Statistical Pattern Recognition - Homework 1
Question 14

Vahid Maleki
Student ID: 40313004

October 11, 2025

Question 14

In an n-dimensional space, there are N samples belonging to one class:

$$X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}.$$

By substituting the mathematical relations, show that the following pseudocode incrementally estimates the sample mean (mean_x) and the sample covariance matrix (S):

Algorithm 1 Online Correlation

```
mean_x = C = n = S = 0
for i = 1 to N do
    n ← n + 1
    d_i ← x_i − mean_x
    mean_x ← mean_x + d_i / n
    C ← C + (1 − 1/i) d_i d_i^T
    S ← C/(i−1) + mean_x mean_x^T
end for
```

Solution

We need to verify that the pseudocode correctly computes the sample mean mean_x and the sample covariance matrix S incrementally for N samples $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$, where each $\mathbf{x}_i$ is an n-dimensional vector.

The sample mean for k samples is:

$$\text{mean}_x^{(k)} = \frac{1}{k}\sum_{i=1}^{k}\mathbf{x}_i.$$

The sample covariance matrix for k samples is:

$$S^{(k)} = \frac{1}{k-1}\sum_{i=1}^{k}(\mathbf{x}_i - \text{mean}_x^{(k)})(\mathbf{x}_i - \text{mean}_x^{(k)})^T.$$

We will show that the pseudocode updates mean_x and S correctly at each iteration.

The pseudocode initializes $\text{mean}_x = 0$, $n = 0$. For each sample $\mathbf{x}_i$: - Increment $n \leftarrow n+1$, so after the i-th iteration, $n = i$. - Compute the difference: $\mathbf{d}_i = \mathbf{x}_i - \text{mean}_x$, where mean_x is the mean after $i-1$ samples, i.e., $\text{mean}_x = \text{mean}_x^{(i-1)}$. - Update the mean: $\text{mean}_x \leftarrow \text{mean}_x + \frac{\mathbf{d}_i}{n}$.

Let's derive the mean update. After $i-1$ samples, the mean is:

$$\text{mean}_x^{(i-1)} = \frac{1}{i-1}\sum_{j=1}^{i-1}\mathbf{x}_j \quad (\text{if } i-1>0, \text{else } 0).$$

At iteration i:

$$\mathbf{d}_i = \mathbf{x}_i - \text{mean}_x^{(i-1)}.$$

The mean update is:

$$\text{mean}_x^{(i)} = \text{mean}_x^{(i-1)} + \frac{\mathbf{d}_i}{i} = \text{mean}_x^{(i-1)} + \frac{\mathbf{x}_i - \text{mean}_x^{(i-1)}}{i}.$$

Rewrite:

$$\text{mean}_x^{(i)} = \frac{(i-1)\text{mean}_x^{(i-1)} + \mathbf{x}_i}{i}.$$

Substitute $\text{mean}_x^{(i-1)} = \frac{1}{i-1}\sum_{j=1}^{i-1}\mathbf{x}_j$:

$$(i-1)\text{mean}_x^{(i-1)} = \sum_{j=1}^{i-1}\mathbf{x}_j.$$

Thus:

$$\text{mean}_x^{(i)} = \frac{\sum_{j=1}^{i-1}\mathbf{x}_j + \mathbf{x}_i}{i} = \frac{1}{i}\sum_{j=1}^{i}\mathbf{x}_j.$$

This matches the definition of the sample mean after i samples, confirming the mean update is correct.

The pseudocode initializes $C = 0$, $S = 0$. For each sample: - Update $C \leftarrow C + \left(1 - \frac{1}{i}\right)\mathbf{d}_i\mathbf{d}_i^T$. - Update $S \leftarrow \frac{C}{i-1} + \text{mean}_x\text{mean}_x^T$.

The sample covariance matrix after k samples is:

$$S^{(k)} = \frac{1}{k-1}\sum_{i=1}^{k}(\mathbf{x}_i - \text{mean}_x^{(k)})(\mathbf{x}_i - \text{mean}_x^{(k)})^T.$$

However, the pseudocode computes S using C, which seems to accumulate a sum related to the covariance. Let's analyze the update for C:

$$C^{(i)} = C^{(i-1)} + \left(1 - \frac{1}{i}\right)\mathbf{d}_i\mathbf{d}_i^T,$$

where $\mathbf{d}_i = \mathbf{x}_i - \text{mean}_x^{(i-1)}$.

To understand C, compute it iteratively: - For $i = 1$, $n = 1$, $\text{mean}_x^{(0)} = 0$, so $\mathbf{d}_1 = \mathbf{x}_1$, and:

$$C^{(1)} = \left(1 - \frac{1}{1}\right) \mathbf{d}_1 \mathbf{d}_1^T = 0.$$

- For $i = 2$, $n = 2$, $\text{mean}_x^{(1)} = \mathbf{x}_1$, so $\mathbf{d}_2 = \mathbf{x}_2 - \mathbf{x}_1$:

$$C^{(2)} = C^{(1)} + \left(1 - \frac{1}{2}\right) \mathbf{d}_2 \mathbf{d}_2^T = \frac{1}{2}(\mathbf{x}_2 - \mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)^T.$$

- For $i = 3$, $\text{mean}_x^{(2)} = \frac{\mathbf{x}_1 + \mathbf{x}_2}{2}$, $\mathbf{d}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_1 + \mathbf{x}_2}{2}$:

$$C^{(3)} = C^{(2)} + \left(1 - \frac{1}{3}\right) \mathbf{d}_3 \mathbf{d}_3^T = \frac{1}{2}(\mathbf{x}_2 - \mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)^T + \frac{2}{3}\left(\mathbf{x}_3 - \frac{\mathbf{x}_1 + \mathbf{x}_2}{2}\right)\left(\mathbf{x}_3 - \frac{\mathbf{x}_1 + \mathbf{x}_2}{2}\right)^T.$$

The general form for $C^{(i)}$:

$$C^{(i)} = \sum_{j=2}^{i} \left(1 - \frac{1}{j}\right) (\mathbf{x}_j - \text{mean}_x^{(j-1)})(\mathbf{x}_j - \text{mean}_x^{(j-1)})^T.$$

Now, examine the covariance update:

$$S^{(i)} = \frac{C^{(i)}}{i-1} + \text{mean}_x^{(i)}(\text{mean}_x^{(i)})^T.$$

The term $\text{mean}_x^{(i)}(\text{mean}_x^{(i)})^T$ suggests an adjustment to the covariance. Let's compare with the sample covariance:

$$S^{(i)} = \frac{1}{i-1} \sum_{j=1}^{i} (\mathbf{x}_j - \text{mean}_x^{(i)})(\mathbf{x}_j - \text{mean}_x^{(i)})^T.$$

The pseudocode's update for S seems incorrect because adding $\text{mean}_x^{(i)}(\text{mean}_x^{(i)})^T$ does not align with the standard covariance formula, and C uses differences from the previous mean $\text{mean}_x^{(i-1)}$, not the current mean $\text{mean}_x^{(i)}$. Let's derive the correct incremental covariance update to clarify.

The correct incremental covariance update should be:

$$S^{(i)} = \frac{1}{i-1} \sum_{j=1}^{i} (\mathbf{x}_j - \text{mean}_x^{(i)})(\mathbf{x}_j - \text{mean}_x^{(i)})^T.$$

Rewrite the sum:

$$\sum_{j=1}^{i} (\mathbf{x}_j - \text{mean}_x^{(i)})(\mathbf{x}_j - \text{mean}_x^{(i)})^T = \sum_{j=1}^{i} \mathbf{x}_j \mathbf{x}_j^T - i\text{mean}_x^{(i)}(\text{mean}_x^{(i)})^T.$$

For $i - 1$:

$$S^{(i-1)} = \frac{1}{i-2} \left(\sum_{j=1}^{i-1} \mathbf{x}_j \mathbf{x}_j^T - (i-1)\text{mean}_x^{(i-1)}(\text{mean}_x^{(i-1)})^T \right).$$

The pseudocode's $C^{(i)}$ resembles:

$$C^{(i)} \approx \sum_{j=1}^{i} (\mathbf{x}_j - \text{mean}_x^{(j-1)})(\mathbf{x}_j - \text{mean}_x^{(j-1)})^T,$$

but with the factor $1 - \frac{1}{j}$. This suggests the pseudocode may be implementing a variant of the covariance update, possibly Welford's algorithm adjusted for matrices. Let's test the correct form:

$$C^{(i)} = (i-2)S^{(i-1)} + \left(1 - \frac{1}{i}\right)(\mathbf{x}_i - \text{mean}_x^{(i-1)})(\mathbf{x}_i - \text{mean}_x^{(i-1)})^T.$$

However, the final S update adds $\text{mean}_x^{(i)}(\text{mean}_x^{(i)})^T$, which overestimates the covariance. The correct $S^{(i)}$ should be:

$$S^{(i)} = \frac{1}{i-1} \sum_{j=1}^{i} (\mathbf{x}_j - \text{mean}_x^{(i)})(\mathbf{x}_j - \text{mean}_x^{(i)})^T.$$

The pseudocode's $S \leftarrow \frac{C}{i-1} + \text{mean}_x \text{mean}_x^T$ suggests a possible error in the algorithm, as the mean term inflates S. Assuming the intent was to compute the standard covariance, the correct update for C should be:

$$C^{(i)} = C^{(i-1)} + (\mathbf{x}_i - \text{mean}_x^{(i)})(\mathbf{x}_i - \text{mean}_x^{(i-1)})^T,$$

and:

$$S^{(i)} = \frac{C^{(i)}}{i-1}.$$

However, since $\mathbf{d}_i = \mathbf{x}_i - \text{mean}_x^{(i-1)}$, the term $1 - \frac{1}{i} = \frac{i-1}{i}$ adjusts the contribution. Let's derive:

$$(\mathbf{x}_i - \text{mean}_x^{(i)}) = \mathbf{x}_i - \frac{1}{i}\sum_{j=1}^{i} \mathbf{x}_j = \mathbf{x}_i - \left(\frac{i-1}{i}\text{mean}_x^{(i-1)} + \frac{\mathbf{x}_i}{i}\right) = \frac{i-1}{i}(\mathbf{x}_i - \text{mean}_x^{(i-1)}) = \frac{i-1}{i}\mathbf{d}_i.$$

Thus:

$$(\mathbf{x}_i - \text{mean}_x^{(i)})(\mathbf{x}_i - \text{mean}_x^{(i)})^T = \frac{(i-1)^2}{i^2}\mathbf{d}_i\mathbf{d}_i^T.$$

The pseudocode uses:

$$C^{(i)} = C^{(i-1)} + \frac{i-1}{i}\mathbf{d}_i\mathbf{d}_i^T.$$

Summing:

$$C^{(i)} = \sum_{j=2}^{i} \frac{j-1}{j}\mathbf{d}_j\mathbf{d}_j^T.$$

This suggests C accumulates weighted differences. The final $S^{(i)}$ should be:

$$S^{(i)} = \frac{1}{i-1} \sum_{j=1}^{i} (\mathbf{x}_j - \text{mean}_x^{(i)})(\mathbf{x}_j - \text{mean}_x^{(i)})^T.$$

The pseudocode's S update is incorrect due to the $\text{mean}_x \text{mean}_x^T$ term. The correct S should be:

$$S^{(i)} = \frac{C^{(i)}}{i-1},$$

if $C^{(i)} = \sum_{j=1}^{i} (\mathbf{x}_j - \text{mean}_x^{(j-1)})(\mathbf{x}_j - \text{mean}_x^{(j-1)})^T$, but the factor $\frac{i-1}{i}$ and the mean term suggest a modified algorithm. Assuming the pseudocode intends Welford's method, the correct covariance update is:

$$S^{(i)} = \frac{i-2}{i-1} S^{(i-1)} + \frac{1}{i} (\mathbf{x}_i - \text{mean}_x^{(i)})(\mathbf{x}_i - \text{mean}_x^{(i)})^T.$$

Given the pseudocode's structure, it's likely a mistake. The correct C update should be:

$$C^{(i)} = C^{(i-1)} + (\mathbf{x}_i - \text{mean}_x^{(i)})(\mathbf{x}_i - \text{mean}_x^{(i-1)})^T,$$

and:

$$S^{(i)} = \frac{C^{(i)}}{i-1}.$$

However, accepting the pseudocode as given, it approximates the covariance with an additional mean term, which we note as a potential error.

In conclusion, the pseudocode correctly updates the sample mean using:

$$\text{mean}_x^{(i)} = \text{mean}_x^{(i-1)} + \frac{\mathbf{x}_i - \text{mean}_x^{(i-1)}}{i}.$$

The covariance update, however, includes an erroneous $\text{mean}_x \text{mean}_x^T$ term. The correct covariance matrix should be:

$$S^{(i)} = \frac{1}{i-1} \sum_{j=1}^{i} (\mathbf{x}_j - \text{mean}_x^{(i)})(\mathbf{x}_j - \text{mean}_x^{(i)})^T,$$

which requires adjusting the C update to use the current mean or correcting the final S computation.