Note: assume "c" classes, "n" datapoints
Also notation $P(\underset{\sim}{x} \mid w_j, \mu_1 \cdots \mu_c)$ mean "prob of $\underset{\sim}{x}$ given $\underset{\sim}{\mu_1} \cdots \underset{\sim}{\mu_c}$ and given we know $\underset{\sim}{x}$ is from class w_j"

$$\text{Log likelihood} = \log P(\text{data} \mid \mu_1 \cdots \mu_c)$$

$$= \log P(x_1 \cdots x_n \mid \mu_1 \cdots \mu_c)$$

$$= \log \prod_{k=1}^{n} P(x_k \mid \mu_1 \cdots \mu_c)$$

$$= \sum_{k=1}^{n} \log P(x_k \mid \mu_1 \cdots \mu_c)$$

$$= \sum_{k=1}^{n} \log \left[\sum_{j=1}^{c} P(x_k \mid w_j, \mu_1 \cdots \mu_c) P(w_j) \right]$$

NOTE $\frac{\partial}{\partial x} \log f(x) = \frac{1}{f(x)} \frac{\partial}{\partial x} f(x)$

$$\frac{\partial \log P}{\partial \mu_i} = \sum_{k=1}^{n} \frac{1}{\cancel{\log} P(x_k \mid \mu_1 \cdots \mu_c)} \frac{\partial}{\partial \mu_i} \sum_{j=1}^{c} P(x_k \mid w_j, \mu_1 \cdots \mu_c) P(w_j)$$

$$= \sum_{k=1}^{n} \frac{P(w_i)}{\cancel{\log} P(x_k \mid \mu_1 \cdots \mu_c)} \frac{\partial}{\partial \mu_i} P(x_k \mid w_i, \mu_1 \cdots \mu_c)$$

$$= \sum_{k=1}^{n} \frac{P(w_i)}{P(x_k \mid \mu_1 \cdots \mu_c)} P(x_k \mid w_i, \mu_1 \cdots \mu_c) \frac{\partial}{\partial \mu_i} \log P(x_k \mid w_i, \mu_1 \cdots \mu_c)$$

$$= \sum_{k=1}^{n} P(w_i \mid x_k, \mu_1 \cdots \mu_c) \frac{\partial}{\partial \mu_i} \log P(x_k \mid w_i, \mu_1 \cdots \mu_c)$$ BY BAYES

$$= \sum_{k=1}^{n} P(w_i \mid x_k, \mu_1 \cdots \mu_c) \frac{\partial}{\partial \mu_i} \log \exp\left(-\frac{1}{2\sigma^2} (x_k - \mu_i)^2 \right)$$

$$= \sum_{k=1}^{n} P(w_i \mid x_k, \mu_1 \cdots \mu_c) \left(-\frac{1}{2\sigma^2} \frac{\partial}{\partial \mu_i} (x_k - \mu_i)^2 \right)$$

$$= \sum_{k=1}^{n} P(w_i \mid x_k, \mu_1 \cdots \mu_c) \frac{(x_k - \mu_i)}{\sigma} = 0$$ for max likelihood, so

$$\underset{\sim}{\mu_i} = \frac{\sum_{k=1}^{n} P(w_i \mid \underset{\sim}{x_k}, \underset{\sim}{\mu_1} \cdots \underset{\sim}{\mu_c}) \underset{\sim}{x_k}}{\sum_{k=1}^{n} P(w_i \mid \underset{\sim}{x_k}, \underset{\sim}{\mu_1} \cdots \underset{\sim}{\mu_c})}$$

$$\text{Log likelihood} = \log P(\text{data} \,|\, \mu_1 \cdots \mu_c)$$

> Note: assume "c" classes, "n" datapoints
> Also notation $P(\underset{\sim}{x} \,|\, w_j, \mu_1 \cdots \mu_c)$ mean "prob of $\underset{\sim}{x}$ given $\underset{\sim}{\mu_1} \cdots \underset{\sim}{\mu_c}$ and given we know $\underset{\sim}{x}$ is from class w_j"

$$= \log P(x_1 \cdots x_n \,|\, \mu_1 \cdots \mu_c)$$

$$= \log \prod_{k=1}^{n} P(x_k \,|\, \mu_1 \cdots \mu_c)$$

$$= \sum_{k=1}^{n} \log P(x_k \,|\, \mu_1 \cdots \mu_c)$$

$$= \sum_{k=1}^{n} \log \left[\sum_{j=1}^{c} P(x_k \,|\, w_j, \mu_1 \cdots \mu_c) P(w_j) \right]$$

NOTE $\frac{\partial}{\partial x} \log f(x) = \frac{1}{f(x)} \frac{\partial}{\partial x} f(x)$

$$\frac{\partial \log P}{\partial \mu_i} = \sum_{k=1}^{n} \frac{1}{\cancel{\log}\, P(x_k|\mu_1 \cdots \mu_c)} \frac{\partial}{\partial \mu_i} \sum_{j=1}^{c} P(x_k \,|\, w_j, \mu_1 \cdots \mu_c) P(w_j)$$

$$= \sum_{k=1}^{n} \frac{P(w_i)}{\cancel{\log}\, P(x_k|\mu_1 \cdots \mu_c)} \frac{\partial}{\partial \mu_i} P(x_k \,|\, w_i, \mu_1 \cdots \mu_c) \cancel{P(w_j)}$$

$$= \sum_{k=1}^{n} \frac{P(w_i)}{P(x_k|\mu_1 \cdots \mu_c)} P(x_k \,|\, w_i, \mu_1 \cdots \mu_c) \frac{\partial}{\partial \mu_i} \log P(x_k \,|\, w_i, \mu_1 \cdots \mu_c)$$

$$= \sum_{k=1}^{n} P(w_i \,|\, x_k, \mu_1 \cdots \mu_c) \frac{\partial}{\partial \mu_i} \log P(x_k \,|\, w_i, \mu_1 \cdots \mu_c) \quad \text{BY BAYES}$$

$$= \sum_{k=1}^{n} P(w_i \,|\, x_k, \mu_1 \cdots \mu_c) \frac{\partial}{\partial \mu_i} \log \exp\left(-\frac{1}{2\sigma^2} (x_k - \mu_i)^2\right)$$

$$= \sum_{k=1}^{n} P(w_i \,|\, x_k, \mu_1 \cdots \mu_c) \left(-\frac{1}{2\sigma^2} \frac{\partial}{\partial \mu_i} (x_k - \mu_i)^2\right)$$

$$= \sum_{k=1}^{n} P(w_i \,|\, x_k, \mu_1 \cdots \mu_c) \frac{(x_k - \mu_i)}{\sigma} = 0$$ for max likelihood, so

$$\underset{\sim}{\mu_i} = \frac{\sum_{k=1}^{n} P(w_i \,|\, \underset{\sim}{x_k}, \underset{\sim}{\mu_1} \cdots \underset{\sim}{\mu_c})\, \underset{\sim}{x_k}}{\sum_{k=1}^{n} P(w_i \,|\, \underset{\sim}{x_k}, \underset{\sim}{\mu_1} \cdots \underset{\sim}{\mu_c})}$$