PCA example computation

Consider the following dataset, where the number of data points is $N = 4$ (each data point $\mathbf{x}_i \in \mathbb{R}^2$ is a row) and the number of features (dimensions) is $D = 2$:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \mathbf{x}_3^\top \\ \mathbf{x}_4^\top \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & -1 \\ 0 & 2 \end{bmatrix}.$$

Now let's compute the PCA representation for this dataset:

1. The first step is to compute the empirical covariance matrix

   $$\Sigma = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})^\top(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}) = \frac{1}{N}\left(\mathbf{X} - \mathbf{1}_N\boldsymbol{\mu}^\top\right)^\top\left(\mathbf{X} - \mathbf{1}_N\boldsymbol{\mu}^\top\right),$$

   where $\widehat{\boldsymbol{\mu}}$ is the empirical mean. In this case, $\widehat{\boldsymbol{\mu}} = \begin{bmatrix}1.5 & 1\end{bmatrix}$, and the empirical covariance matrix is given by (verify this computation; its simple!):

   $$\Sigma = \frac{1}{4}\begin{bmatrix} 5 & -5 \\ -5 & 6 \end{bmatrix} = \begin{bmatrix} 1.25 & -1.25 \\ -1.25 & 1.5 \end{bmatrix}$$

   TIP: If your computed covariance matrix is *not* symmetric, you know your computation is necessarily wrong as covariance matrices are symmetric positive definite by construction (this is an easy check to ensure you are on the right track).

2. Next, we compute the eigendecomposition i.e. the eigenvalues and eigenvectors of Σ. The eigenvalues of a 2×2 symmetric matrix $A(= \Sigma)$ can be computed using

   $$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}, \quad \det(A - \lambda I) = 0 \implies \lambda = \frac{a + c \pm \sqrt{(a-c)^2 + 4b^2}}{2}.$$

   In this case, the larger eigenvalue is clearly given by

   $$\lambda_1 = \frac{a + c + \sqrt{(a-c)^2 + 4b^2}}{2} \approx 2.631 > \lambda_2 = \frac{a + c - \sqrt{(a-c)^2 + 4b^2}}{2} \approx 0.119.$$

   The eigenvectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ corresponding to $\{\lambda_1, \lambda_2\}$ are given by (again, simply solving the equation $A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$:

   $$\boldsymbol{v}_1 = \begin{bmatrix} \frac{\lambda_1 - c}{b} \\ 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{v}_2 = \begin{bmatrix} \frac{\lambda_2 - c}{b} \\ 1 \end{bmatrix}.$$

   Next, we need to check orthonormality of the eigenvectors. It is easy to verify that they are already orthogonal, and now we just need to unit-normalize them:

   $$\overline{\boldsymbol{v}}_1 = \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|} = \frac{1}{\sqrt{1 + (\frac{\lambda_1 - c}{2})^2}}\begin{bmatrix} \frac{\lambda_1 - c}{b} \\ 1 \end{bmatrix}, \quad \text{and similarly for } \overline{\boldsymbol{v}}_2 \ldots$$

3. To pick the top principal component ($K = 1$), we need to choose $\mathbf{U} = \begin{bmatrix}\overline{\boldsymbol{v}}_1\end{bmatrix} \in \mathbb{R}^{D \times 1}$, and then compute the latent (PCA) representation for each data point $\mathbf{x}_i$ as

   $$z_i = \mathbf{U}^\top(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}) \implies \mathbf{Z} = \left(\mathbf{X} - \mathbf{1}_N\boldsymbol{\mu}^\top\right)\mathbf{U} = \ldots \quad \text{(finish this!)}$$

   NOTE: The above computation can easily be extended to $K > 1$ simply by taking the top–K (largest orthonormal) eigenvectors and stacking them into $\mathbf{U} \in \mathbb{R}^{D \times K}$.

On the exam, of course we will try to make the numbers pretty.