CSC872 Pattern Analysis and Machine Intelligence, Spring 2024

HW #4: (Due in Two Week on 4/16)

Lectures 8 - 9

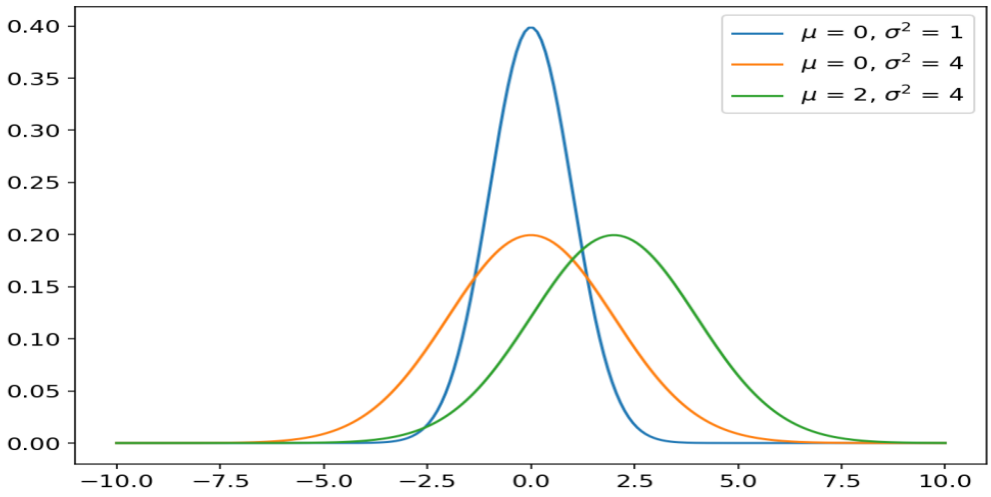
Total of 100 pts (10% of the total course credit)

Name: Ashmitha Dale Pais

SID: 923069586

1.1

Suppose lets say P(X) takes a form of a weighted sum of K different gaussian components.

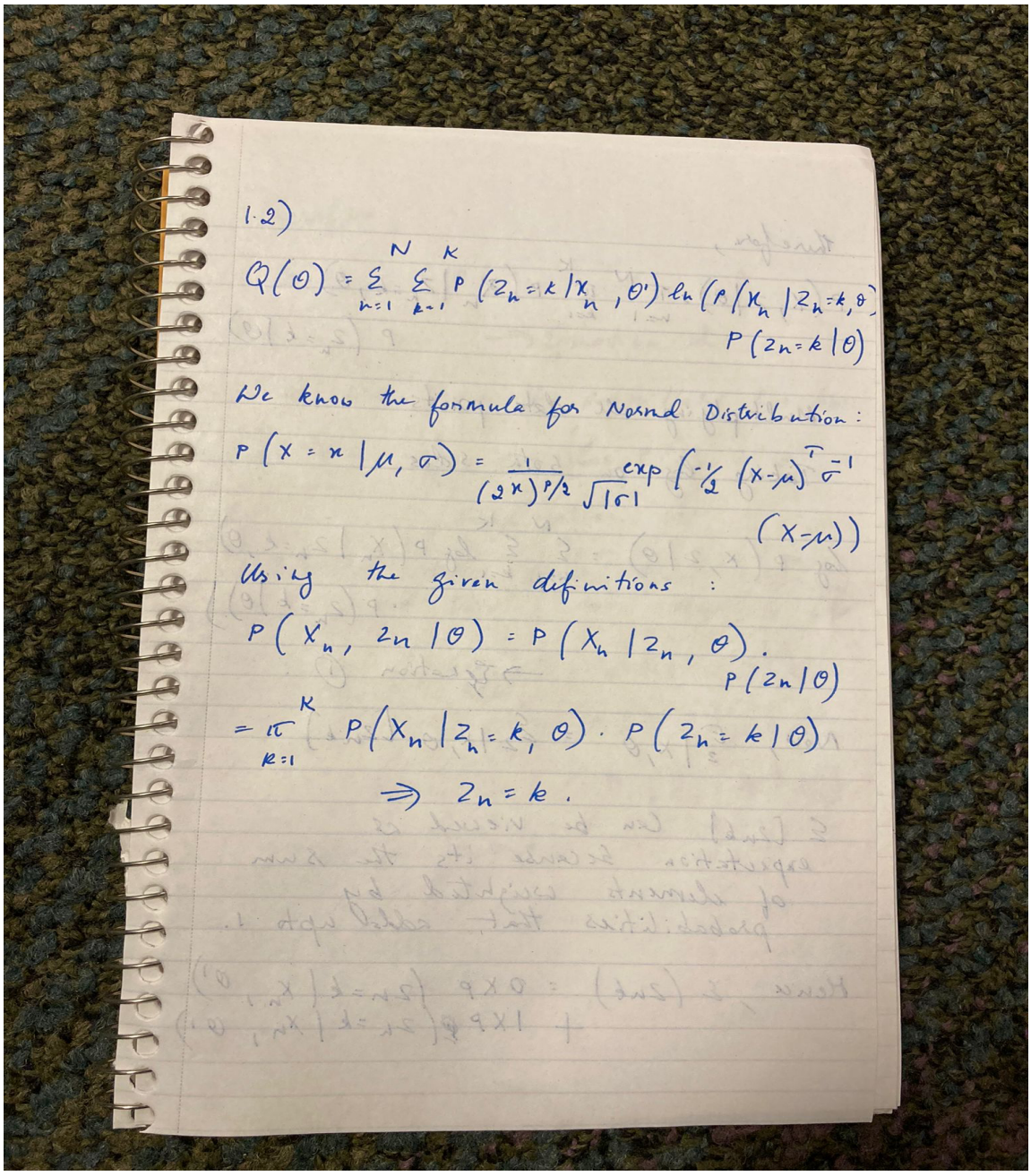


$P(X_n) = \Pi 1\ N(x \mid \mu_1, \sigma_1) + \Pi 2\ N(x \mid \mu_2, \sigma_2) + \Pi 3\ N(x \mid \mu_3, \sigma_3)$

Then, $P(X_n) = \sum_{k=1}^{K} \Pi k\ N(Xn \mid \mu k, \sigma k)$

Where $\Pi k =$ mixing weights of $\sum_{k=1}^{K} \Pi k = 1$

1.2



1.2)

$$Q(\theta) = \sum_{n=1}^{N} \sum_{k=1}^{K} P(z_n = k \mid x_n, \theta') \ln \left(\frac{P(x_n \mid z_n = k, \theta)}{P(z_n = k \mid \theta)} \right)$$

We know the formula for Normal Distribution:

$$P(x = x \mid \mu, \sigma) = \frac{1}{(2\pi)^{p/2} \sqrt{|\sigma|}} \exp\left(-\frac{1}{2} (x-\mu)^T \sigma^{-1} (x-\mu) \right)$$

Using the given definitions:

$$P(x_n, z_n \mid \theta) = P(x_n \mid z_n, \theta) \cdot P(z_n \mid \theta)$$

$$= \prod_{k=1}^{K} P(x_n \mid z_n = k, \theta) \cdot P(z_n = k \mid \theta)$$

$$\Rightarrow z_n = k.$$

Therefore,

$$P(X, Z|\theta) = \prod_{n=1}^{N} \prod_{k=1}^{K} P(X_n | Z_n = k, \theta) \cdot P(Z_n = k | \theta)$$

Multiplying all data points

Taking log on both sides

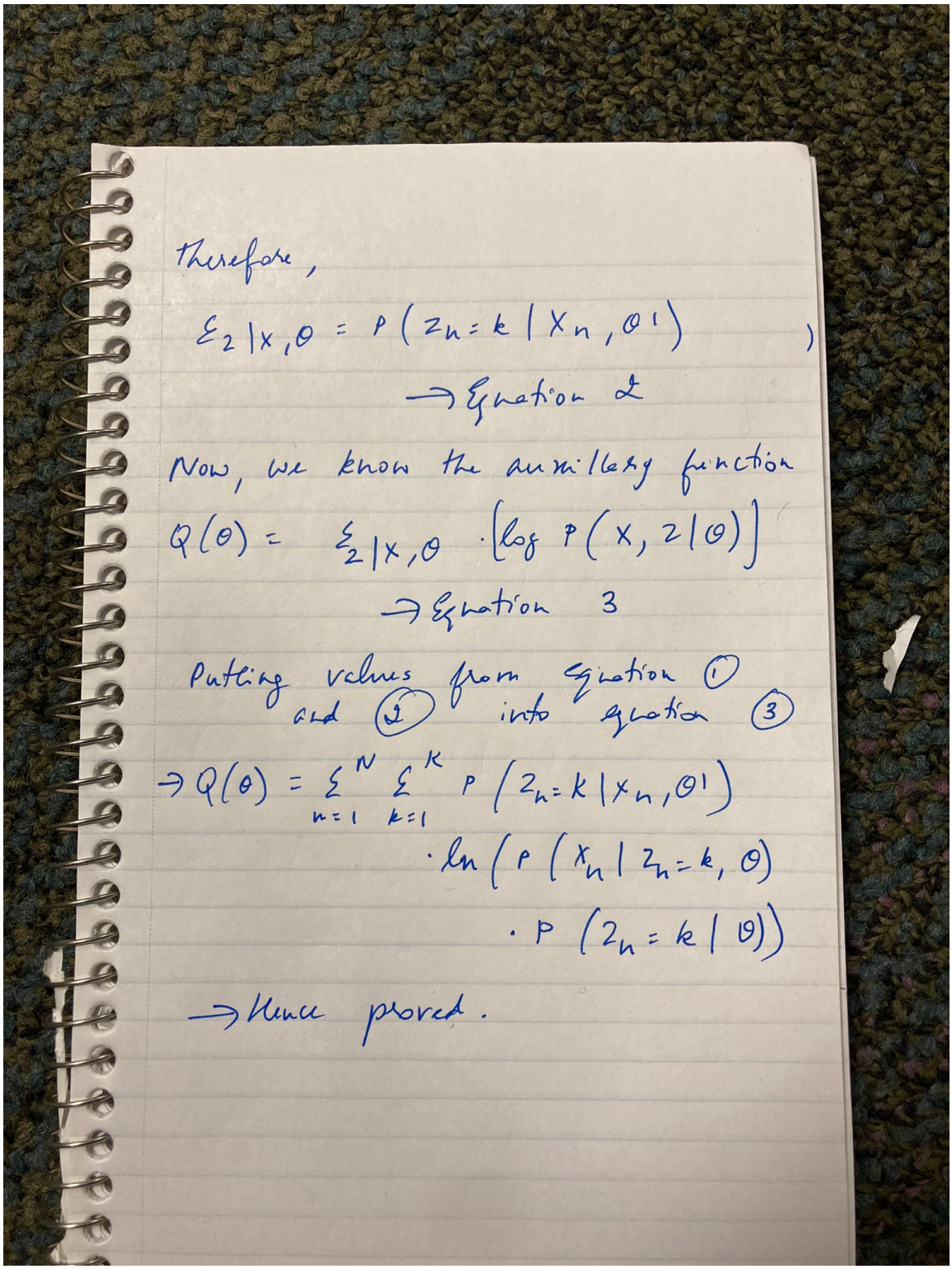
$$\log P(X, Z|\theta) = \sum_{n=1}^{N} \sum_{k=1}^{K} \log P(X_n | Z_n = k, \theta) \cdot P(Z_n = k | \theta))$$

→ Equation (1).

Now, $E_{Z|X,\theta} = E_{Z|X,\theta}[Z_{nk}]$

$E[Z_{nk}]$ Can be viewed as expectation because its the sum of elements weighted by probabilities that adds upto 1.

Hence, $E(Z_{nk}) = 0 \times P(Z_n = k | X_n, \theta') + 1 \times P(Z_n = k | X_n, \theta')$



Therefore,

$$\mathcal{E}_{2|X,\theta} = P(z_n = k \mid X_n, \theta^1)$$

→ Equation 2

Now, we know the auxiliary function

$$Q(\theta) = \mathcal{E}_{2|X,\theta} \cdot [\log P(X, z|\theta)]$$

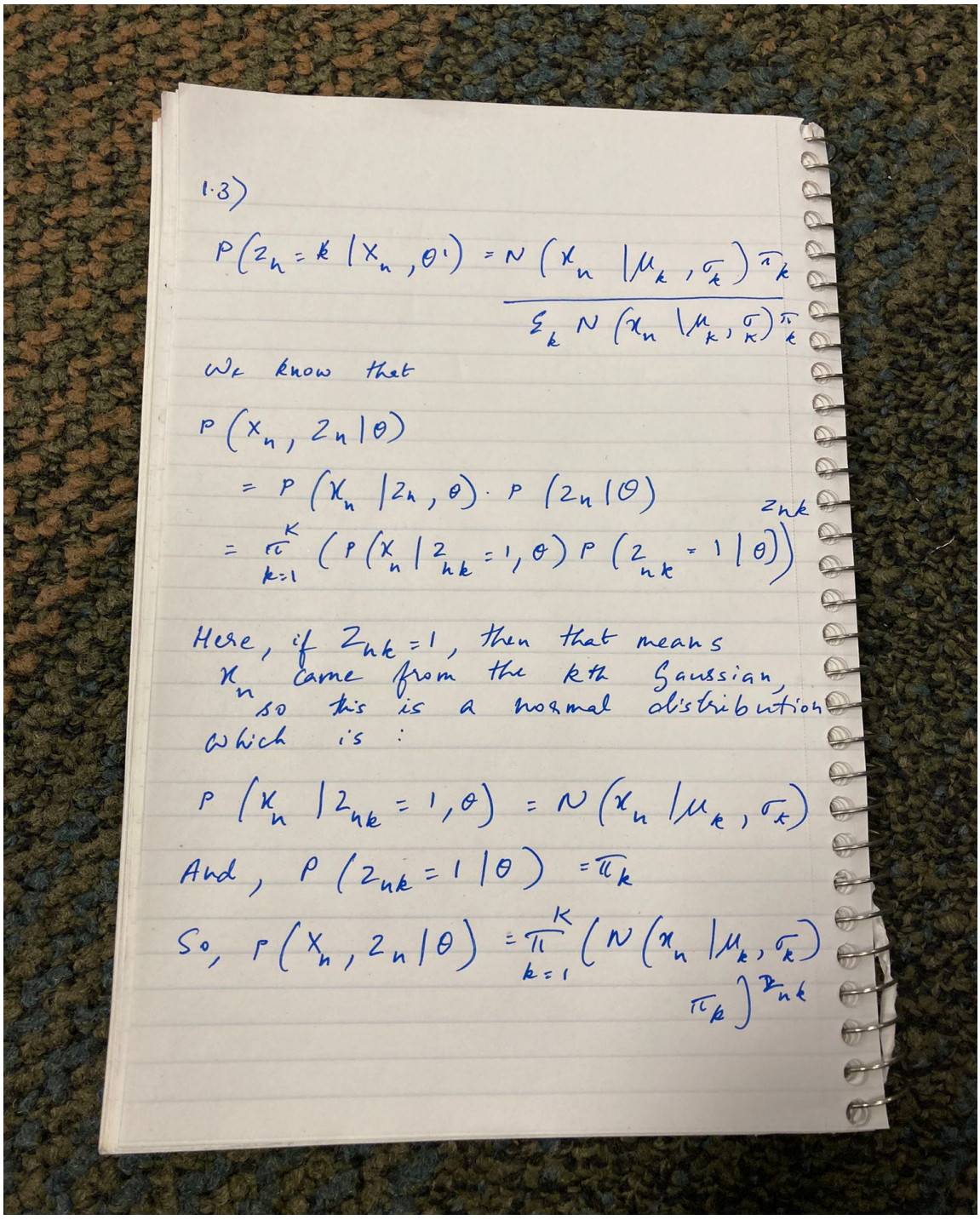
→ Equation 3

Putting values from equation (1) and (2) into equation (3)

$$\rightarrow Q(\theta) = \sum_{n=1}^{N} \sum_{k=1}^{K} P(z_n = k \mid X_n, \theta^1) \cdot \ln\left(P(X_n \mid z_n = k, \theta) \cdot P(z_n = k \mid \theta)\right)$$

→ Hence proved.

1.3



1.3)

$$P(z_n = k \mid x_n, \theta) = \frac{N(x_n \mid \mu_k, \sigma_k)\,\pi_k}{\sum_k N(x_n \mid \mu_k, \sigma_k)\,\pi_k}$$

We know that

$$P(x_n, z_n \mid \theta) = P(x_n \mid z_n, \theta) \cdot P(z_n \mid \theta)$$

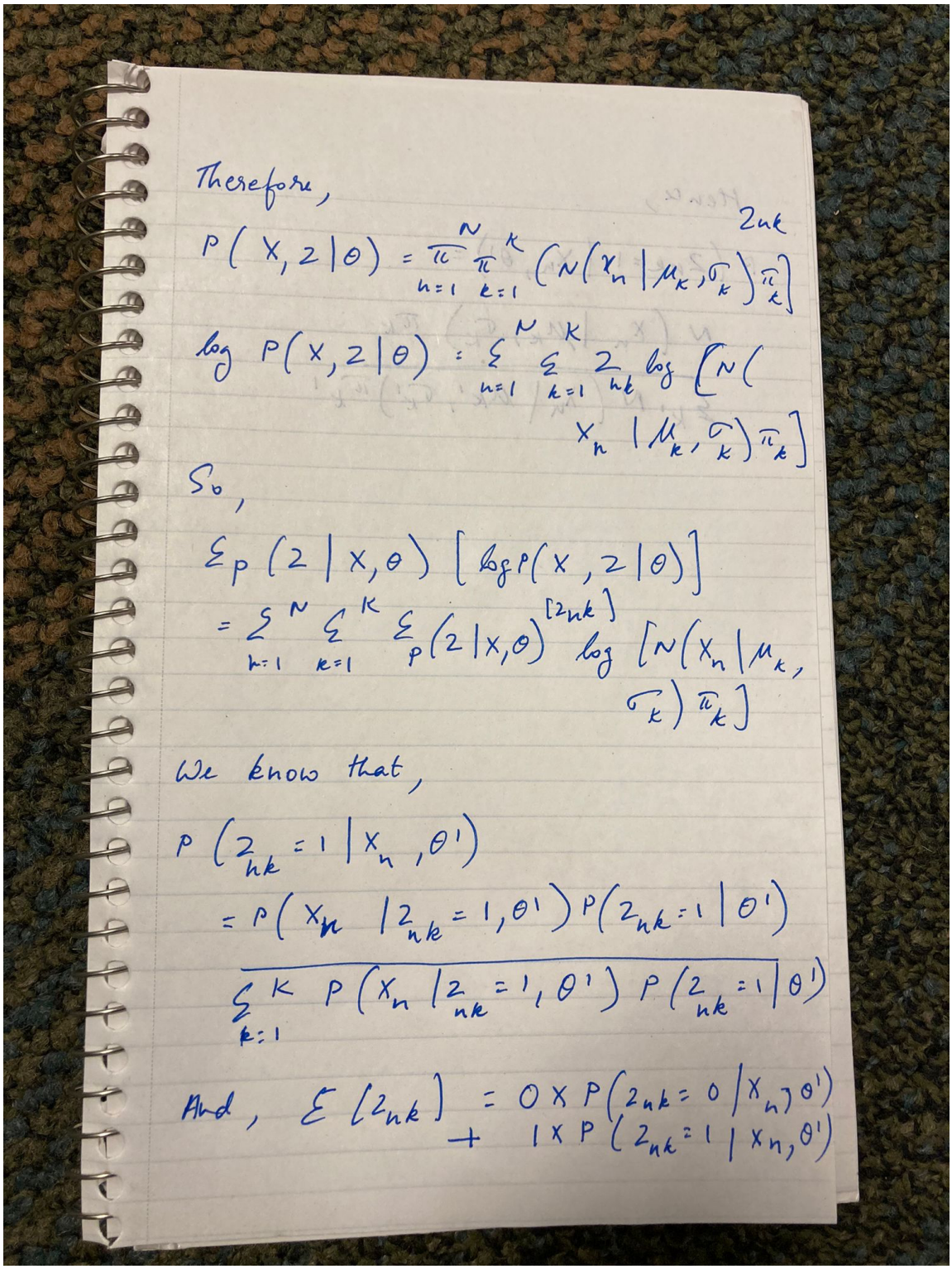
$$= \prod_{k=1}^{K} \left(P(x_n \mid z_{nk} = 1, \theta)\, P(z_{nk} = 1 \mid \theta) \right)^{z_{nk}}$$

Here, if $z_{nk} = 1$, then that means x_n came from the kth Gaussian, so this is a normal distribution which is:

$$P(x_n \mid z_{nk} = 1, \theta) = N(x_n \mid \mu_k, \sigma_k)$$

And, $P(z_{nk} = 1 \mid \theta) = \pi_k$

So, $P(x_n, z_n \mid \theta) = \prod_{k=1}^{K} \left(N(x_n \mid \mu_k, \sigma_k)\, \pi_k \right)^{z_{nk}}$



Therefore,

$$P(X, Z|\theta) = \prod_{n=1}^{N} \prod_{k=1}^{K} \left[\mathcal{N}(x_n | \mu_k, \sigma_k) \pi_k \right]^{z_{nk}}$$

$$\log P(X, Z|\theta) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \log \left[\mathcal{N}(x_n | \mu_k, \sigma_k) \pi_k \right]$$

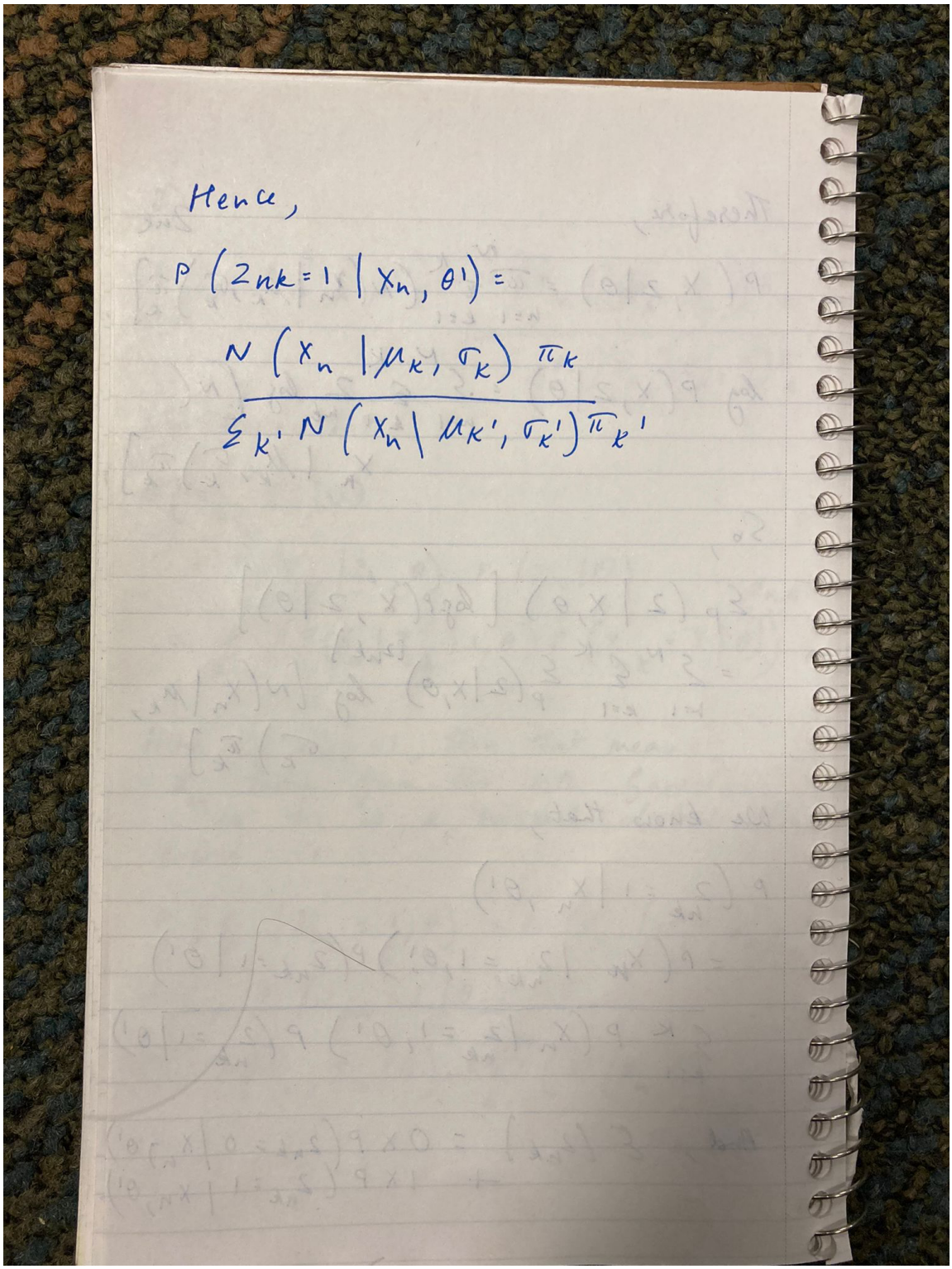
So,

$$\mathcal{E}_{p(Z|X,\theta)} \left[\log p(X, Z|\theta) \right] = \sum_{n=1}^{N} \sum_{k=1}^{K} \mathcal{E}_{p(Z|X,\theta)}[z_{nk}] \log \left[\mathcal{N}(x_n | \mu_k, \sigma_k) \pi_k \right]$$

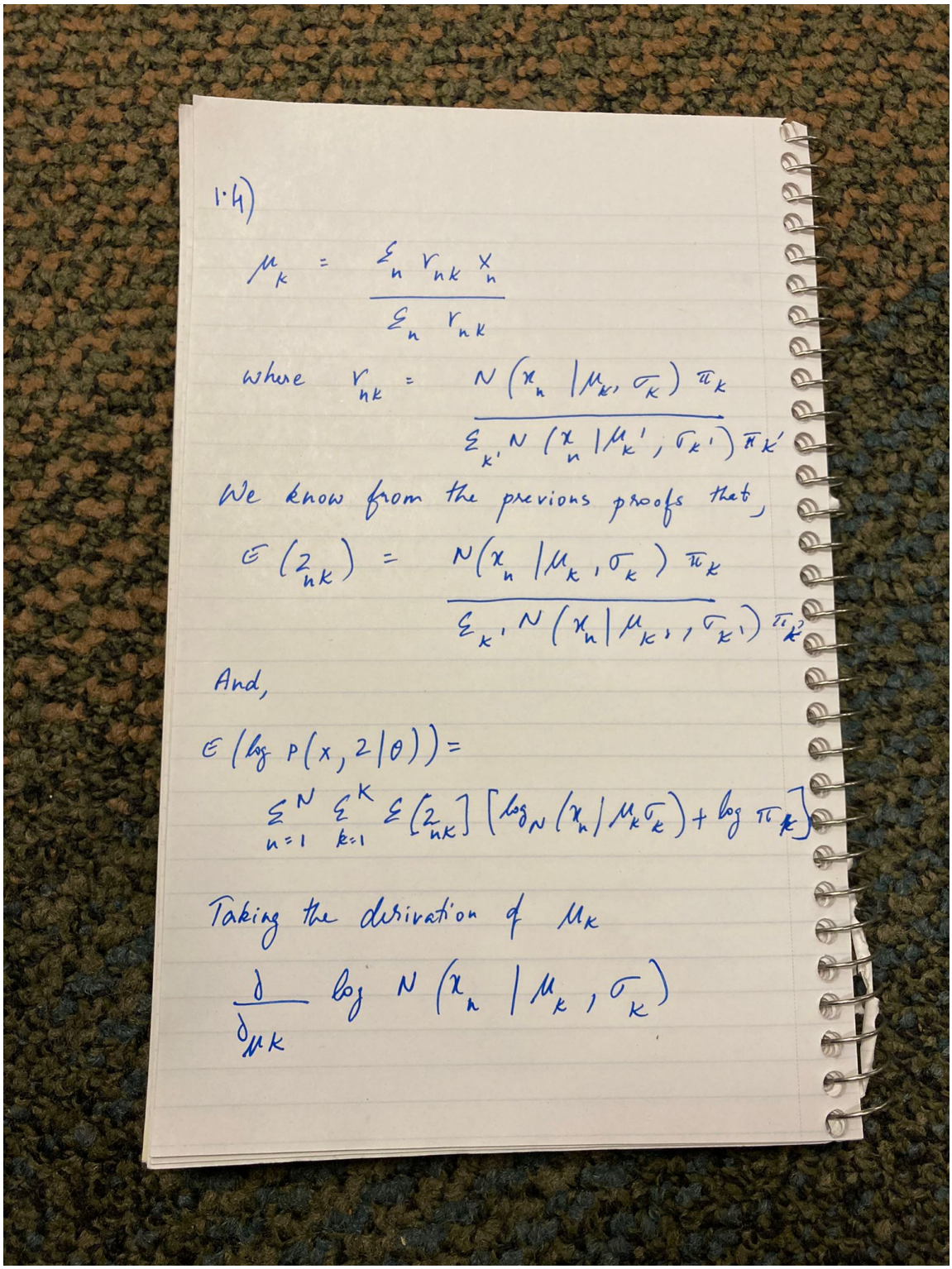
We know that,

$$p(z_{nk} = 1 | x_n, \theta') = \frac{p(x_n | z_{nk} = 1, \theta') p(z_{nk} = 1 | \theta')}{\sum_{k=1}^{K} p(x_n | z_{nk} = 1, \theta') p(z_{nk} = 1 | \theta')}$$

And, $\mathcal{E}[z_{nk}] = 0 \times p(z_{nk} = 0 | x_n, \theta') + 1 \times p(z_{nk} = 1 | x_n, \theta')$


Hence,
$p(z_{nk}=1 \mid x_n, \theta') =$
$\frac{N(x_n \mid \mu_k, \sigma_k)\,\pi_k}{\sum_{k'} N(x_n \mid \mu_{k'}, \sigma_{k'})\,\pi_{k'}}$

1.4



1.4)

$$\mu_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}}$$

where $r_{nk} = \frac{N(x_n | \mu_k, \sigma_k)\,\pi_k}{\sum_{k'} N(x_n | \mu_{k'}, \sigma_{k'})\,\pi_{k'}}$

We know from the previous proofs that,

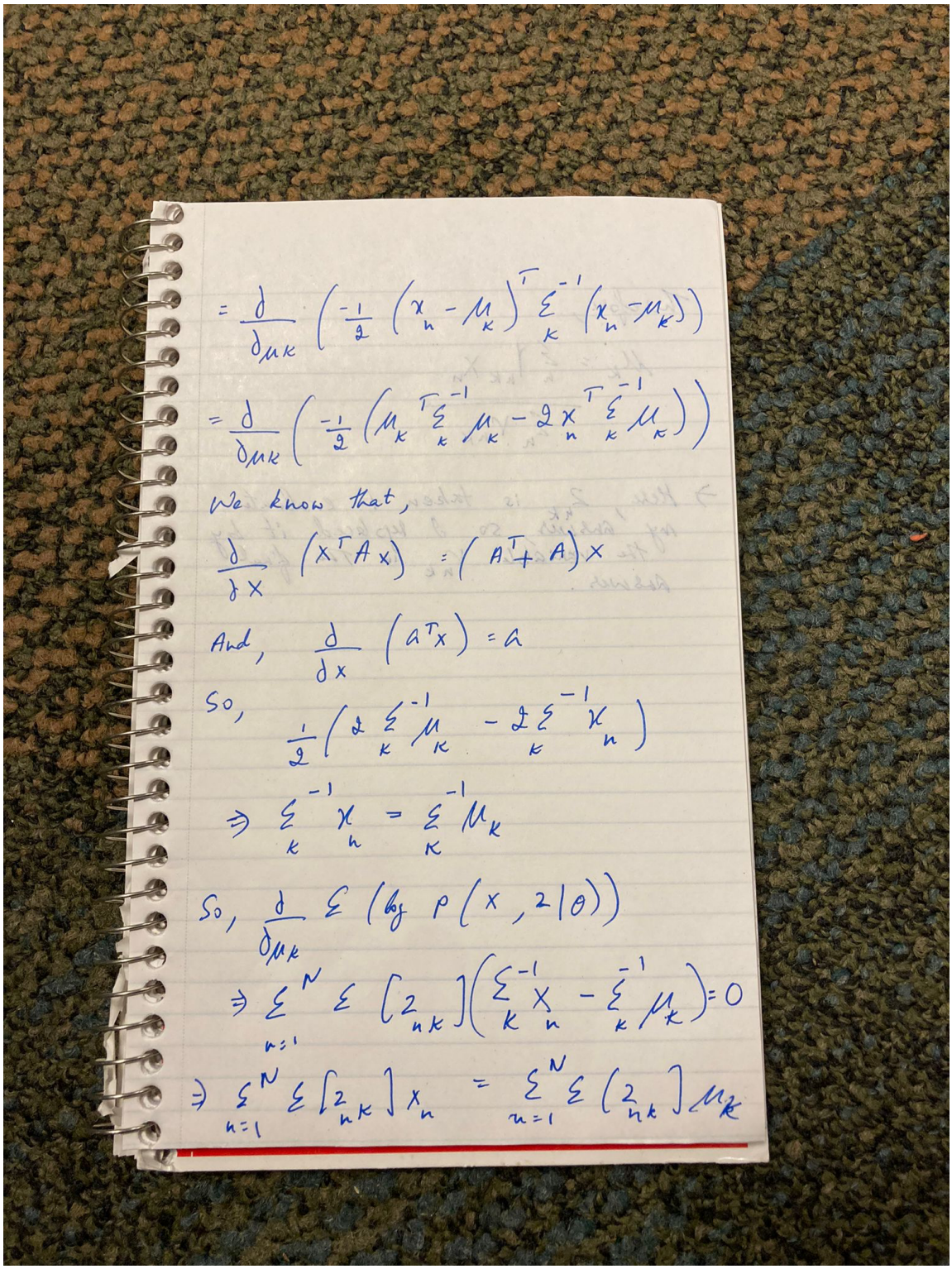
$$E(z_{nk}) = \frac{N(x_n | \mu_k, \sigma_k)\,\pi_k}{\sum_{k'} N(x_n | \mu_{k'}, \sigma_{k'})\,\pi_{k'}}$$

And,

$$E(\log p(x, z | \theta)) = \sum_{n=1}^{N} \sum_{k=1}^{K} E[z_{nk}] \left[\log_N (x_n | \mu_k \sigma_k) + \log \pi_k\right]$$

Taking the derivation of μ_k

$$\frac{\partial}{\partial \mu_k} \log N(x_n | \mu_k, \sigma_k)$$



$$= \frac{\partial}{\partial \mu_k}\left(\frac{-1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)\right)$$

$$= \frac{\partial}{\partial \mu_k}\left(\frac{-1}{2}\left(\mu_k^T \Sigma_k^{-1}\mu_k - 2x_n^T \Sigma_k^{-1}\mu_k\right)\right)$$

We know that,

$$\frac{\partial}{\partial x}(x^T A x) = (A^T + A)x$$

And, $\frac{\partial}{\partial x}(a^T x) = a$

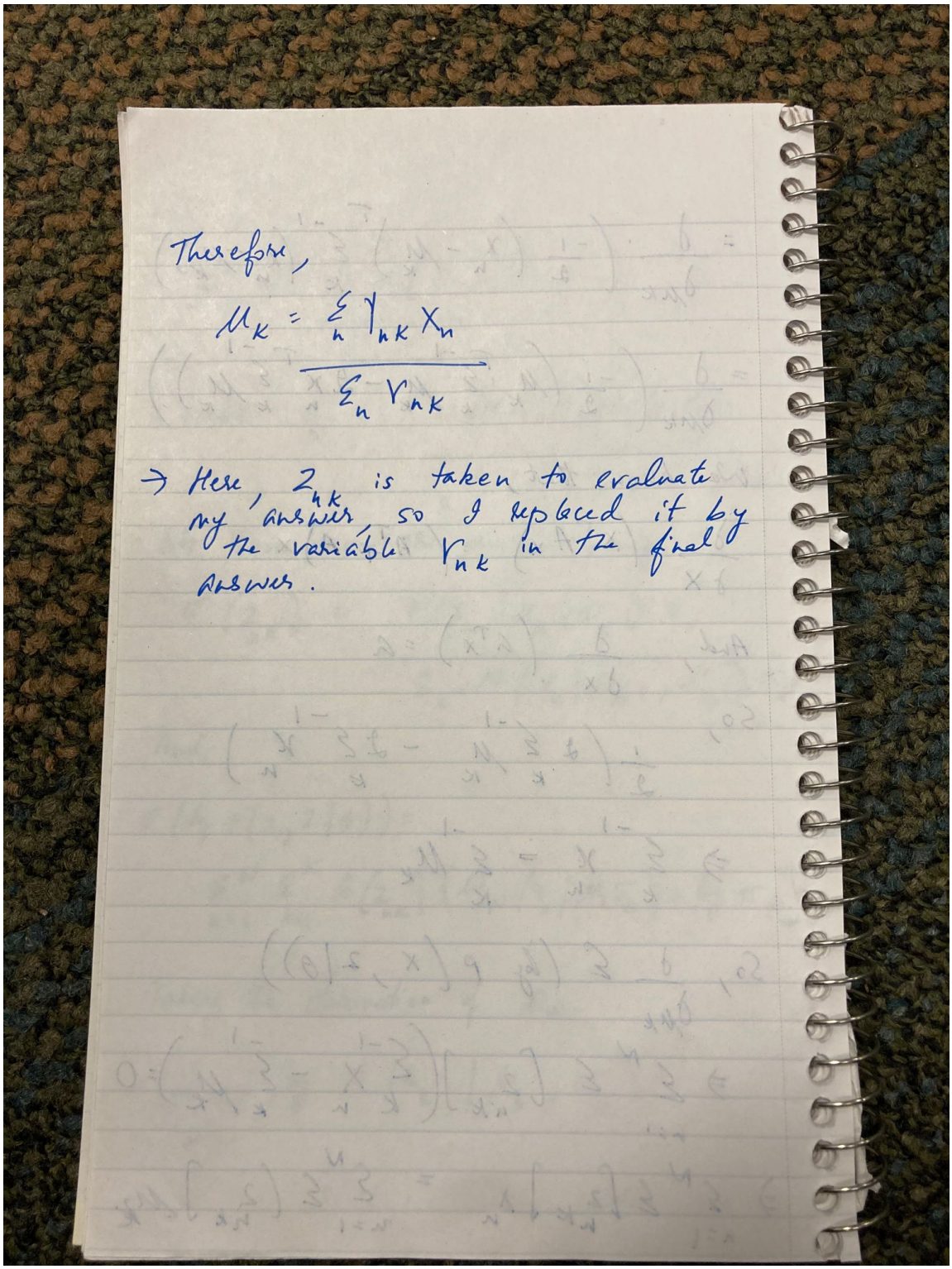
So, $\frac{1}{2}\left(2\Sigma_k^{-1}\mu_k - 2\Sigma_k^{-1}x_n\right)$

$$\Rightarrow \Sigma_k^{-1}x_n = \Sigma_k^{-1}\mu_k$$

So, $\frac{\partial}{\partial \mu_k}\mathcal{E}(\log p(x, z|\theta))$

$$\Rightarrow \sum_{n=1}^{N}\mathcal{E}[z_{nk}]\left(\Sigma_k^{-1}x_n - \Sigma_k^{-1}\mu_k\right) = 0$$

$$\Rightarrow \sum_{n=1}^{N}\mathcal{E}[z_{nk}]x_n = \sum_{n=1}^{N}\mathcal{E}[z_{nk}]\mu_k$$



Therefore,

$$\mu_k = \frac{\sum_n \gamma_{nk} x_n}{\sum_n \gamma_{nk}}$$

→ Here, z_{nk} is taken to evaluate my answer, so I replaced it by the variable γ_{nk} in the final answer.

2

The process can be divided into estimation and maximization steps:

Estimation Step (E-step):

1. Initially, we set our model parameters: the mean (μk), covariance matrix (Σk), and mixing coefficients (πk).
2. For each data point, we compute the posterior probabilities of the data points belonging to each centroid using the current parameter values. These probabilities are often represented by the latent variables γk.
3. Finally, we estimate the values of the latent variables γk based on the current parameter values.

Maximization Step:

1. Here, we update the parameter values: the mean (μk), covariance matrix (Σk), and mixing coefficients (πk) using the estimated latent variable γk.
2. We update the mean of the cluster point (μk) by computing the weighted average of data points using the corresponding latent variable probabilities.
3. The covariance matrix (Σk) is updated by calculating the weighted average of the squared differences between the data points and the mean, using the corresponding latent variable probabilities.
4. The mixing coefficients (πk) are updated by averaging the latent variable probabilities for each component.

We repeat the E-step and M-step until convergence:

1. We iterate between the estimation step and maximization step until the change in the log-likelihood or the parameters falls below a predefined threshold or until a maximum number of iterations is reached.
2. Essentially, in the estimation step, we update the latent variables based on the current parameter values.
3. Conversely, in the maximization step, we update the parameter values using the estimated latent variables.
4. This iterative process continues until our model converges.