The E-step in EM Algorithm

prior probability of being in group k

likelihood of seeing x in group k

$$p(\text{cluster } k \mid x) = \frac{w_k \; \mathcal{N}(x \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} w_j \; \mathcal{N}(x \mid \mu_j, \Sigma_j)}$$

posterior probability of being in cluster k

Responsibilities are the posterior probability of a data point being in cluster

How likely is the cluster? Many/few data points there?

How well does that data fit with that cluster?

$$r_{nk} = \frac{w_k \; \mathcal{N}(x \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} w_j \; \mathcal{N}(x \mid \mu_j, \Sigma_j)}$$

normalize to get a probability

Responsibility is high if the data point is likely to belong to that cluster rather than other clusters

this is soft assignment

GMM: EM Algorithm

1. Choose **k** random points to be cluster centers (or estimate using k-means...)
2. For each data point, calculate the **probability** of belonging to each cluster
3. Using these probability weights, recalculate the **means + variances** (and weights)
4. Repeat 2 and 3 until **distributions converge**