

Dimension reduction

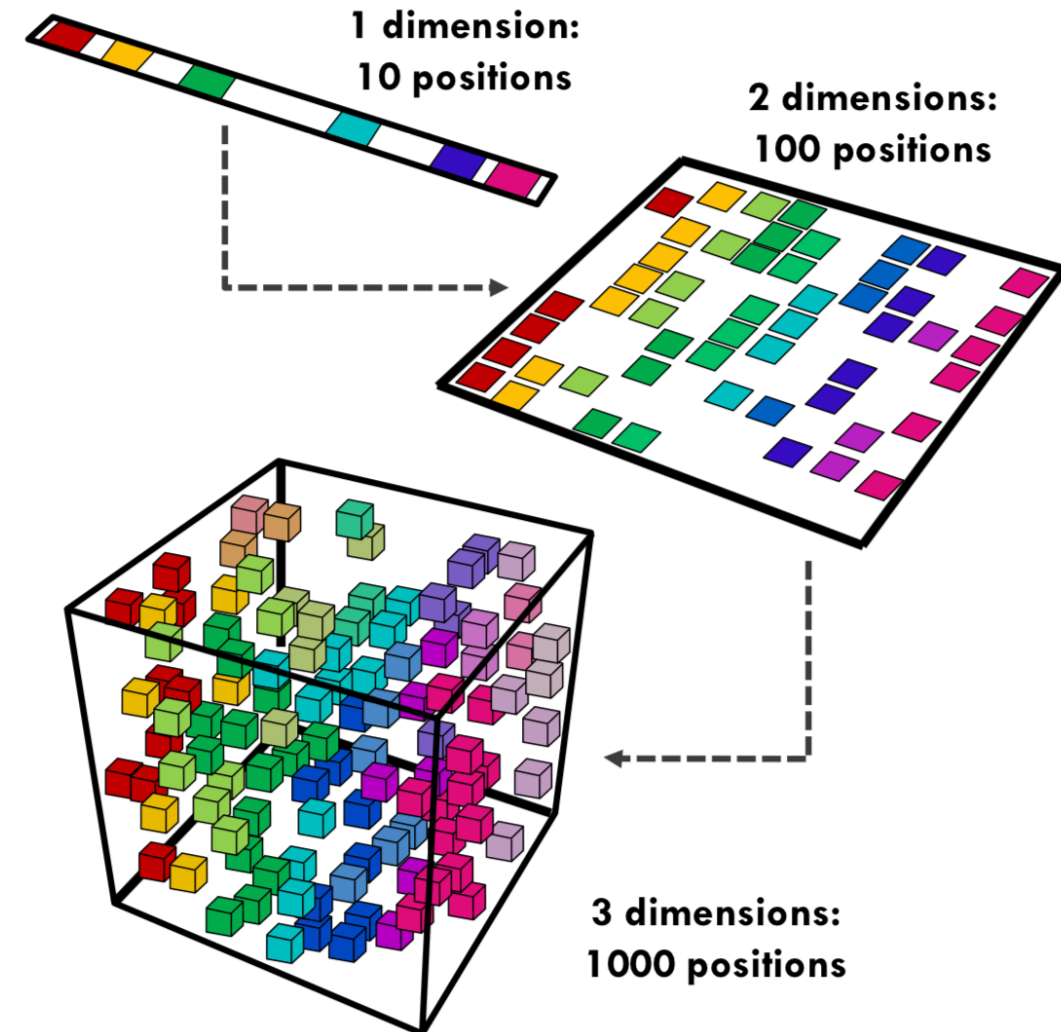
Kookjin Lee

(kookjin.Lee@asu.edu)

The contents of this course, including lectures and other instructional materials, are copyrighted materials. Students may not share outside the class, including uploading, selling or distributing course content or notes taken during the conduct of the course. Any recording of class sessions is authorized only for the use of students enrolled in this course during their enrollment in this course. Recordings and excerpts of recordings may not be distributed to others. (see ACD 304 – 06 , “ Commercial Note Taking Services ” and ABOR Policy 5 - 308 F.14 for more information).

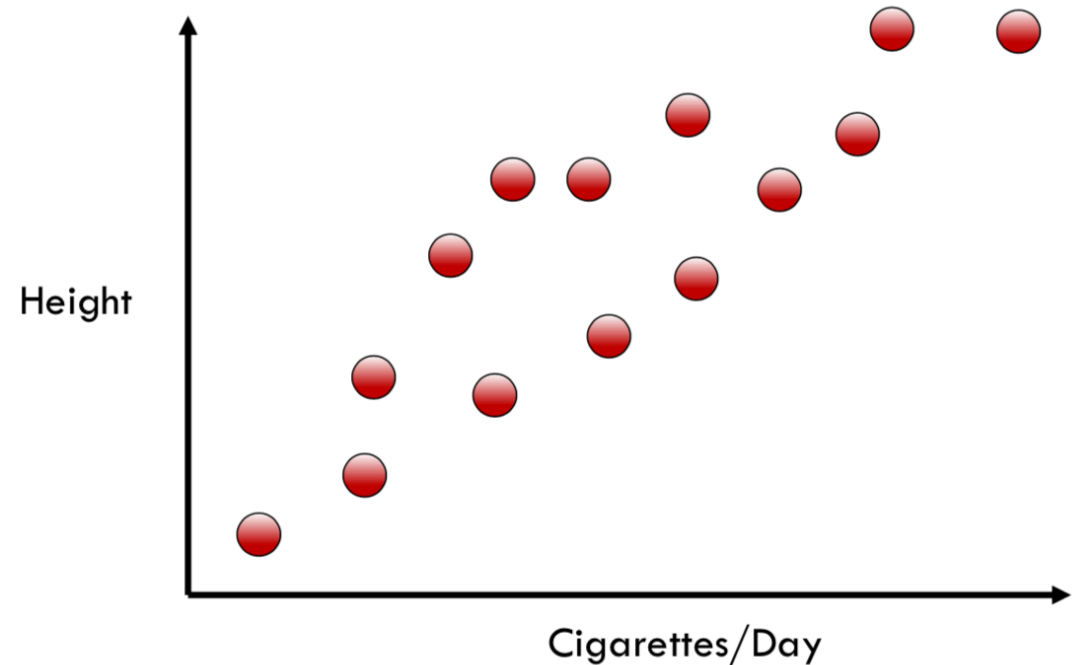
Curse of dimensionality

- Theoretically, increasing features should improve performance
- In practice, too many features leads to worse performance
- Number of training examples required increases exponentially with dimensionality



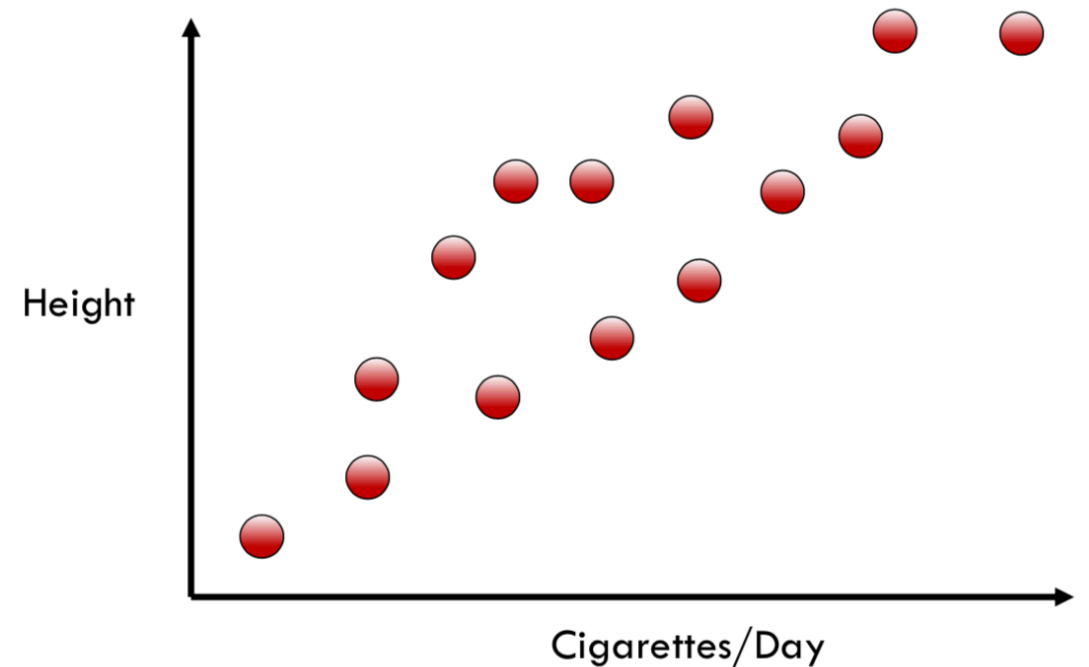
Solution: dimension reduction

- Data can be represented by fewer dimensions (features)
- Reduce dimensionality by selecting subset (feature elimination)
- Combine with linear and non-linear transformations



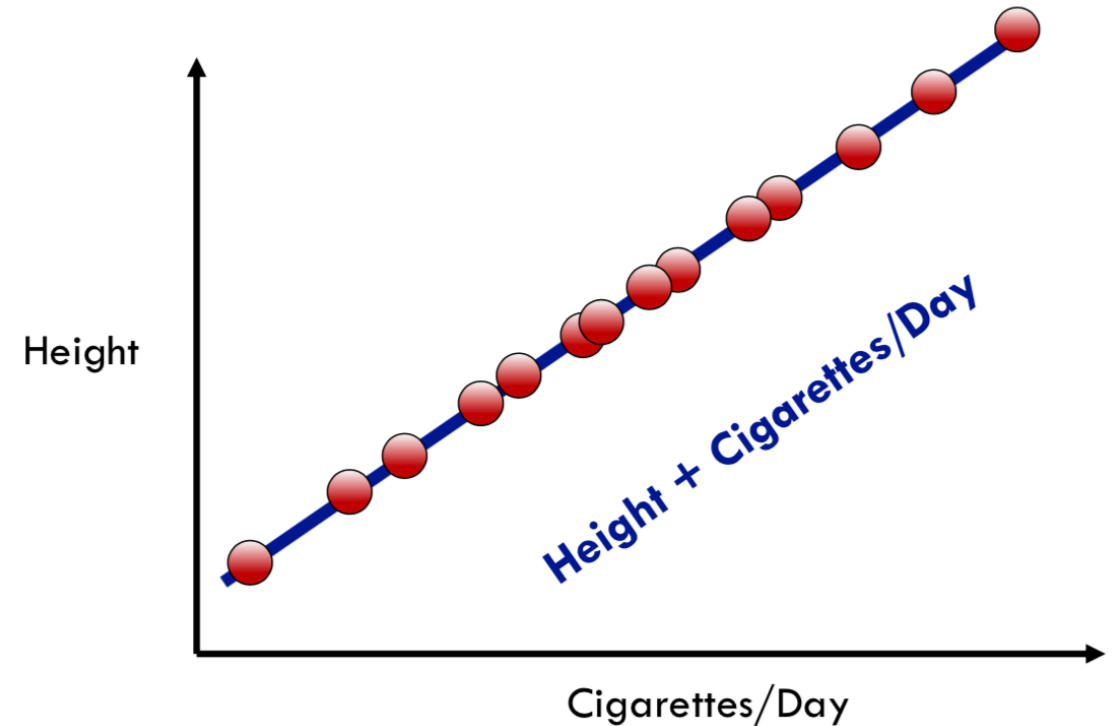
Solution: dimension reduction

- Two features: height and cigarettes per day
- Both features increase together (correlated)
- Can we reduce number of features to one?



Solution: dimension reduction

- Create single feature that is combination of height and cigarettes
- Principal Component Analysis (PCA)

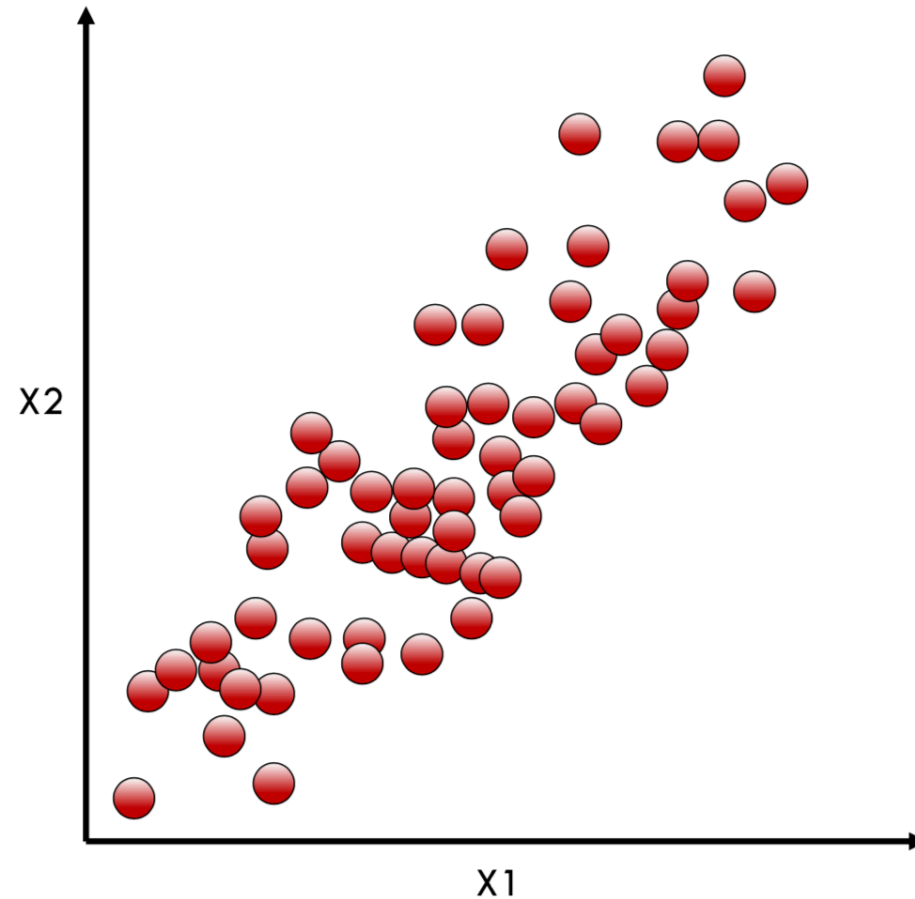


Dimensionality reduction

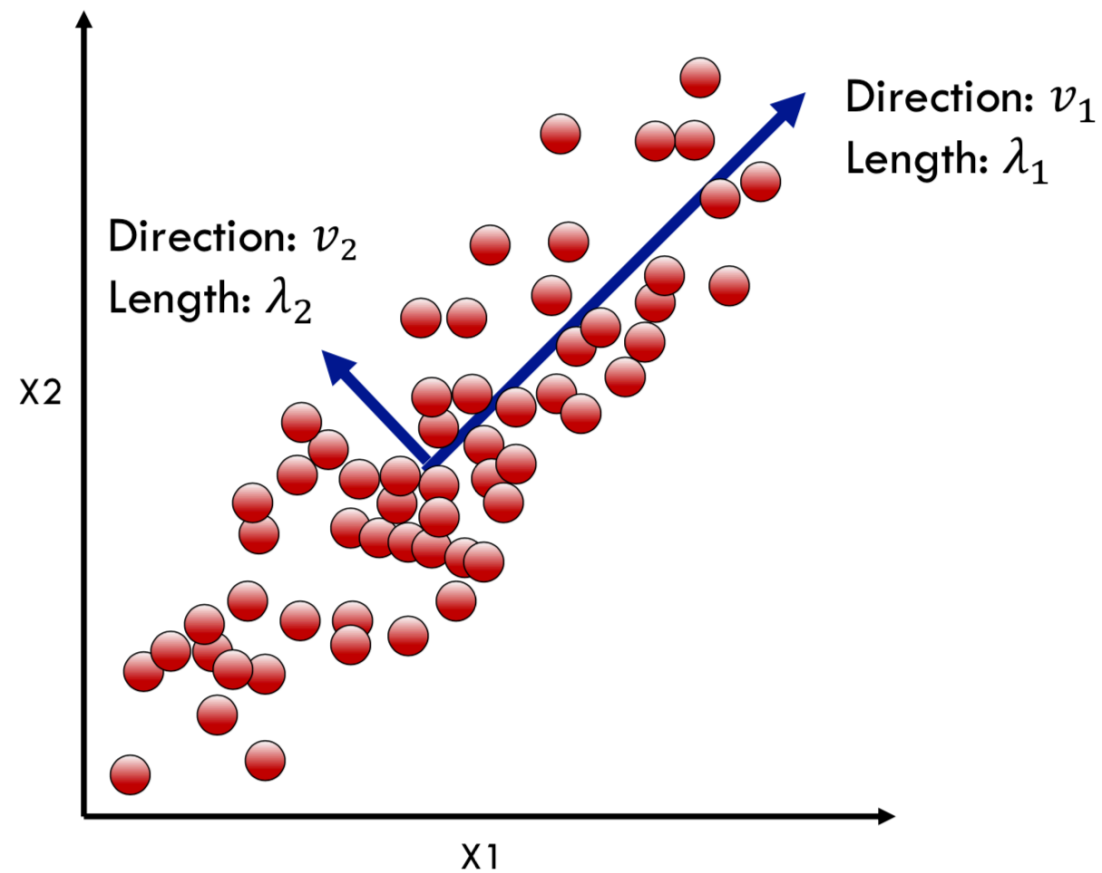
- Given an N-dimensional data set, find an N x k matrix:

$$y^{(i)} = U^T x^{(i)}$$

Principal Component Analysis (PCA)



Principal Component Analysis (PCA)



Mathematical derivation

- A simple example

Mathematical derivation

- Projection

$$uu^{\top}x^{(i)}$$

- Distance from origin to projected quantity

$$u^{\top}x^{(i)}$$

- Maximize the variance?

$$\frac{1}{n} \sum_{i=1}^n (u^{\top}x^{(i)})^2$$

Mathematical derivation

- Projection

$$uu^{\top}x^{(i)}$$

- Distance from origin to projected quantity

$$u^{\top}x^{(i)}$$

- Maximize the variance?

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n (u^{\top}x^{(i)})^2 &= \frac{1}{n} \sum_{i=1}^n u^{\top}x^{(i)}x^{(i)\top}u \\ &= u^{\top} \left(\frac{1}{n} \sum_{i=1}^n x^{(i)}x^{(i)\top} \right) u\end{aligned}$$

Mathematical derivation

- The principal component = principal eigenvector

$$u^{\top} \left(\frac{1}{n} \sum_{i=1}^n x(i) x(i)^{\top} \right) u = u^{\top} \Sigma u$$

$$\Sigma u = \lambda u$$

Singular Value Decomposition

- SVD is a numerical algorithm commonly used for PCA (e.g., scikit-learn)

$$\begin{bmatrix} \star & \star & \star \\ \star & \star & \star \\ \star & \star & \star \\ \star & \star & \star \\ \star & \star & \star \end{bmatrix} = \begin{bmatrix} \star & \star & \star & \star & \star \\ \star & \star & \star & \star & \star \\ \star & \star & \star & \star & \star \\ \star & \star & \star & \star & \star \\ \star & \star & \star & \star & \star \end{bmatrix} \begin{bmatrix} \star & 0 & 0 \\ 0 & \star & 0 \\ 0 & 0 & \star \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \star & \star & \star \\ \star & \star & \star \\ \star & \star & \star \end{bmatrix}$$

$A_{m \times n} \qquad U_{m \times m} \qquad S_{m \times n} \qquad V_{n \times n}^T$

Singular Value Decomposition

- Truncated SVD

$$\begin{array}{c}
 \begin{bmatrix} \star & \star & \star \\ \star & \star & \star \\ \star & \star & \star \\ \star & \star & \star \\ \star & \star & \star \end{bmatrix} \approx \begin{bmatrix} \star & \star & \star & \star & \star \\ \star & \star & \star & \star & \star \\ \star & \star & \star & \star & \star \\ \star & \star & \star & \star & \star \\ \star & \star & \star & \star & \star \end{bmatrix} \begin{bmatrix} \mathbf{9} & 0 & \mathbf{0} \\ 0 & \mathbf{7} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \star & \star & \star \\ \star & \star & \star \\ \star & \star & \star \\ \star & \star & \star \end{bmatrix} \\
 A_{m \times n} \qquad U_{m \times k} \qquad S_{k \times k} \qquad V_{k \times n}^T
 \end{array}$$

3D example

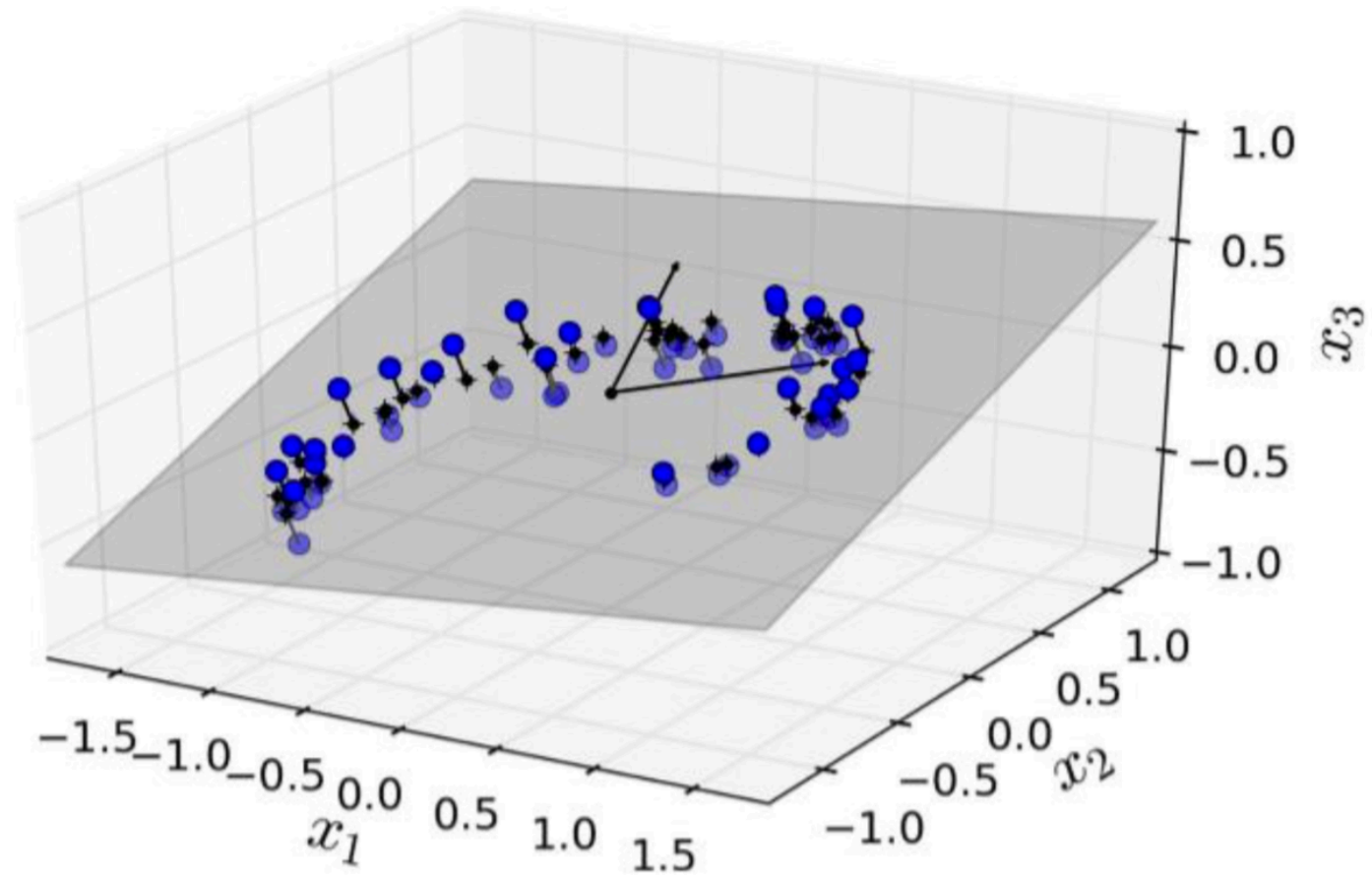


Figure 8-2. A 3D dataset lying close to a 2D subspace

3D example

- Projected data

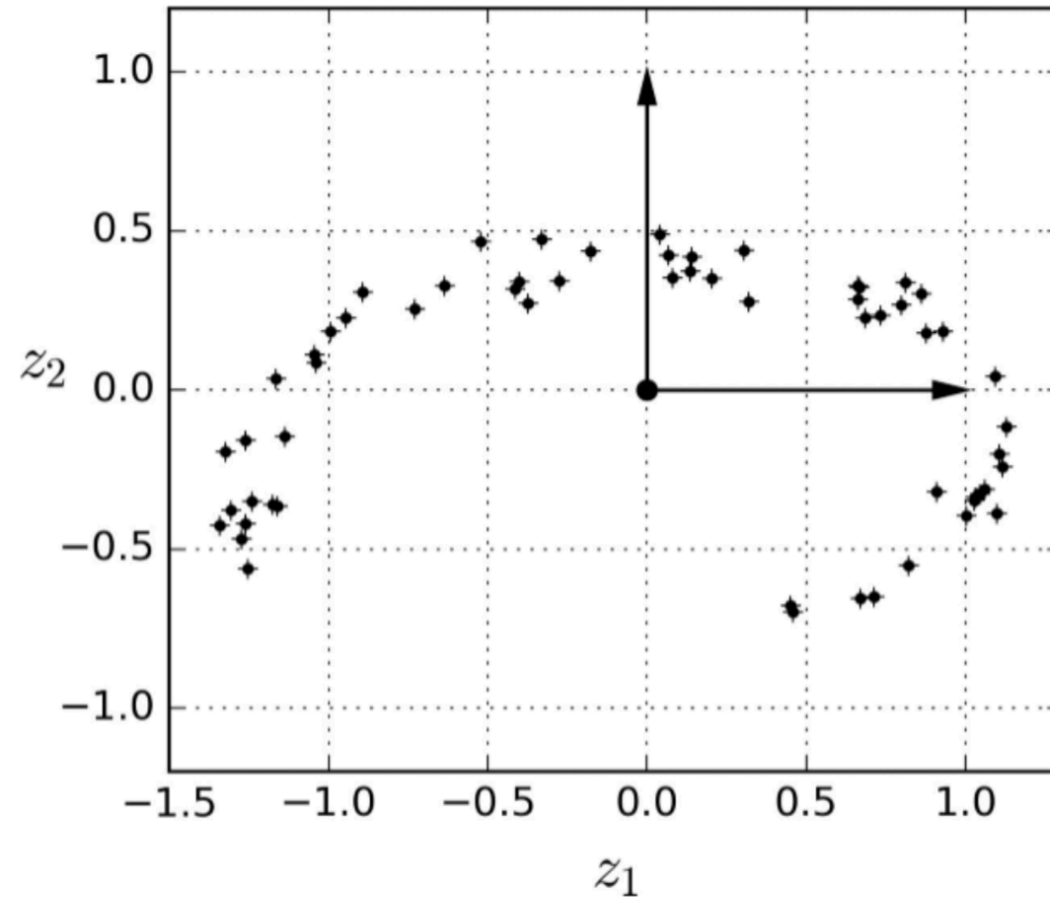
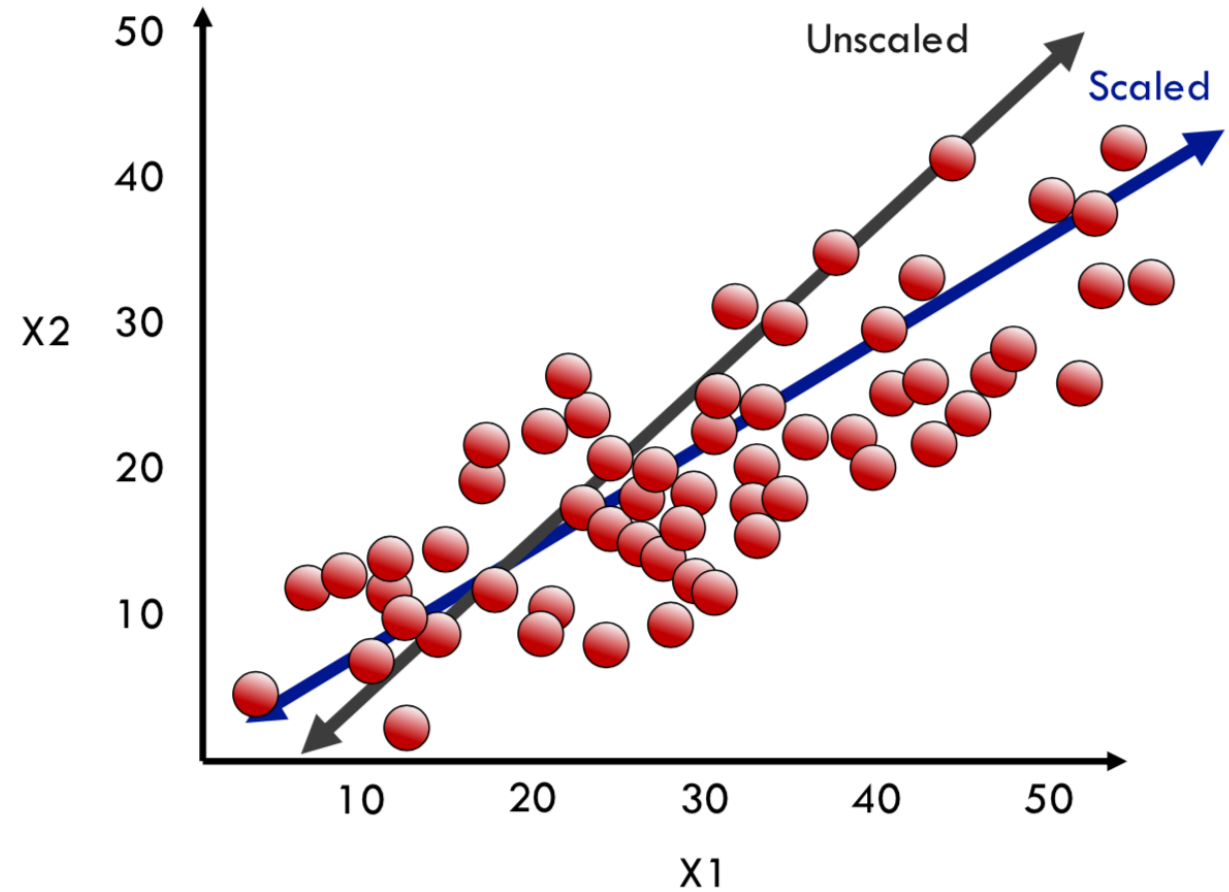
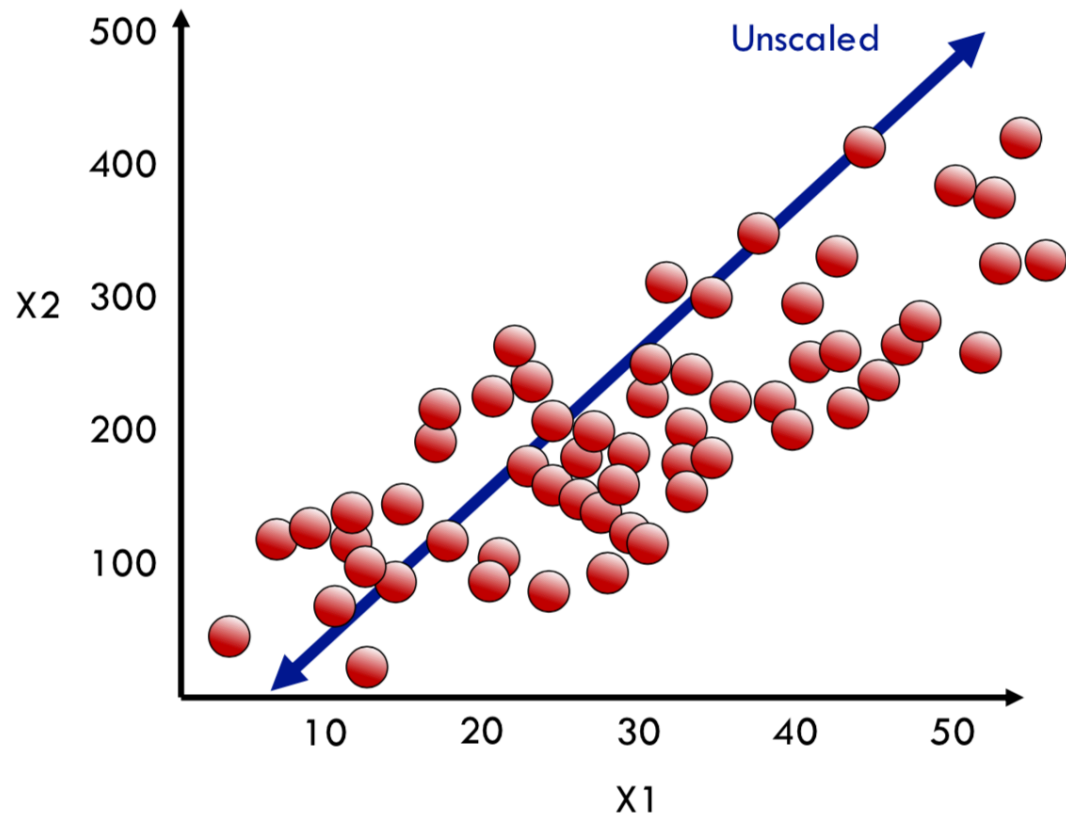


Figure 8-3. The new 2D dataset after projection

Importance of feature scaling

- PCA (and SVD) seek to find the vectors that capture the most variance
- Variance is sensitive to axis scale
- Must scale data

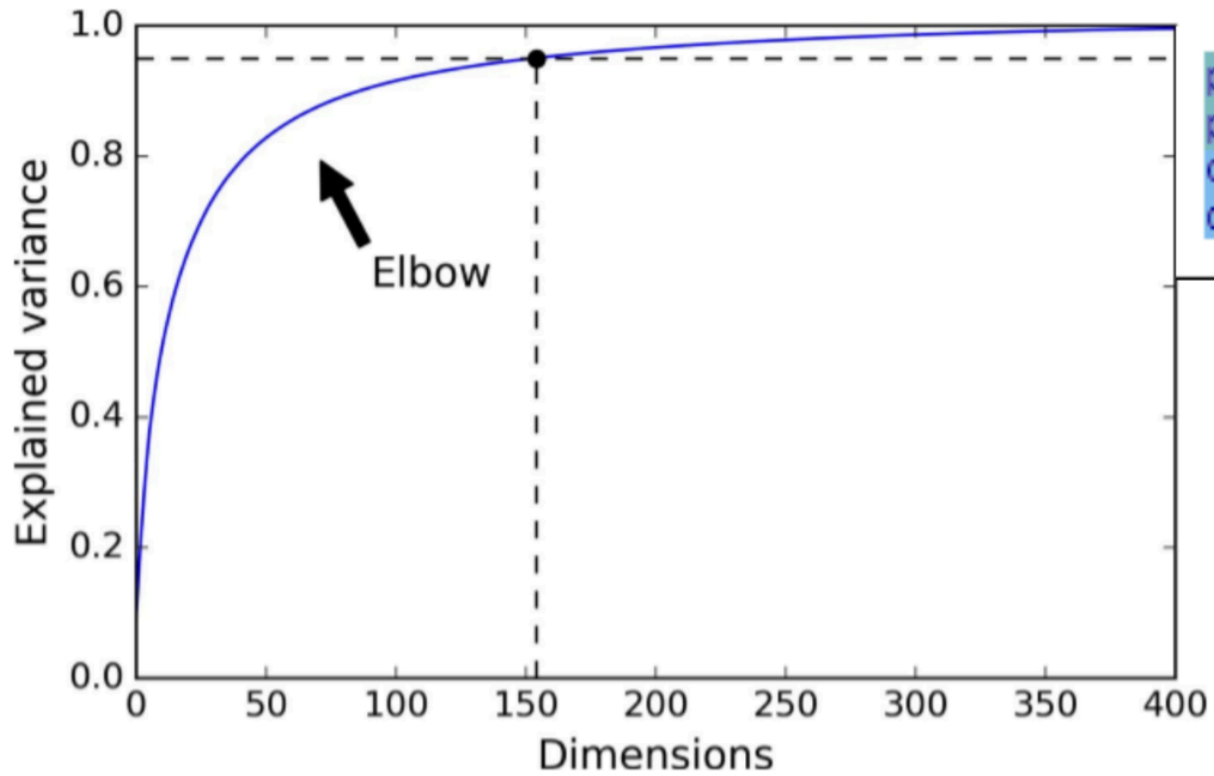
Importance of feature scaling



How to scale?

- To have zero mean and unit variance
 - Standard Scaler

Explained variance



```
pca = PCA()  
pca.fit(X)  
cumsum = np.cumsum(pca.explained_variance_ratio_)  
d = np.argmax(cumsum >= 0.95) + 1
```