# Dimensionality Reduction

**Course: Data Mining**

**Professor: Dr. Tahaei**

**Herbod Pourali, Ilya Farhangfar**

**December 2024**

**Problem 1.** In a two-class classification problem with two features, each class has two data points:

$$\omega_1 : \begin{bmatrix}4\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \quad \omega_2 : \begin{bmatrix}2\\5\end{bmatrix}, \begin{bmatrix}3\\4\end{bmatrix}.$$

**(a)** Compute the mean vector and covariance matrix, then use PCA to reduce the data to a one-dimensional space. Analyze the class separability in the reduced space. Solve the problem for both principal directions.

**(b)** In part (a), propose a vector that, if added to all data points, ensures that the first principal component does not change.

**Solution:** **(a)** To calculate the mean vector, the mean of all data points, irrespective of their class, must be computed:

$$\bar{x} = \frac{1}{4}\left(\begin{bmatrix}4\\1\end{bmatrix} + \begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}2\\5\end{bmatrix} + \begin{bmatrix}3\\4\end{bmatrix}\right) = \begin{bmatrix}2.5\\2.5\end{bmatrix}.$$

Next, the covariance matrix must be calculated:

$$\tilde{X} = \begin{bmatrix}4-2.5 & 1-2.5\\1-2.5 & 0-2.5\\2-2.5 & 5-2.5\\3-2.5 & 4-2.5\end{bmatrix} = \begin{bmatrix}1.5 & -1.5\\-1.5 & -2.5\\-0.5 & 2.5\\0.5 & 1.5\end{bmatrix}.$$

$$S = \frac{1}{N}\tilde{X}^T\tilde{X} = \frac{1}{4}\begin{bmatrix}5 & 1\\1 & 17\end{bmatrix}.$$

Now, the eigenvectors and eigenvalues of the covariance matrix must be determined:

$$Sv = \lambda v$$

The eigenvalues are:

$$\lambda_1 = -\sqrt{37} + 11, \quad \lambda_2 = \sqrt{37} + 11$$

The corresponding eigenvectors are:

$$v_1 = \begin{bmatrix}-\sqrt{37} - 6\\1\end{bmatrix}, \quad v_2 = \begin{bmatrix}\sqrt{37} - 6\\1\end{bmatrix}.$$

Now, project the data onto the first principal direction:

$$X' = XA = Xv_1$$

Using a numerical approximation:

$$\sqrt{37} \approx 6.08.$$

Projecting the data onto the first principal direction:

$$X' = \begin{bmatrix}4 & 1\\1 & 0\\2 & 5\\3 & 4\end{bmatrix}\begin{bmatrix}0.08\\1\end{bmatrix} = \begin{bmatrix}1.32\\0.08\\5.16\\4.24\end{bmatrix}.$$

As is evident, the two classes are separable. Now, repeat this process for the second principal direction:

Project the data onto the second principal direction:

$$X' = XA = Xv_2$$

$$X' = \begin{bmatrix} 4 & 1 \\ 1 & 0 \\ 2 & 5 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -12.08 \\ 1 \end{bmatrix} = \begin{bmatrix} -47.32 \\ -12.08 \\ -19.16 \\ -32.24 \end{bmatrix}.$$

As you can see, the two classes are not separable.

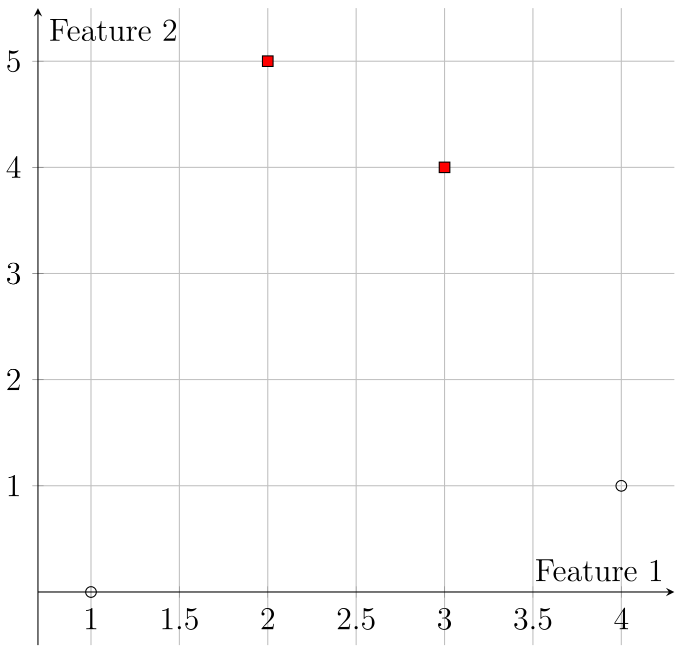
**Projection of Data onto 1D Space**

Here we show the projection of two classes onto a 1D space after dimension reduction.

**Original 2D Plot**

The following plot shows the original data points from the two classes.

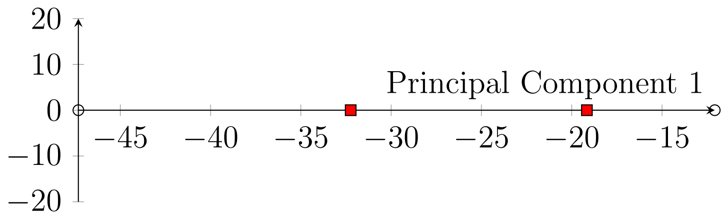
Original 2D Data Representation

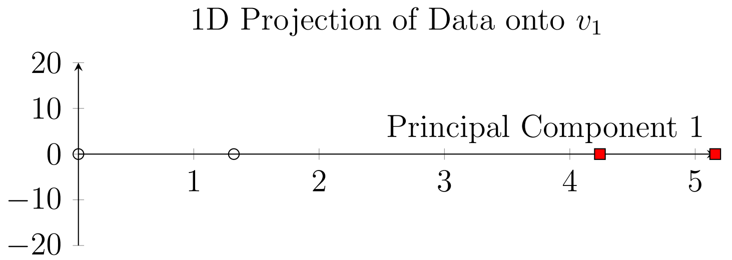



**1D Projection of Data**

After applying PCA, the data is projected onto a 1D space. The plot below shows the projection of the two classes onto this 1D space.

1D Projection of Data onto $v_2$






**Part (b):** We show that adding any arbitrary vector $\mathbf{a}$ from the eigenspace to all the data points keeps the first principal component (PC) unchanged.

Let $\bar{X}_{\text{new}}$ be the new data after adding the vector $\mathbf{a}$ to each data point:

$$\bar{X}_{\text{new}} = \frac{1}{n}\sum_{i=1}^{n}(x_i + \mathbf{a}) = \bar{X} + \mathbf{a}$$

Therefore, the covariance of the new data becomes:

$$\text{Cov}(X_{\text{new}}) = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i + \mathbf{a} - \bar{X}_{\text{new}}\right)\left(x_i + \mathbf{a} - \bar{X}_{\text{new}}\right)^T$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i + \mathbf{a} - \bar{X} - \mathbf{a}\right)\left(x_i + \mathbf{a} - \bar{X} - \mathbf{a}\right)^T$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{X}\right)\left(x_i - \bar{X}\right)^T$$

Thus, we have:

$$\text{Cov}(X_{\text{new}}) = \text{Cov}(X)$$

**Problem 2.** Consider a set of points $x^{(1)}, \ldots, x^{(m)}$. Assume that the data are normalized and have zero mean and unit variance in each dimension. Furthermore, assume that $f_u(x)$ is the projection of the point $x$ in the direction of the vector $u$. In other words, we have:

$$V = \{au : a \in \mathbb{R}\}$$

then:

$$f_u(x) = \arg\min_{v \in V} \|x - v\|^2$$

Show that the vector $\mathbf{u}$ which minimizes the MSE error between the points and their projections is the first principal component (PC1). In other words, show that:

$$\arg\min_{u:u^\top u=1} \frac{1}{m}\sum_{i=1}^{m}\|x^{(i)} - f_u(x^{(i)})\|^2$$

**Solution: (b)**

$$f_u(x) = \arg\min \|x - au\|^2 \quad \text{where} \quad v = au$$

$$\frac{\partial \|x - au\|^2}{\partial a} = \frac{\partial}{\partial a}\left((x - au)^T(x - au)\right)$$

$$= \frac{\partial}{\partial a}\left(x^T x - 2au^T x + a^2 u^T u\right)$$

$$= -2u^T x + 2au^T u = 0$$

$$\Rightarrow a = \frac{u^T x}{u^T u}$$

$$f_u(x) = au = \frac{u^T x}{u^T u} u$$

Now, let's rewrite the expression:

$$\arg\min_{\|u\|=1} \sum_{i=1}^{m} \|x^{(i)} - f_u(x^{(i)})\|^2$$

$$= \arg\min_{\|u\|=1} \sum_{i=1}^{m} \|x^{(i)} - uu^T x^{(i)}\|^2$$

$$= \arg\min_{\|u\|=1} \sum_{i=1}^{m} \left(x^{(i)} - uu^T x^{(i)}\right)^T \left(x^{(i)} - uu^T x^{(i)}\right)$$

$$= \arg\min_{\|u\|=1} \sum_{i=1}^{m} \left((x^{(i)})^T x^{(i)} - u^T x^{(i)} (x^{(i)})^T u\right)$$

$$= \arg\min_{\|u\|=1} \sum_{i=1}^{m} \left((x^{(i)})^T x^{(i)} - 2(u^T x^{(i)})^2 + (u^T x^{(i)})^2\right)$$

$$= \arg\min_{\|u\|=1} \sum_{i=1}^{m} -(u^T x^{(i)})^2$$

$$= \arg\max_{\|u\|=1} \sum_{i=1}^{m} (u^T x^{(i)})^2$$

$$= \arg\max_{\|u\|=1} u^T \left(\sum_{i=1}^{m} x^{(i)} (x^{(i)})^T\right) u$$

To solve this, we use the method of Lagrange multipliers:

$$\arg\max_{\|u\|=1} u^T \Sigma u$$

Using Lagrange multipliers:

$$L(u, \lambda) = u^T \Sigma u - \lambda(u^T u - 1)$$

To solve this equation, we compute the derivative of the Lagrangian:

$$\frac{\partial L(u,\lambda)}{\partial u} = 2\Sigma u - 2\lambda u = 0 \quad \Rightarrow \quad \Sigma u = \lambda u$$

Thus, we can conclude that the solution to this problem is the eigenvectors of the matrix $\Sigma$.

$$u^T \Sigma u = u^T \lambda u = \lambda u^T u = \lambda$$

As a result, this optimal value corresponds to the eigenvectors of the matrix $\Sigma$, which is equivalent to the first principal component in PCA.