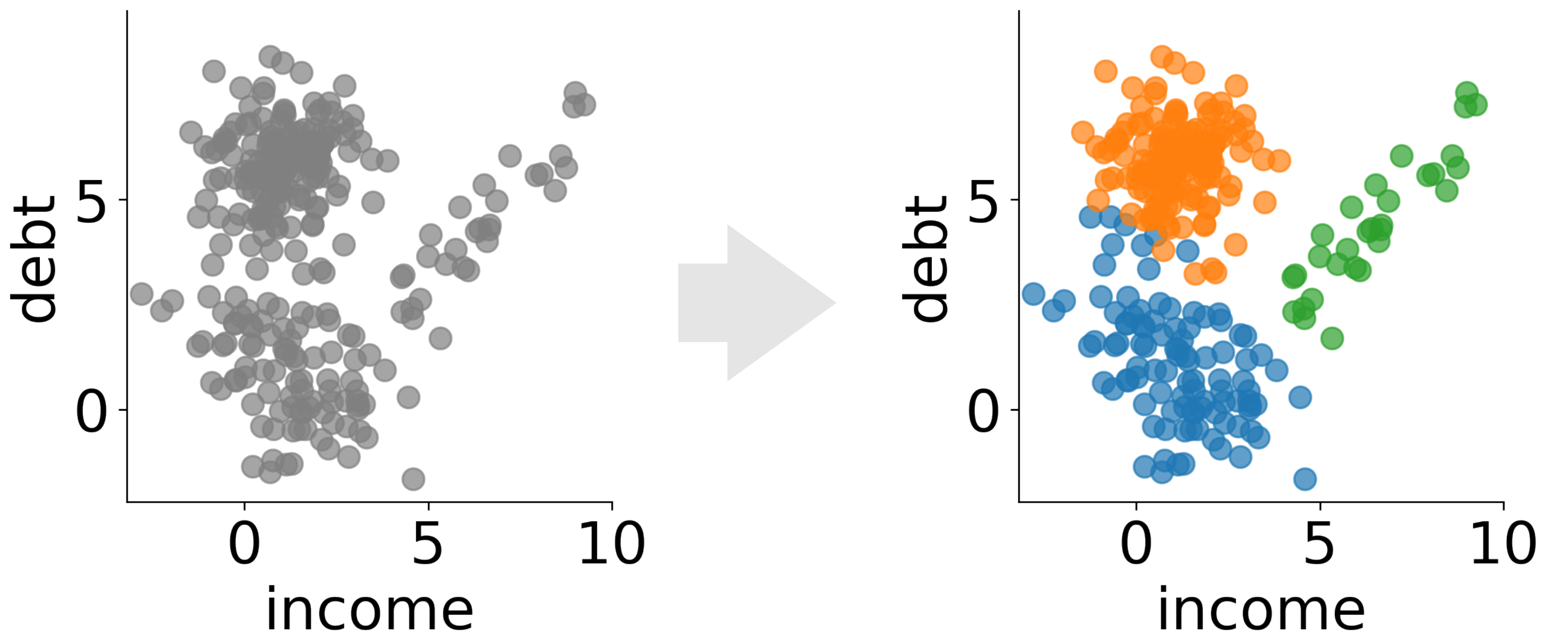
K-Means connection



K-Means

NOTE that we determine the number of means C (this is why its called K-means), just like GMM.

1. Randomly initialize parameters $\theta = \{\mu_1, \ldots, \mu_C\}$
2. Until convergence repeat:

   a) For each point compute closest centroid

   $$c_i = \arg\min_c \|x_i - \mu_c\|^2$$

   b) Update centroids

   $$\mu_c = \frac{\sum_{i:c_i=c} x_i}{\#\{i : c_i = c\}}$$

ok. simple as.

K-Means from GMM perspective

From GMM to K-means:

- Fix covariances to be identical $\Sigma_c = I$ (Identity matrix)
- Fix weights to be uniform $\pi_c = \dfrac{1}{\# \text{ of Guassians}}$

This makes it so that our gaussian model only has the means (mus) as the parameters.

$$p(x_i \mid t_i = c, \theta) = \frac{1}{Z} \exp\left(-0.5\|x_i - \mu_c\|^2\right)$$

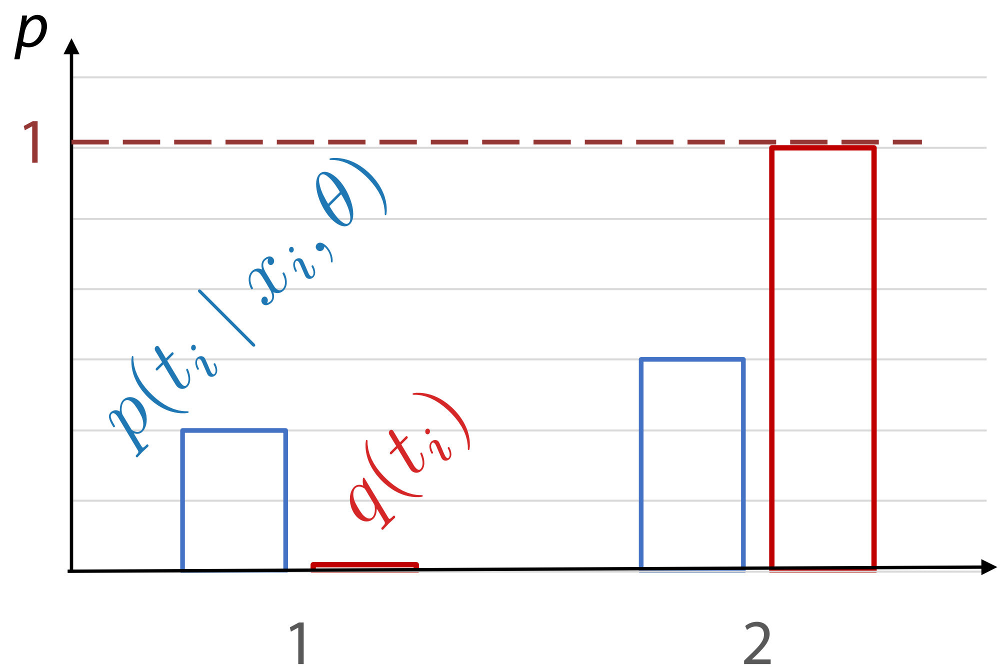
^squared euclidean distance.

K-Means from EM perspective

E-step

$$q^{k+1} = \arg\min_{q \in Q} \mathcal{KL}\left[q(T) \parallel p(T \mid X, \theta^k)\right]$$

Where Q is the set of delta-functions



K-Means from EM perspective

E-step

$$q^{k+1}(t_i) = \begin{cases} 1 & \text{if } t_i = c_i \\ 0 & \text{otherwise} \end{cases}$$

$$c_i = \arg\max_c p(t_i = c \mid x_i, \theta)$$

$$\begin{aligned} p(t_i \mid x_i, \theta) &= \frac{1}{Z} p(x_i \mid t_i, \theta)\, p(t_i \mid \theta) \\ &= \frac{1}{Z} \exp\left(-0.5\|x_i - \mu_c\|^2\right) \pi_c \end{aligned}$$

what is pi here?
It is the prior for a given class c.

K-Means from EM perspective

E-step

$$q^{k+1}(t_i) = \begin{cases} 1 & \text{if } t_i = c_i \\ 0 & \text{otherwise} \end{cases}$$

$$c_i = \arg\max_c p(t_i = c \mid x_i, \theta) = \arg\min_c \|x_i - \mu_c\|^2$$

$$\begin{aligned} p(t_i \mid x_i, \theta) &= \frac{1}{Z} p(x_i \mid t_i, \theta)\, p(t_i \mid \theta) \\ &= \frac{1}{Z} \exp\left(-0.5\|x_i - \mu_c\|^2\right) \pi_c \end{aligned}$$

K-Means from EM perspective

E-step

$$q^{k+1}(t_i) = \begin{cases} 1 & \text{if } t_i = c_i \\ 0 & \text{otherwise} \end{cases}$$

$$c_i = \arg\min_c \|x_i - \mu_c\|^2$$

Exactly like in K-Means!