Homework 4 Part 1

This is an individual assignment.

Solve all problems by hand. You may type your answers in markdown cells or `push` a PDF file with your handwritten answers.

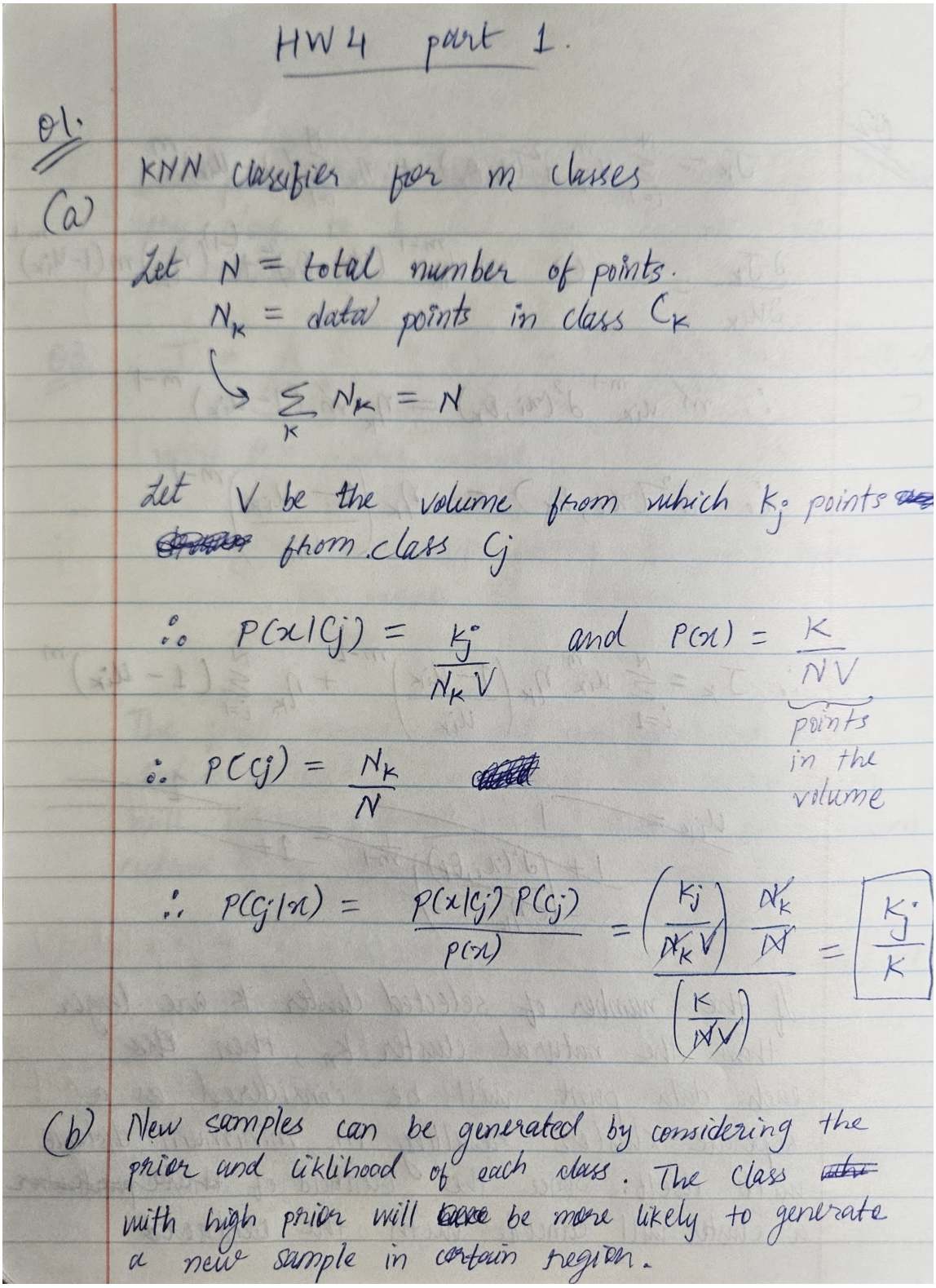
Problem 1 (10 points)

Consider the k-Nearest Neighbors (kNN) classifier for m classes.

Define the class prior probability $P(C_j)$, data likelihood $p(\mathbf{x}|C_j)$, and show that the KNN's posterior probability is

$$P(C_j|\mathbf{x}) = \frac{k_j}{k}$$

where k_j is the number of neighbors from class C_j. In addition, explain how you can generate new samples.



Problem 2 (15 points)

Consider the objective function for the Possibilistic C-Means (PCM)

$$J(\Theta, \mathbf{U}, \mathbf{H}) = \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}^{m} d^2(\mathbf{x}_i, \theta_k) + \sum_{k=1}^{K}\eta_k \sum_{i=1}^{N}(1-u_{ik})^m$$

where $m > 1$ is the *fuzzifier*, $u_{ik} \in [0, 1]$ is the membership value of sample $\mathbf{x}_i$ in cluster centroid θ_k, $d^2(\mathbf{x}_i, \theta_k)$ is the squared distance between sample $\mathbf{x}_i$ and cluster centroid θ_k, and $\eta_k > 0$ determines the relative significance of the two terms.

1. (5 points) Explain the role of the second term in the objective function for PCM.

The role of second term is to basically ensure that the membership value u_{ik} is as close to 1 as possible. So the algorithm can mitigate the zero membership assignment problems of FCM. η_k determines the relative degree to which the second term is important compared with the first term.

1. (10 points) Observe that, for each vector $\mathbf{x}_i$, the u_{ik}'s, $k = 1, \ldots, K$, are independent of each other, we can write $J(\Theta, \mathbf{U}, \mathbf{H})$ as

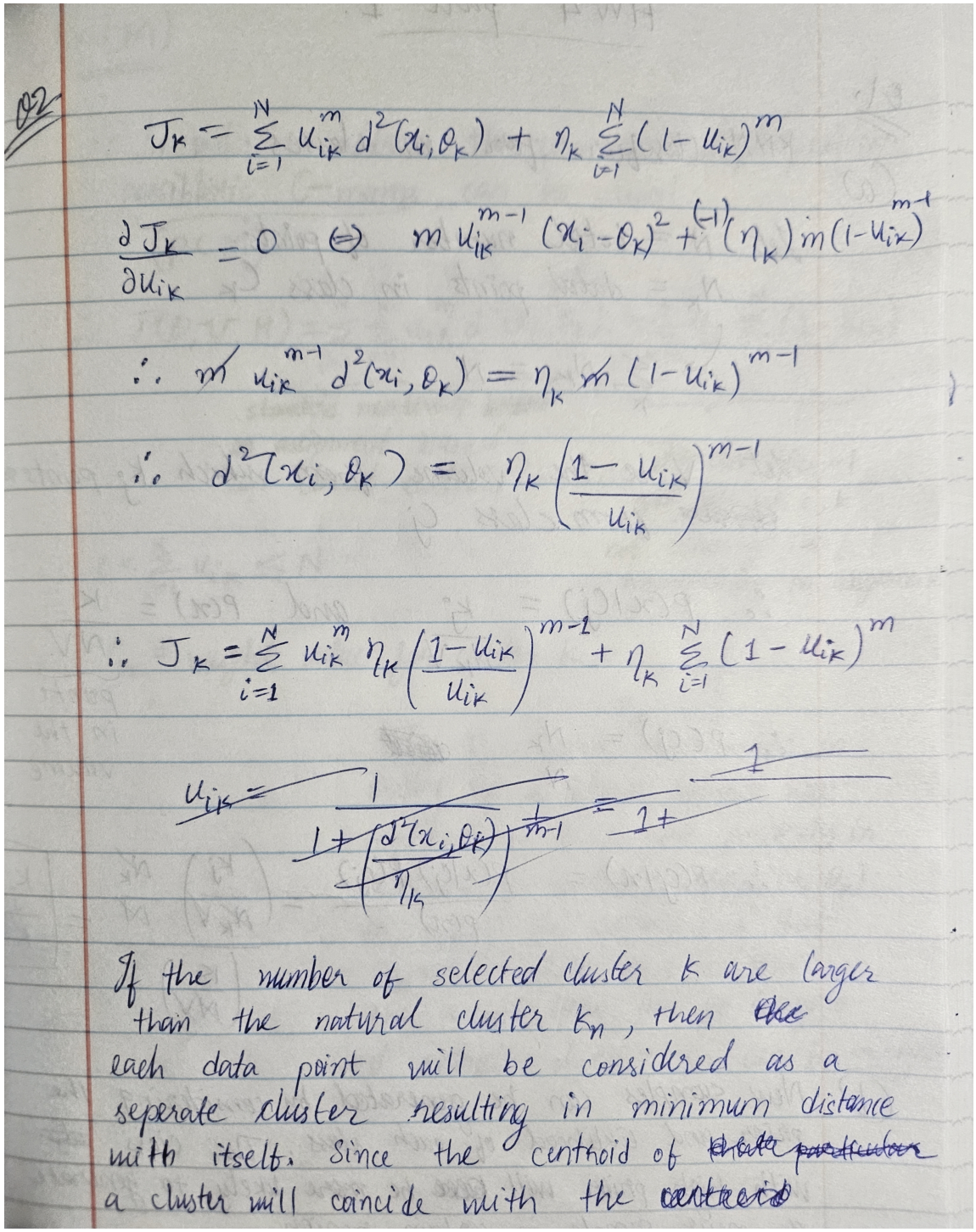
$$J(\Theta, \mathbf{U}, \mathbf{H}) = \sum_{k=1}^{K} J_k$$

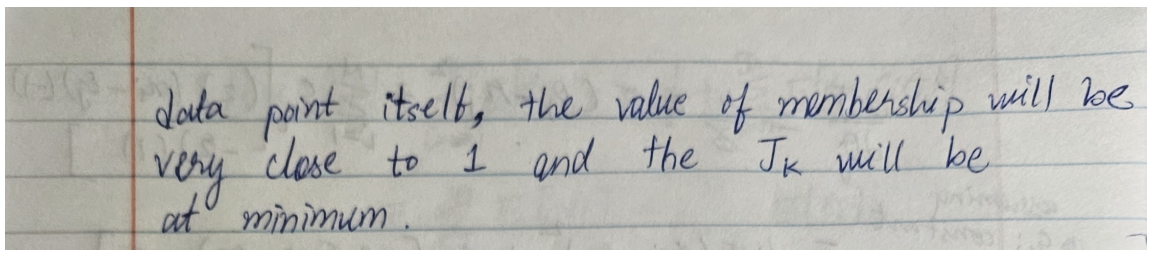
where

$$J_k = \sum_{i=1}^{N} u_{ik}^m d^2(\mathbf{x}_i, \theta_k) + \eta_k \sum_{i=1}^{N} (1 - u_{ik})^m$$

Each J_k corresponds to a different cluster and the minimization of $J(\Theta, \mathbf{U}, \mathbf{H})$ with respect to the u_{ik}'s can be carried out separately for each J_k.

Solve for J_k as a function of u_{ik}, η_k and m. For a fixed η_k, describe the placement of the centroids θ_k in the minimization of J. Discuss the implications of this behavior when the number of selected clusters K is *larger* then the number K_n of natural clusters in $\mathbf{X}$ (i.e. $K > K_n$).





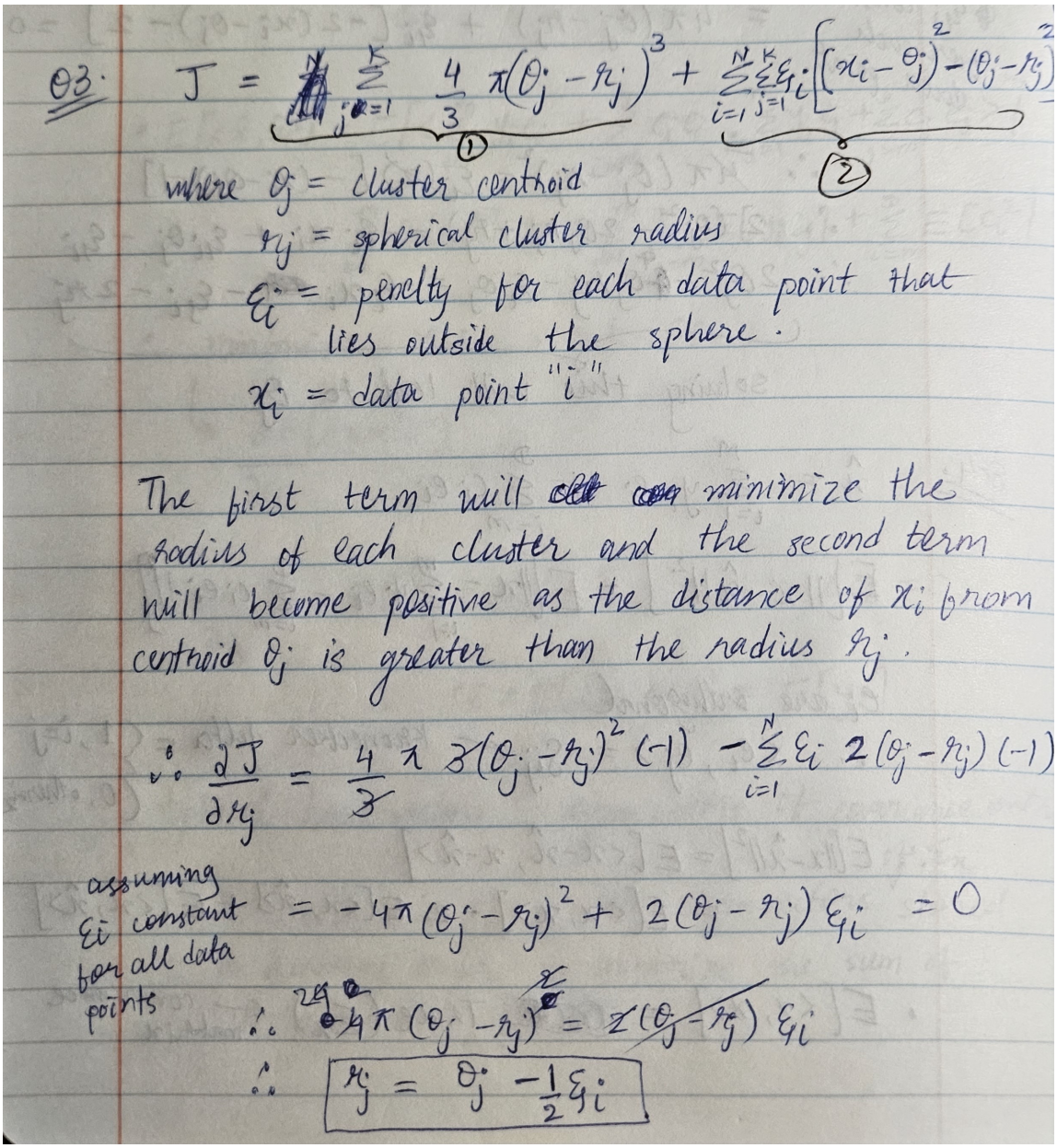
Problem 3 (10 points)

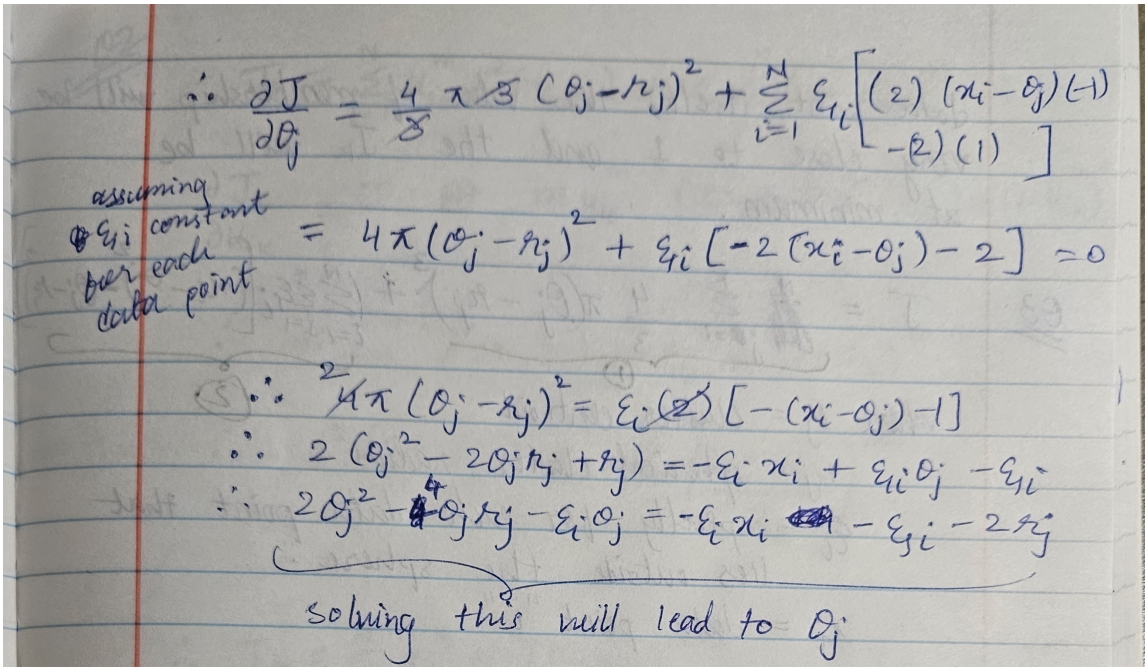
Suppose that you are interested in designing a clustering method that will cluster a dataset into spherical clusters with radius r_j centered at centroid θ_j, $j = 1, \ldots, K$.

Conditions:

- The solution must find $r_j, j = 1, \ldots, K$ so that all K clusters have **minimum volume**.
- Allow a **minimal number of points** to lie outside the volume of its respective sphere/cluster.
- Introduce a penalty $\xi_i > 0, i = 1, \ldots, N$ for points lying outside sphere.

Design an objective function for this clustering algorithm satisfying the conditions above. Define all hyperaparameters (if any), parameters, and solve for parameter solutions. Explain your reasoning.





Problem 4 (10 points)

Let $\mathbf{e}_i, i = 1, \ldots, D$, be any orthonormal basis in the D-dimensional space. Consider a D-dimensional random vector $\mathbf{x}$, which is approximated by

$$\hat{\mathbf{x}} = \sum_{i=1}^{m} y_i \mathbf{e}_i + \sum_{i=m}^{D} c_i \mathbf{e}_i$$

where c_i are non-random constants. Show that the minimum mean square error $\mathbb{E}\left[\|\mathbf{x} - \hat{\mathbf{x}}\|^2\right]$ is achieved if

1. $c_i = \mathbb{E}[y_i], i = m, \ldots, D$,
2. the orthonormal basis consists of the eigenvectors of Σ_x (covariance of $\mathbf{X}$); and
3. $\mathbf{e}_i, i = m, \ldots, D$, correspond to the eigenvectors associated with the $D - m$ smallest eigenvalues.

Q4. $\hat{x} = \sum_{i=1}^{M} y_i e_i + \sum_{i=m}^{D} c_i e_i$

$$E\left[\|x - \hat{x}\|^2\right] = E\left[\|x - \sum_{i=1}^{m} y_i e_i - \sum_{i=m}^{D} c_i e_i\|^2\right]$$

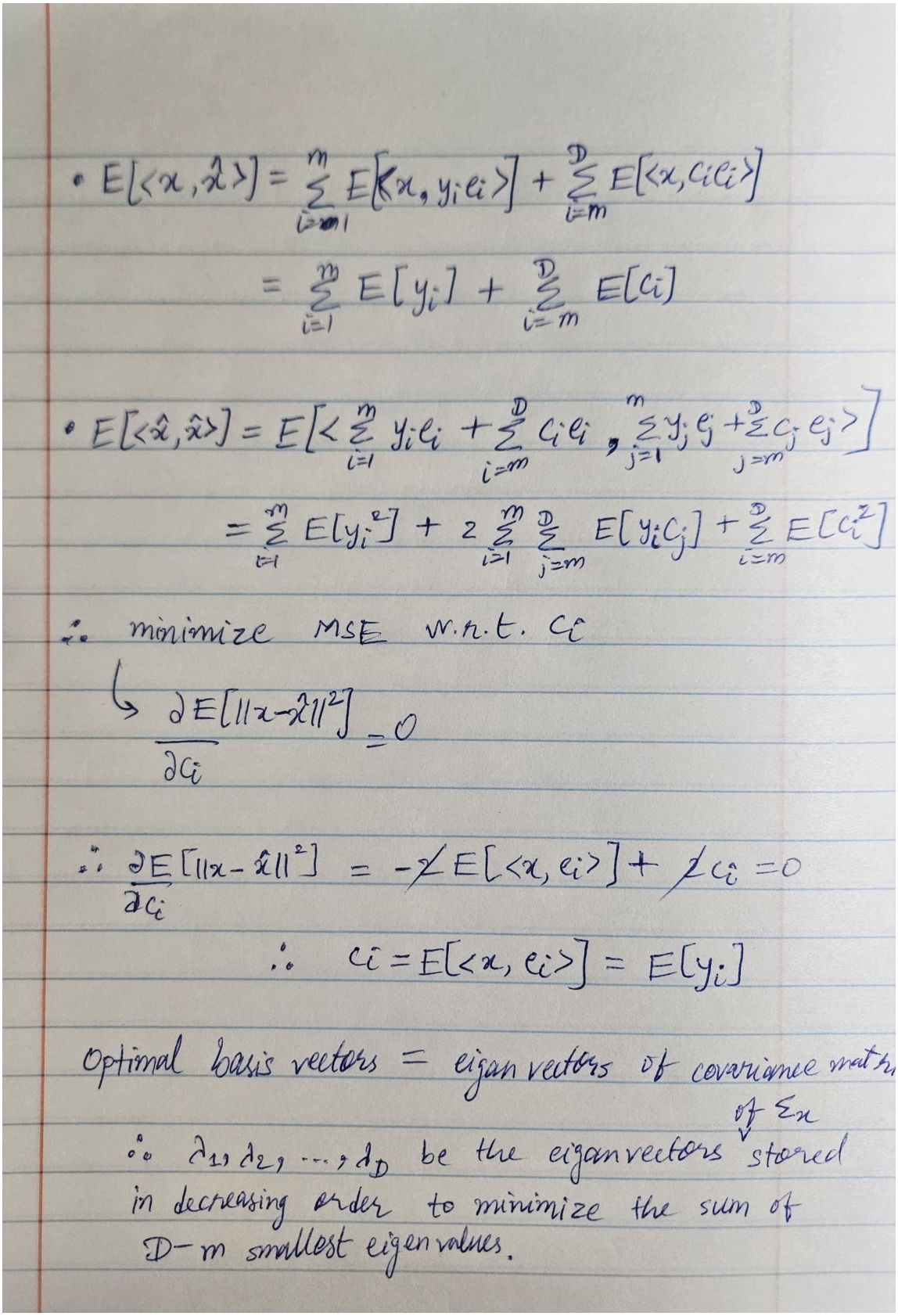
e_i are orthogonal

$\therefore \langle e_i, e_j \rangle = \delta_{ij}$ ← Kronecker delta $= \begin{cases} 1, & i=j \\ 0, & \text{otherwise} \end{cases}$

$\therefore E[\|x-\hat{x}\|^2] = E[\langle x-\hat{x}, x-\hat{x}\rangle]$

$= E[\langle x, x\rangle] - 2E[\langle x, \hat{x}\rangle] + E[\langle \hat{x}, \hat{x}\rangle]$

- $E[\langle x, x\rangle] =$ Trace (Σ_x) ← covariance matrix.



On-Time (5 points)

Submit your assignment before the deadline.

Submit Your Solution

Confirm that you've successfully completed the assignment.

Along with the Notebook, include a PDF of the notebook with your solutions.

`add` and `commit` the final version of your work, and `push` your code to your GitHub repository.

Submit the URL of your GitHub Repository as your assignment submission on Canvas.
