Exercise: Given the following training data:

$$\mathbf{x}^{(1)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \mathbf{x}^{(4)} = \begin{pmatrix} 6 \\ 3 \end{pmatrix}$$

a) Compute the K-L transformation.

Not rocket science.

Solution:
First, we compute the mean vector:

$$\mu = \frac{1}{4}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 4 \\ 0 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 6 \\ 3 \end{pmatrix}\right) = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

Now, we can compute the covariance matrix. First, we compute the individual contributions from each point:

$$\begin{aligned}
\mathbf{x}^{(1)} &\rightarrow \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix}\right)\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix}\right)^T &= \begin{pmatrix} -3 \\ -1 \end{pmatrix}\begin{pmatrix} -3 & -1 \end{pmatrix} &= \begin{pmatrix} 9 & 3 \\ 3 & 1 \end{pmatrix} \\
\mathbf{x}^{(2)} &\rightarrow \left(\begin{pmatrix} 4 \\ 0 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix}\right)\left(\begin{pmatrix} 4 \\ 0 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix}\right)^T &= \begin{pmatrix} 1 \\ -1 \end{pmatrix}\begin{pmatrix} 1 & -1 \end{pmatrix} &= \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \\
\mathbf{x}^{(3)} &\rightarrow \left(\begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix}\right)\left(\begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix}\right)^T &= \begin{pmatrix} -1 \\ 0 \end{pmatrix}\begin{pmatrix} -1 & 0 \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \\
\mathbf{x}^{(4)} &\rightarrow \left(\begin{pmatrix} 6 \\ 3 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix}\right)\left(\begin{pmatrix} 6 \\ 3 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix}\right)^T &= \begin{pmatrix} 3 \\ 2 \end{pmatrix}\begin{pmatrix} 3 & 2 \end{pmatrix} &= \begin{pmatrix} 9 & 6 \\ 6 & 4 \end{pmatrix}
\end{aligned}$$

Now, we can get the covariance matrix:

$$C = \frac{1}{4-1}\left(\begin{pmatrix} 9 & 3 \\ 3 & 1 \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 9 & 6 \\ 6 & 4 \end{pmatrix}\right) = \frac{1}{3}\begin{pmatrix} 20 & 8 \\ 8 & 6 \end{pmatrix} = \begin{pmatrix} \frac{20}{3} & \frac{8}{3} \\ \frac{8}{3} & 2 \end{pmatrix}$$

Git covariance matrix.

To get the intended transformation we need to compute C's eigenvectors. We know that the eigenvalues λ of a matrix are the roots of the characteristic polynomial:

$$|\mathbf{C} - \lambda\mathbf{I}| = 0$$

Git eigenvalues of the covariance matrix. Use algebra.

$$\begin{aligned}
|\mathbf{C} - \lambda\mathbf{I}| &= \left|\begin{pmatrix} \frac{20}{3} & \frac{8}{3} \\ \frac{8}{3} & 2 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right| \\
&= \left|\begin{pmatrix} \frac{20}{3} - \lambda & \frac{8}{3} - 0 \\ \frac{8}{3} - 0 & 2 - \lambda \end{pmatrix}\right| \\
&= \left(\frac{20}{3} - \lambda\right)(2 - \lambda) - \left(\frac{8}{3} - 0\right)\left(\frac{8}{3} - 0\right) \\
&= \frac{40}{3} - \frac{20}{3}\lambda - 2\lambda + \lambda^2 - \frac{64}{9} \\
&= \lambda^2 - \frac{26}{3}\lambda + \frac{56}{9} \\
&= 0
\end{aligned}$$

Solving the second degree equation, we get:

$$\lambda_1 = 0.79 \vee \lambda_2 = 7.88$$

Then, git eigenvectors and normalize them.

Having the eigenvalues, we can get the eigenvectors. If we have that $\lambda_1 = 0.79$, then the corresponding eigenvector $u_1 = \begin{pmatrix} u_{11} \\ u_{12} \end{pmatrix}$ will verify the following:

$$\mathbf{C}\mathbf{u}_1 = \lambda_1 \mathbf{u}_1 \iff (\mathbf{C} - \lambda_1 \mathbf{I})\,\mathbf{u}_1 = 0$$

Which yields:

$$(\mathbf{C} - \lambda_1 \mathbf{I})\,\mathbf{u}_1 = \mathbf{0}$$

$$\left[\begin{pmatrix} \frac{20}{3} & \frac{8}{3} \\ \frac{8}{3} & 2 \end{pmatrix} - \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_1 \end{pmatrix}\right]\begin{pmatrix} u_{11} \\ u_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} \frac{20}{3} - \lambda_1 & \frac{8}{3} \\ \frac{8}{3} & 2 - \lambda_1 \end{pmatrix}\begin{pmatrix} u_{11} \\ u_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} \left(\frac{20}{3} - \lambda_1\right) u_{11} + \frac{8}{3} u_{12} \\ \frac{8}{3} u_{11} + (2 - \lambda_1)\, u_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

So, the following condition must be met:

$$u_{12} = -\frac{3}{8}\left(\frac{20}{3} - \lambda_1\right) u_{11}$$

With that, a solution for the system will be:

$$\mathbf{u}_1 = \begin{pmatrix} u_{11} \\ -\frac{3}{8}\left(\frac{20}{3} - \lambda_1\right) u_{11} \end{pmatrix}$$

We can choose for instance $u_{11} = 1$ and get:

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ -\frac{3}{8}\left(\frac{20}{3} - 0.79\right) \end{pmatrix} = \begin{pmatrix} 1 \\ -2.2 \end{pmatrix}$$

However, it is usual to work with normalized eigenvectors:

$$\mathbf{u}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|_2} = \begin{pmatrix} 0.4138 \\ -0.9104 \end{pmatrix}$$

For $\lambda_2 = 7.88$, the corresponding eigenvector $\mathbf{u}_2 = \begin{pmatrix} u_{21} \\ u_{22} \end{pmatrix}$ will verify the following:

$$\mathbf{C}\mathbf{u}_2 = \lambda_2 \mathbf{u}_2 \iff (\mathbf{C} - \lambda_2 \mathbf{I})\,\mathbf{u}_2 = 0$$

Which yields:

$$(\mathbf{C} - \lambda_2 \mathbf{I})\,\mathbf{u}_2 = \mathbf{0}$$

$$\left[\begin{pmatrix} \frac{20}{3} & \frac{8}{3} \\ \frac{8}{3} & 2 \end{pmatrix} - \begin{pmatrix} \lambda_2 & 0 \\ 0 & \lambda_2 \end{pmatrix}\right]\begin{pmatrix} u_{21} \\ u_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} \frac{20}{3} - \lambda_2 & \frac{8}{3} \\ \frac{8}{3} & 2 - \lambda_2 \end{pmatrix}\begin{pmatrix} u_{21} \\ u_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} \left(\frac{20}{3} - \lambda_2\right) u_{21} + \frac{8}{3} u_{22} \\ \frac{8}{3} u_{21} + (2 - \lambda_2)\, u_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Again, the following condition must be met:

$$u_{22} = -\frac{3}{8}\left(\frac{20}{3} - \lambda_2\right) u_{21}$$

So, a solution for the system will be:

$$\mathbf{u}_2 = \begin{pmatrix} u_{21} \\ -\frac{3}{8}\left(\frac{20}{3} - \lambda_1\right) u_{21} \end{pmatrix}$$

We can choose for instance $u_{21} = 1$ and get:

$$\mathbf{u}_2 = \begin{pmatrix} 1 \\ -\frac{3}{8}\left(\frac{20}{3} - 7.88\right) \end{pmatrix} = \begin{pmatrix} 1 \\ 0.45 \end{pmatrix}$$

However, it is usual to work with normalized eigenvectors:

$$\mathbf{u}_2 = \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|_2} = \begin{pmatrix} 0.9119 \\ 0.4104 \end{pmatrix}$$

Git final transformation matrix.

Having the eigenvectors, we can place them on the columns of a matrix to build the K-L tranformation:

$$U_{K-L} = \begin{pmatrix} 0.4138 & 0.9119 \\ -0.9104 & 0.4104 \end{pmatrix}$$

b) What is the rotation applied to go from the original coordinate system to the eigenvector coordinate system?

Huge brain question.

Solution:

$$\mathbf{U}\mathbf{e}_1 = \begin{pmatrix} 0.4138 & 0.9119 \\ -0.9104 & 0.4104 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mathbf{u}_1$$

Git angle.

$$\mathbf{u}_1\mathbf{e}_1 = \left\|\begin{pmatrix} 0.4138 \\ -0.9104 \end{pmatrix}\right\|_2 \left\|\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\|_2 \cos\alpha = 0.4138$$

$$\alpha = \arccos(0.4138) = 1.1442rad \simeq 65deg$$

c) Which eigenvector is most significant?

The one with the biggest eigenvalue. In this case, u2.

d) Can we apply the Kaiser criterion?

Solution:

Yes, since $\lambda_1 < 1$ we can apply the criterion and discard it.

e) Map the points onto the most significant dimension.

Solution:

We can map the points to the eigenspace through $\mathbf{x}_{eig} = \mathbf{U}^T\mathbf{x}$:

$$\mathbf{x}_{eig}^{(1)} = \begin{pmatrix} 0.4138 & 0.9119 \\ -0.9104 & 0.4104 \end{pmatrix}^T \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\mathbf{x}_{eig}^{(2)} = \begin{pmatrix} 0.4138 & 0.9119 \\ -0.9104 & 0.4104 \end{pmatrix}^T \begin{pmatrix} 4 \\ 0 \end{pmatrix} = \begin{pmatrix} 1.66 \\ 3.65 \end{pmatrix}$$

$$\mathbf{x}_{eig}^{(3)} = \begin{pmatrix} 0.4138 & 0.9119 \\ -0.9104 & 0.4104 \end{pmatrix}^T \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.08 \\ 2.23 \end{pmatrix}$$

$$\mathbf{x}_{eig}^{(4)} = \begin{pmatrix} 0.4138 & 0.9119 \\ -0.9104 & 0.4104 \end{pmatrix}^T \begin{pmatrix} 6 \\ 3 \end{pmatrix} = \begin{pmatrix} -0.25 \\ 6.70 \end{pmatrix}$$

PCA corresponds to keeping only the most significant dimensions. In this case, we discard the first dimension and get:

$$\begin{aligned} \mathbf{x}_{PCA}^{(1)} &= \begin{pmatrix} 0 \end{pmatrix} \\ \mathbf{x}_{PCA}^{(2)} &= \begin{pmatrix} 3.65 \end{pmatrix} \\ \mathbf{x}_{PCA}^{(3)} &= \begin{pmatrix} 2.23 \end{pmatrix} \\ \mathbf{x}_{PCA}^{(4)} &= \begin{pmatrix} 6.70 \end{pmatrix} \end{aligned}$$