Q1-1 Use Bayes Rule to derive an expression for
$p(y=k \mid \vec{x}, \vec{\mu}, \vec{\sigma})$

Using Bayes Rule for general probabilities:

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

We know $p(\vec{x} \mid y=k, \vec{\mu}, \vec{\sigma})$ from the question above.

Therefore

~~p(x|y=k)~~

$$p(y=k \mid \vec{x}, \vec{\mu}, \vec{\sigma}) = \frac{P(y=k, \vec{x}, \vec{\mu}, \vec{\sigma})}{P(\vec{x}, \vec{\mu}, \vec{\sigma})} \qquad (I)$$

Similarly, for $p(\vec{x} \mid y=k, \vec{\mu}, \vec{\sigma})$, we get:

$$p(\vec{x} \mid y=k, \vec{\mu}, \vec{\sigma}) = \frac{P(y=k, \vec{x}, \vec{\mu}, \vec{\sigma})}{P(y=k, \vec{\mu}, \vec{\sigma})} \qquad (II)$$

Therefore, from re-writing (II), we get

$$p(\vec{x} \mid y=k, \vec{\mu}, \vec{\sigma}) \times P(y=k, \vec{\mu}, \vec{\sigma}) = P(y=k, \vec{x}, \vec{\mu}, \vec{\sigma}) \qquad (III)$$

Therefore, substituting the LHS of (III) with the numerator in the RHS of (I), we obtain:

$$P(y=k \mid \vec{x}, \vec{\mu}, \vec{\sigma}) = \frac{P(\vec{x} \mid y=k, \vec{\mu}, \vec{\sigma}) \times P(y=k, \vec{\mu}, \vec{\sigma})}{\cancel{P(y=k, \vec{\mu}, \vec{\sigma})}\ P(\vec{x}, \vec{\mu}, \vec{\sigma})} \qquad (IV)$$

M1

Expanding this equation's subparts using Bayes' Theorem leads to:

$$P(y=k|\vec{x},\vec{\mu},\vec{\sigma}) = \frac{P(\vec{x}|y=k,\vec{\mu},\vec{\sigma}) \times P(y=k,\vec{\mu},\vec{\sigma})}{P(\vec{x},\vec{\mu},\vec{\sigma})} \quad \text{(V)}$$

But

$$P(\vec{x},\vec{\mu},\vec{\sigma}) = P(\vec{x}|\vec{\mu},\vec{\sigma}) \times P(\vec{\mu}|\vec{\sigma}) \times P(\vec{\sigma})$$ from Bayes Thm.

and

$$P(y=k,\vec{\mu},\vec{\sigma}) = P(y=k|\vec{\mu},\vec{\sigma}) \times P(\vec{\mu}|\vec{\sigma}) \times P(\vec{\sigma})$$ from Bayes Thm.

Therefore eqn (V) can be expanded to:

$$P(y=k|\vec{x},\vec{\mu},\vec{\sigma}) = \frac{P(\vec{x}|y=k,\vec{\mu},\vec{\sigma}) \times P(y=k|\vec{\mu},\vec{\sigma}) \times P(\vec{\mu}|\vec{\sigma}) \times P(\vec{\sigma})}{P(\vec{x}|\vec{\mu},\vec{\sigma}) \times P(\vec{\mu}|\vec{\sigma}) \times P(\vec{\sigma})}$$

Cancelling out like terms gets us:

$$P(y=k|\vec{x},\vec{\mu},\vec{\sigma}) = \frac{P(\vec{x}|y=k,\vec{\mu},\vec{\sigma}) \times P(y=k|\vec{\mu},\vec{\sigma})}{P(\vec{x}|\vec{\mu},\vec{\sigma})} \quad \text{(VI)}$$

Now, using the hint provided (to use law of total probability for the denominator term, we obtain:

$$P(\vec{x}|\vec{\mu},\vec{\sigma}) = \sum_{k} P(x|y=k,\vec{\sigma},\vec{\mu}) \times P(y=k)$$

Therefore equation (VI) becomes:

$$P(y=k|\vec{x},\vec{\mu},\vec{\sigma}) = \frac{P(\vec{x}|y=k,\vec{\mu},\vec{\sigma}) \times P(y=k|\vec{\mu},\vec{\sigma})}{\sum_{k}(P(x|y=k,\vec{\sigma},\vec{\mu}) \times P(y=k))} \quad \text{(VII)}$$

Mo

Making some further substitutions:

$P(y=k \mid \vec{\mu}, \vec{\sigma}) = P(y=k)$ as $P(y=k) = a_k$ regardless of given prior information about the distribution's mean and variance (from the question above).

Therefore, eqn (VII) becomes:

$$P(y=k \mid \vec{x}, \vec{\mu}, \vec{\sigma}) = \frac{P(\vec{x} \mid y=k, \vec{\mu}, \vec{\sigma}) \times P(y=k)}{\sum_{k} P(x \mid y=k, \vec{\mu}, \vec{\sigma}) \times P(y=k)}$$

$$\boxed{P(y=k \mid \vec{x}, \vec{\mu}, \vec{\sigma} = \frac{P(x \mid y=k, \vec{\mu}, \vec{\sigma}) \times a_k}{\sum_{k} \left(P(x \mid y=k, \vec{\mu}, \vec{\sigma}) \times a_k \right)}} \qquad \text{(VIII)}$$

A further representation can be performed by replacing the respective probabilities with their Gaussian Equations:

$$\boxed{P(y=k \mid \vec{x}, \vec{\mu}, \vec{\sigma}) = \frac{\left(\prod_{i=1}^{i=D} 2\pi \sigma_i^2 \right)^{-1/2} \times \exp\left\{ -\sum_{i=1}^{i=D} \frac{1}{2\sigma_i^2} (x_i - \mu_{ki})^2 \right\} \times a_k}{\sum_{k=1}^{k=K} \left[\left(\prod_{i=1}^{i=D} 2\pi \sigma_i^2 \right)^{-1/2} \times \exp\left\{ -\sum_{i=1}^{i=D} \frac{1}{2\sigma_i^2} (x_i - \mu_{ki})^2 \right\} \times a_k \right]}} \qquad \text{(IX)}$$

"capital K"

"small k"

The Bayes formula/version of the solution is eqn (VIII).
The full Gaussian representation of the solution is eqn (IX).

m=

A further simplification can be performed by removing the

$$\left(\prod_{i=1}^{i=D} 2\pi\sigma_i \right)^{-1/2}$$

out of the main equation (in the denominator).

This cancels the same value in the numerator. giving us:

$$\boxed{p(y=k \mid \vec{x},\vec{\mu},\vec{\sigma}) = \frac{\exp\left\{-\sum_{i=1}^{i=D} \frac{1}{2\sigma_i^2}(x_i-\mu_{ki})^2\right\} * a_k}{\sum_{k=1}^{k=K}\left(\exp\left\{-\sum_{i=1}^{i=D}\frac{1}{2\sigma_i}(x_i-\mu_{ki})^2\right\} * a_k\right)}}$$

m2

Q1-2 Write down an expression for the negative log likelihood

NLL (negative log likelihood) = $\ell(\vec{\theta}; D) = -\log p(y^{(1)}, \vec{x}^{(1)}, y^{(2)}, \vec{x}^{(2)}, \ldots$
$\ldots, y^{(N)}, \vec{x}^{(N)} \mid \vec{\theta})$

for $\vec{\theta} = \{\vec{\alpha}, \vec{\mu}, \vec{\sigma}\}$

Note: taking log base as e (i.e ln)

Therefore

$\ell(\vec{\theta}; D) = -\log \cancel{(p^{(1)}, x}\ p(y^{(1)}, \vec{x}^{(1)}, \ldots y^{(N)}, \vec{x}^{(N)} \mid \vec{\alpha}, \vec{\mu}, \vec{\sigma})$

Since the data is i.i.d, it can be treated as the product of each sample:

$$= -\log \left(\prod_{n=1}^{n=N} p(y^{(n)}, \vec{x}^{(n)}, \vec{\alpha}, \vec{\mu}, \vec{\sigma}) \right)$$

By Bayes Rule, this becomes:

$$= -\log \left(\prod_{n=1}^{n=N} p(y^{(n)} \mid \vec{x}^{(n)}, \vec{\alpha}, \vec{\mu}, \vec{\sigma}) \times p(\vec{x}^{(n)} \mid \vec{\alpha}, \vec{\mu}, \vec{\sigma}) \right)$$

From ~~applying~~ $\log(A \times B \times C) = \log A + \log B + \log C$ (identity), we can remove ~~shift~~ the multiplication operator and replace it w/ a sigma (Σ).

$$= \sum_{n=1}^{n=N} \left(\log \left(p(y^{(n)} \mid \vec{x}^{(n)}, \vec{\alpha}, \vec{\mu}, \vec{\sigma}) \times p(\vec{x}^{(n)} \mid \vec{\alpha}, \vec{\mu}, \vec{\sigma}) \right) \right)$$

From the previous question (Q1-1) we know what each value is.

Therefore, substituting it:

$$= \sum_{n=1}^{n=N} \log\left(\frac{\cancel{\left(\prod_{i=1}^{i=D} 2\pi\sigma_i^2\right)^{-1/2}} \times \exp\left\{-\sum_{i=1}^{i=D} \frac{1}{2\sigma_i^2}(x_i - \mu_{ki})^2\right\} \times a_k}{\sum_{k=1}^{k=K} \exp\left\{-\sum_{i=1}^{i=D} \frac{1}{2\sigma_i^2}(x_i \ldots\right\}} \right.$$

"kay"

can't fit in the page

$$\left. \times \left(\prod_{i=1}^{i=D} 2\pi\sigma_i^2\right)^{-1/2} \sum_{k=1}^{k=K}\left[\alpha_k \times \exp\left\{-\sum_{i=1}^{i=D} \frac{1}{2\sigma_i^2}(x_i - \mu_i)^2\right\}\right] \right)$$

By cancelling out like terms, we obtain

$$= \sum_{n=1}^{n=N} \log\left[\left(\prod_{i=1}^{i=D} 2\pi\sigma_i^2\right)^{-1/2} \alpha_k(d) \exp\left\{\sum_{i=1}^{i=D} \frac{1}{2\sigma_i^2}\left(x_i - \mu_{k(\hat{d})i}\right)^2\right\}\right]$$

This gives us the answer of:

$$\boxed{l(\vec{\theta}; D) = \sum_{n=1}^{n=N}\left[\sum_{i=1}^{i=D}\left(\log(2\pi\sigma_i^2) + \frac{1}{2\sigma_i}\left(x_i - \mu_{k(\hat{d})i}\right)^2\right) - \log\left(\alpha_{k(\hat{d})}\right)\right]}$$

"kay"

Q1-3 Partial derivitaves w.r.t μ_{ki} and σ_i^2 of log likelihood

$NLL = L(\vec{\theta}; D)$

$$\frac{\partial NLL}{\partial \mu_{ki}} = \frac{\partial}{\partial \mu_{ki}} \left(\sum_{n=1}^{n=N} \left(\sum_{i=1}^{i=D} + \left(\log(\cancel{2\pi\sigma_i^2})^{0}\right) + \frac{1}{2\sigma_i^2}\left(x_i^{(n)} - \cancel{\mu_{k^{(n)}i}}\right)^2 - \left(\log \cancel{\alpha_{k^{(n)}}}^{0}\right)\right)\right)$$

$$\left(\frac{1}{2\sigma_i^2}\left(x_i^{(n)} - \mu_{k^{(n)}i}\right)^{1}(2)(-1)\right)$$

$\mu_{k^n i}$

$$\frac{\partial NLL}{\partial \mu_{ki}} = \cancel{\frac{1}{2}} \sum_{n=1}^{n=N} \sum_{i=1}^{i=D} \frac{1}{\sigma_i^2}\left(x_i^{(n)} - \mu_{k^{(n)}i}\right)$$

However for μ_{ki} only exists when $k^{(n)} = k$. Therefore, this becomes

↳ i.e. 1 else 0

$$\boxed{\frac{\partial NLL}{\partial \mu_{ki}} = \sum_{n=1}^{n=N} \left(\mathbb{1}(k^{(n)}, k) \frac{1}{\sigma_i^2}\left(x^{(n)} - \mu_{ki}\right)\right)}$$

↳ 1 when $k^n == k$ else 0.

Similarly, for $\frac{\partial NLL}{\partial \sigma_i^2}$, we get

$$\frac{1}{2\pi\sigma_i^2}(2\pi)$$

$$\frac{\partial NLL}{\partial \sigma_i^2} = \frac{\partial}{\partial \sigma_i^2} \sum_{n=1}^{n=N} \sum_{i=1}^{i=D} + \log(\cancel{2\pi\sigma_i^2}) + \frac{1}{2\sigma_i^2}\left(x_i^{n} - \cancel{\mu_{k^n i}}\right)^2 - \log(\cancel{\alpha_{k^n}})^{0}$$

$$\frac{-1}{2\sigma_i^4}\left(x_i^{(n)} - \mu_{k^{(n)}i}\right)^2$$

$$= \sum_{n=1}^{n=N} \left(\frac{1}{2\sigma_i^2} - \frac{1}{2\sigma_i^4} \left(x_i^{(n)} - \mu_{k^n i} \right)^2 \right)$$

$$\boxed{\frac{\partial NLL}{\partial \sigma_i^2} = \frac{N}{2\sigma_i^2} - \frac{1}{2\sigma_i^4} \sum_{n=1}^{n=N} \left(x_i^n - \mu_{k^n i} \right)^2}$$

$x_i^n \in x^n$

α

Q1-4 To get the MLEs for $\vec{\mu}$ and $\vec{\sigma}$, we have to set the partial derivatives of μ_{ki} and σ_i^2 (obtained from the previous question to 0).

<u>Note that since the MLE behaves well under transformations, and since the mapping of S.D σ to Variance σ^2 is such a one-to-one function, the MLE (σ) will be the same as the MLE (σ^2).</u>

Therefore:

$$MLE_{\mu_{ki}} = \frac{\partial NLL}{\partial \mu_{ki}} = 0$$

$$\Rightarrow \sum_{n=1}^{n=N} \left[\mathbb{1}(k^n, k) \frac{1}{\sigma_i^2} (x_i^n - \mu_{ki}) \right] = 0$$

$$\Rightarrow \boxed{\mu_{ki} = \frac{\sum_{n=1}^{n=N} \left(\mathbb{1}(k^n, k)\, x_i^n \right)}{\sum_{n=1}^{n=N} \left(\mathbb{1}(k^n, k) \right)}}$$

Similarly,

$$MLE_{\sigma} = \frac{\partial NLL}{\partial \sigma^2} = 0$$

$$\sum_{n=1}^{n=N} \frac{1}{2\sigma_i^2} - \frac{1}{2\sigma_i^2} \sum_{n=1}^{n=N} \left(x_i^n - \mu_{k^n i} \right)^2 = 0$$

$$\therefore \boxed{\sigma_i^2 = \frac{1}{N} \sum_{n=1}^{n=N} \left(x_i^n - \mu_{k^n i} \right)}$$

Therefore

$$MLE_{\sigma} = \frac{1}{N} \sum_{n=1}^{n=N} \left(x_i^n - \mu_{k^n i} \right)$$