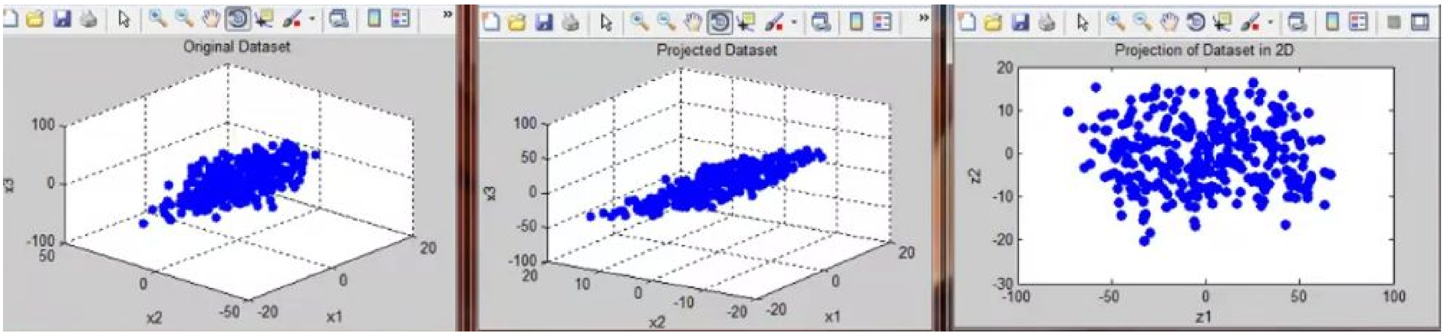
Plan

1. **Dimensional reduction - reduce the number of features to simplify data**
    a. **Improving performance**
    b. **Visualization**
2. **Principal Component Analysis - PCA algorithm overview**
    a. **Understanding PCA problem**
    b. **PCA Algorithm**
    c. **Comparing PCA with linear regression**
3. **Applying PCA - recommended ways to use PCA**
    a. **Reconstruction from compressed representation**
    b. **Choosing the number of principal components**
    c. **Advice for applying pca**



Reduce from n-dimension to k-dimension: Find k vectors $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ onto which to project the data, so as to minimize the projection error.

$$\texttt{Sigma} = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)})(x^{(i)})^T$$

```
[U,S,V] = svd(Sigma);
Ureduce = U(:,1:k);
z = Ureduce'*x;
```

Typically, choose k to be smallest value so that

$$\frac{\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)} - x_{approx}^{(i)}\|^2}{\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)}\|^2} \leq 0.01 \qquad (1\%)$$

"99% of variance is retained"

Application of PCA

- Compression
  - Reduce memory/disk needed to store data
  - Speed up learning algorithm

- Visualization

Questions:

1. **What is dimensional reduction?**
2. **What's the main reasons to use PCA?**
3. **Which parameter illustrates percentage of retained variance?**