

E-step

Compute “expected” classes of all datapoints

$$P(Y_j = k | x_j, \mu_1 \dots \mu_K) \propto \exp\left(-\frac{1}{2\sigma^2} \|x_j - \mu_k\|^2\right) \cancel{P(Y_j = k)}$$

M-step

Compute most likely new μ s given class expectations

$$\mu_k = \frac{\sum_{j=1}^m P(Y_j = k | x_j) x_j}{\sum_{j=1}^m P(Y_j = k | x_j)}$$

$$\mu_k = \frac{\delta(Y_j = k, x_j) x_j}{\sum_{j=1}^m \delta(Y_j = k, x_j)}$$

δ represents hard assignment to “most likely” or nearest cluster