E-step / Compute cluster assignment
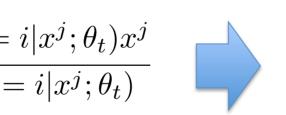
Compute "expected" classes → set most likely class

$$p(y = i|x^j; \theta_t) = \exp\left(-\frac{1}{2\sigma^2}\|x^j - \mu_i\|_2^2\right)$$

➡

$$a^i = \arg\min_j dist(x^i, c^j)$$

$$dist(x, x') = \|x - x'\|_2^2$$

M-step / Recompute cluster mean

Compute most likely new **μ**s → averages over hard assignments

$$\mu_i = \frac{\sum_{j=1}^m p(y = i|x^j; \theta_t)x^j}{\sum_{j=1}^m p(y = i|x^j; \theta_t)}$$

➡

$$c^i = \frac{1}{|\{j|a^j = i\}|} \sum_{\{j|a^j = i\}} x^j$$

With hard assignments and unit variance, EM is equivalent to k-means clustering algorithm!!!