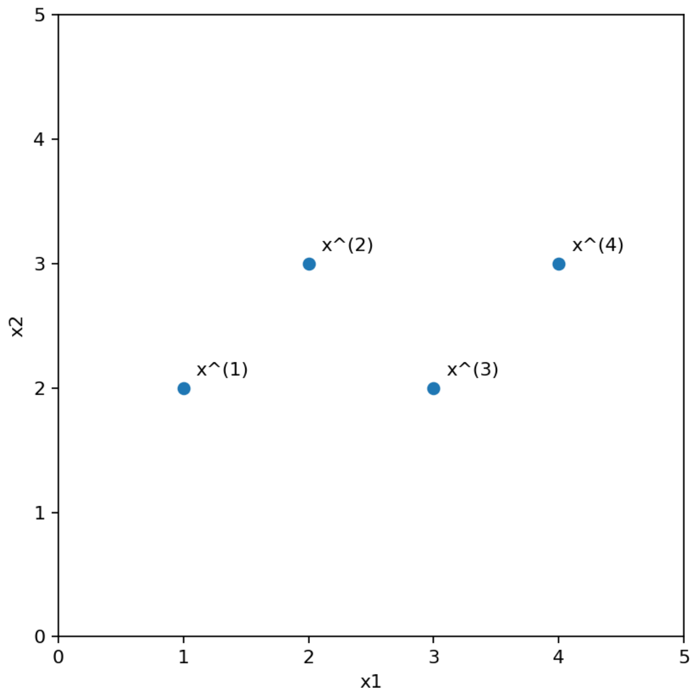
1 PCA

Consider dataset $\mathcal{D} = \{\mathbf{x}^{(1)} = \begin{bmatrix}1\\2\end{bmatrix}, \mathbf{x}^{(2)} = \begin{bmatrix}2\\3\end{bmatrix}, \mathbf{x}^{(3)} = \begin{bmatrix}3\\2\end{bmatrix}, \mathbf{x}^{(4)} = \begin{bmatrix}4\\3\end{bmatrix}\}$. A visualization of the dataset is as below.



1.1 Centering Data

Centering is crucial for PCA. We must preprocess data so that all features have zero mean before applying PCA, i.e.

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}^{(i)} = \vec{0}$$

Compute the centered dataset:

$\mathbf{x}^{(1)} =$ ____________ $\mathbf{x}^{(2)} =$ ____________

$\mathbf{x}^{(3)} =$ ____________ $\mathbf{x}^{(4)} =$ ____________

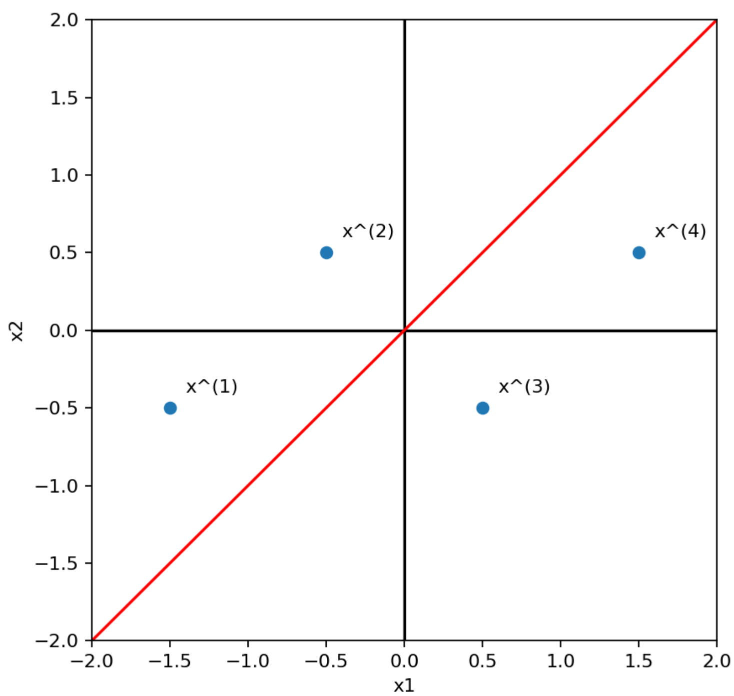
1.2 Unit vector

In order to easily compute the projected coordinates of data, we need to make the projected directions unit vectors. Suppose we want to project our data onto the vector $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Normalize $\mathbf{v}$ to be a unit vector.

$$\mathbf{v} = \_\_\_\_\_\_\_\_\_\_$$

1.3 Project Data

The centered data should now look like the following:



Suppose we want to project the centered data onto $\mathbf{v}$, where $\mathbf{v}$ goes through the origin.

Compute the magnitude of the projections, i.e. compute $z^{(i)} = \mathbf{v}^T \mathbf{x}^{(i)}, \forall 1 \leq i \leq N$.

$z^{(1)} = \_\_\_\_\_\_\_\_\_\_$ $z^{(2)} = \_\_\_\_\_\_\_\_\_\_$

$z^{(3)} = \_\_\_\_\_\_\_\_\_\_$ $z^{(4)} = \_\_\_\_\_\_\_\_\_\_$

Let $\mathbf{x}^{(i)'}$ be the projected point of $\mathbf{x}^{(i)}$. Note that $\mathbf{x}^{(i)'} = \mathbf{v}^T\mathbf{x}^{(i)}\mathbf{v} = z^{(i)}\mathbf{v}$. Compute the projected coordinates:

$\mathbf{x}^{(1)'} =$ ____________ $\mathbf{x}^{(2)'} =$ ____________

$\mathbf{x}^{(3)'} =$ ____________ $\mathbf{x}^{(4)'} =$ ____________

1.4 Reconstruction Error

One of the two goals of PCA is to find new directions to project our dataset onto such that it **minimizes the reconstruction error**, where the reconstruction error is defined as following:

$$\text{Reconstruction Error} = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{x}^{(i)'} - \mathbf{x}^{(i)}\|_2^2$$

What is the reconstruction error in our case?

Reconstruction Error = ____________

1.5 Variance of Projected Data

Another goal is to find new directions to project our dataset onto such that it **maximizes the variance of the projections**, where the variance of projections is defined as following:

$$\begin{aligned}
\text{variance of projection} &= \frac{1}{N}\sum_{i=1}^{N}(z^{(i)} - \hat{E}[z])^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{v}^T\mathbf{x}^{(i)} - \frac{1}{N}\sum_{j=1}^{N}\{\mathbf{v}^T\mathbf{x}^{(j)})^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{v}^T\mathbf{x}^{(i)} - \mathbf{v}^T(\frac{1}{N}\sum_{j=1}^{N}\mathbf{x}^{(j)}))^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{v}^T\mathbf{x}^{(i)} - \mathbf{v}^T\vec{0})^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{v}^T\mathbf{x}^{(i)})^2
\end{aligned}$$

What is the variance of the projections?

variance = ____________

1.6 Principal component of PCA

What is the first principal component of $X = [x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}]^T$?

2 K-means

2.1 Practical Example

Consider a dataset $\mathcal{D}$ with 5 points as shown below. Perform a K-means clustering on this dataset with $K = 2$ using the Euclidean distance as the distance function.

Remember that in the K-means algorithm, an iteration consists of performing following tasks: assigning each data point to it's nearest cluster center followed by recomputing those centers based on all the data points assigned to it.

Initially, the 2 cluster centers are chosen as $\mu_1 = [5.3, 3.5]^T$, $\mu_2 = [5.1, 4.2]^T$.

$$\mathcal{D} = \begin{bmatrix} 5.5 & 3.1 \\ 5.1 & 4.8 \\ 6.3 & 3.0 \\ 5.5 & 4.4 \\ 6.8 & 3.5 \end{bmatrix}$$

1. What will be the coordinates of the center for cluster 1 after the first iteration?
2. What will be the coordinates of the center for cluster 2 after the first iteration?
3. How many points will belong to cluster 1 after the first iteration?
4. How many points will belong to cluster 2 after the first iteration?
5. In general (not directly pertaining to this example), is it possible to have empty clusters, once the algorithm converges? Is K-means guaranteed to converge?

Recap: Lloyd's algorithm

1. Fixing the cluster centers, assign points to nearest clusters.
2. Given the point assignments, re-estimate cluster centers.
3. Termination: No points change clusters in next iteration.

Suppose we have the following data points where the x and o points are the initial cluster means (the cluster means are **not** data points). What clusters and cluster means will Lloyd's algorithm learn?