Week 8 Exercise Sheet Solutions

The following exercises have different levels of difficulty indicated by (*), (**), (***). An exercise with (*) is a simple exercise requiring less time or effort to solve compared to an exercise with (***), which is a more complex exercise.

Unsupervised Learning

1. (*) We want to use PCA to reduce dimensionality from 3 to 2. The covariance matrix of the data is

$$\mathbf{C} = \begin{pmatrix} 4 & -2 & 0 \\ -2 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \tag{1}$$

and the corresponding eigenvectors are

$$\mathbf{w}_1 = \begin{pmatrix} -0.872 \\ 0.466 \\ 0.152 \end{pmatrix}, \mathbf{w}_2 = \begin{pmatrix} -0.390 \\ -0.847 \\ 0.361 \end{pmatrix}, \mathbf{w}_3 = \begin{pmatrix} 0.297 \\ -0.256 \\ 0.920 \end{pmatrix}. \tag{2}$$

Using the eigenvalue equation $\lambda_i = \mathbf{w}_i \mathbf{C} \mathbf{w}_i$, show that the eigenvalues are $\lambda_1 = 5.070, \lambda_2 = -0.346, \lambda_3 = 2.278$ (to 3 decimal places).

Solution:
Note: There was a mistake where the third eigenvector should be

$$\mathbf{w}_3 = \begin{pmatrix} 0.297 \\ 0.256 \\ 0.920 \end{pmatrix}.$$

and there were some rounding errors that have made the calculated eigenvalues slightly different. To find the eigenvalues from the eigenvectors we can simply apply the formula

$$\lambda_i = \mathbf{w}_i^T \mathbf{C} \mathbf{w}_i. \tag{3}$$

For the first eigenvector:

$$\begin{aligned}
\lambda_1 &= \begin{pmatrix}-0.872 & 0.466 & 0.152\end{pmatrix}\begin{pmatrix}4 & -2 & 0\\ -2 & 1 & 1\\ 0 & 1 & 2\end{pmatrix}\begin{pmatrix}-0.872\\ 0.466\\ 0.152\end{pmatrix}\\
&= \begin{pmatrix}-0.872 & 0.466 & 0.152\end{pmatrix}\begin{pmatrix}-4.42\\ 2.362\\ 0.77\end{pmatrix}\\
&= -0.872 \times -4.42 + 0.466 \times 2.362 + 0.152 \times 0.77 = 5.071(972)
\end{aligned}$$

For the second eigenvector:

$$\begin{aligned}
\lambda_1 &= \begin{pmatrix}-0.390 & -0.847 & 0.361\end{pmatrix}\begin{pmatrix}4 & -2 & 0\\ -2 & 1 & 1\\ 0 & 1 & 2\end{pmatrix}\begin{pmatrix}-0.390\\ -0.847\\ 0.361\end{pmatrix}\\
&= \begin{pmatrix}-0.390 & -0.847 & 0.361\end{pmatrix}\begin{pmatrix}0.134\\ 0.294\\ -0.125\end{pmatrix}\\
&= -0.346(403)
\end{aligned}$$

For the third eigenvector:

$$\begin{aligned}
\lambda_1 &= \begin{pmatrix}0.297 & -0.256 & 0.920\end{pmatrix}\begin{pmatrix}4 & -2 & 0\\ -2 & 1 & 1\\ 0 & 1 & 2\end{pmatrix}\begin{pmatrix}0.297\\ 0.256\\ 0.920\end{pmatrix}\\
&= \begin{pmatrix}0.297 & 0.256 & 0.920\end{pmatrix}\begin{pmatrix}0.676\\ 0.582\\ 2.096\end{pmatrix}\\
&= 2.278(084).
\end{aligned}$$

---

2. (**) Following from 1., which 2 eigenvectors should be used when applying PCA to reduce the dimensionality to 2? If we have 2 datapoints $\mathbf{x}_1 = (2,3,3)^T$ and $\mathbf{x}_2 = (4,1,0)^T$. Apply the PCA transformation to calculate the transformed datapoints. Show your steps and you assume that the datapoints have already had the mean subtracted.

Solution:
Please note while in these question we have used $\mathbf{w}$ as the eigenvectors, this corresponds to the $\mathbf{u}$ vectors used in the lab. To reduce the dimensionality using PCA the eigenvectors with the 2 highest eigenvalues should be used. For this example

eigenvalue 1 and 3 are the largest and should be selected. This means that the transformation matrix will be (we will use the incorrect eigenvalues defined in the question):

$$\mathbf{U} = \begin{pmatrix}\mathbf{w}_1 & \mathbf{w}_3\end{pmatrix} = \begin{pmatrix} -0.872 & 0.297 \\ 0.466 & -0.256 \\ 0.152 & 0.920 \end{pmatrix}.$$

The transformed data points can be calculated using

$$\mathbf{y}_i = \mathbf{U}^T \mathbf{x}_i$$

Applying this to the data points gives:

$$\begin{aligned}\mathbf{y}_1 &= \begin{pmatrix} -0.872 & 0.466 & 0.152 \\ 0.297 & -0.256 & 0.920 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} 0.11 \\ 4.122 \end{pmatrix}\end{aligned}$$

$$\begin{aligned}\mathbf{y}_w &= \begin{pmatrix} -0.872 & 0.466 & 0.152 \\ 0.297 & -0.256 & 0.920 \end{pmatrix} \begin{pmatrix} 4 \\ 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} -3.022 \\ 1.444 \end{pmatrix}\end{aligned}$$

So the new data points have now been transformed into 2D.

---

3. (***) An alternative to derive PCA is to minimise the reconstruction error. Consider the first principal component $\mathbf{u}_1$ such that a transformed data point is $y_{n1} = \mathbf{u}_1^T \mathbf{x}_n$ and the reconstructed data point is $\tilde{\mathbf{x}}_n = \mathbf{u}_1 y_{n1}$. Show that the reconstruction error

$$E = \frac{1}{2N} \sum_{n=1}^{N} |\mathbf{x}_n - \tilde{\mathbf{x}}_n|^2 \tag{4}$$

is equal to

$$E = -\mathbf{u}_1^T \mathbf{C} \mathbf{u}_1 + \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n^T \mathbf{x}_n, \tag{5}$$

where $\mathbf{C} = \sum_{n=1}^{N} \mathbf{x}\mathbf{x}^T / N$ is the covariance matrix. You will need to use the definition that the square of a vector $\mathbf{a}$ is the inner product $|\mathbf{a}|^2 = \mathbf{a}^T \mathbf{a}$ and that the principal component is normalised $\mathbf{u}_1^T \mathbf{u}_1 = 1$.

Solution:

When applying PCA using the first principal component the transformed data point is $y_{n1} = \mathbf{u}_1^T \mathbf{x}_n$ and the reconstructed data point is then given by $\tilde{\mathbf{x}}_n = \mathbf{u}_1 y_{n1}$. If we combine these together then we get

$$\tilde{\mathbf{x}}_n = \mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n$$

Notice that we have the outer product of the principal component vector and not the inner product. The reconstruction error is defined as

$$E = \frac{1}{2N} \sum_{n=1}^{N} |\mathbf{x}_n - \tilde{\mathbf{x}}_n|^2$$

so inserting our expression for the reconstructed data point gives

$$\begin{aligned}
E &= \frac{1}{2N} \sum_{n=1}^{N} \left|\mathbf{x}_n - \mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n\right|^2 \\
&= \frac{1}{2N} \sum_{n=1}^{N} \left(\mathbf{x}_n - \mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n\right)^T \left(\mathbf{x}_n - \mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n\right) \\
&= \frac{1}{2N} \sum_{n=1}^{N} \left[\mathbf{x}_n^T \mathbf{x}_n - \mathbf{x}_n^T \mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n - \left(\mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n\right)^T \mathbf{x}_n + \left(\mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n\right)^T \mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n\right]
\end{aligned}$$

To apply the transpose to the vector products we must reorder the order of the products (e.g $(AB)^T = B^T A^T$) so

$$\left(\mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n\right)^T = \mathbf{x}_n^T \left(\mathbf{u}_1 \mathbf{u}_1^T\right)^T = \mathbf{x}_n^T \mathbf{u}_1 \mathbf{u}_1^T$$

Using this in our expression for the reconstruction error gives

$$\begin{aligned}
E &= \frac{1}{2N} \sum_{n=1}^{N} \left[\mathbf{x}_n^T \mathbf{x}_n - \mathbf{x}_n^T \mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n - \mathbf{x}_n^T \mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n + \mathbf{x}_n^T \mathbf{u}_1 \mathbf{u}_1^T \mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n\right] \\
&= \frac{1}{2N} \sum_{n=1}^{N} \left[|\mathbf{x}_n|^2 - 2\mathbf{x}_n^T \mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n + \mathbf{x}_n^T \mathbf{u}_1 (\mathbf{u}_1^T \mathbf{u}_1) \mathbf{u}_1^T \mathbf{x}_n\right]
\end{aligned}$$

and since the middle of the last term has magnitude of the eigenvector $\mathbf{u}_1^T \mathbf{u}_1 = 1$ this simplifies to

$$\begin{aligned}
E &= \frac{1}{2N} \sum_{n=1}^{N} \left[|\mathbf{x}_n|^2 - 2\mathbf{x}_n^T \mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n + \mathbf{x}_n^T \mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n\right] \\
&= \frac{1}{2N} \sum_{n=1}^{N} \left[|\mathbf{x}_n|^2 - \mathbf{x}_n^T \mathbf{u}_1 \mathbf{u}_1^T \mathbf{x}_n\right].
\end{aligned}$$

If we recognise that $\mathbf{x}_n^T\mathbf{u}_1$ and $\mathbf{u}_1^T\mathbf{x}_n$ are scalars so the order that we have them can be swapped with no issue and because $\mathbf{u}_1$ is independent of the sum over n they can be factored out of the sum. Using this

$$\begin{aligned} E &= \frac{1}{2N}\sum_{n=1}^{N}\left[|\mathbf{x}_n|^2 - \mathbf{u}_1^T\mathbf{x}_n\mathbf{x}_n^T\mathbf{u}_1\right] \\ &= -\mathbf{u}_1^T\left(\frac{1}{2N}\sum_{n=1}^{N}\mathbf{x}_n\mathbf{x}_n^T\right)\mathbf{u}_1 + \frac{1}{2N}\sum_{n=1}^{N}|\mathbf{x}_n|^2 \end{aligned}$$

Recall that the definition of the covariance (or scatter) matrix is

$$\mathbf{C} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n\mathbf{x}_n^T$$

so our final result is

$$E = -\frac{1}{2}\mathbf{u}_1^T\mathbf{C}\mathbf{u}_1 + \frac{1}{2N}\sum_{n=1}^{N}|\mathbf{x}_n|^2$$

which is the required result (the 1/2 is correct here and missing in the original question but since it is a constant it does not affect the result). Importantly this is telling us that minimising the reconstruction is also minimising the negative variance (remember that $\mathbf{u}_1^T\mathbf{C}\mathbf{u}_1$ is the variance of the data along direction $\mathbf{u}_1$). The second term is a constant as it is simply the average magnitude of the data point vectors and does not depend on the principal component directions.
