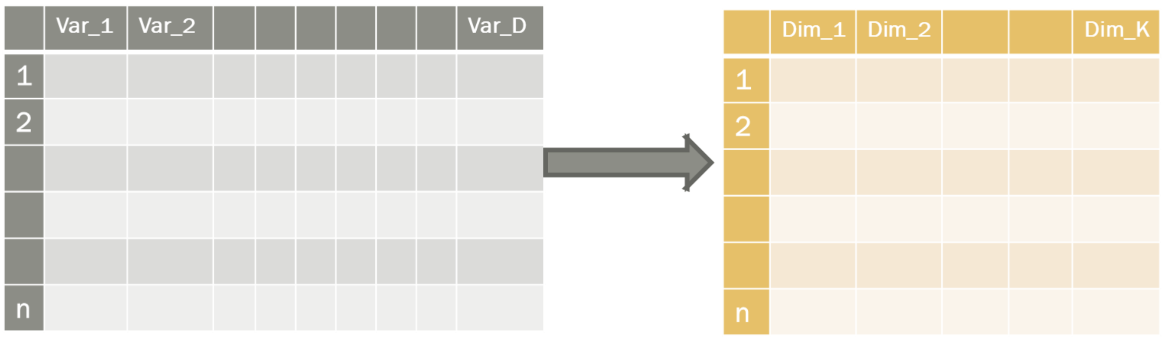
Dimensionality Reduction

From a theoretical point of view, increasing the number of features should lead to better performance. In practice, the inclusion of more features leads to worse performance (i.e., curse of dimensionality). The number of training examples required increases exponentially with dimensionality.

Significant improvements can be achieved by first mapping the data into a lower-dimensional space.

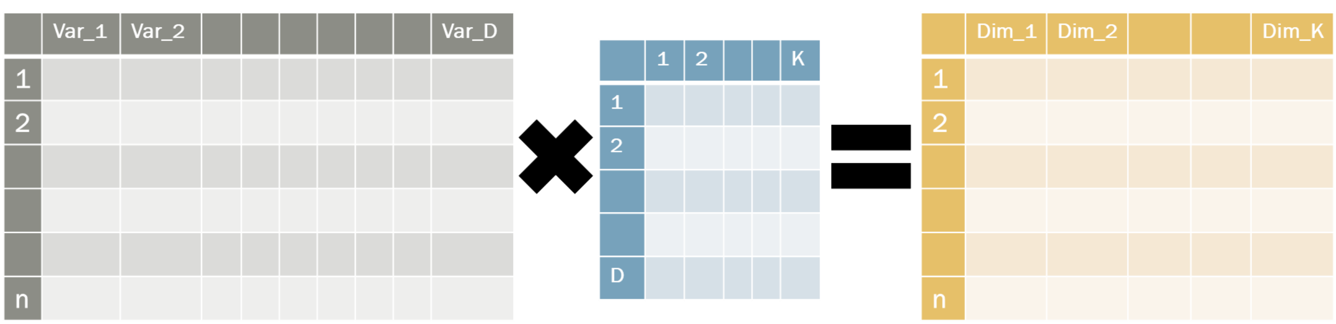


K << D

Dimensionality can be reduced by:

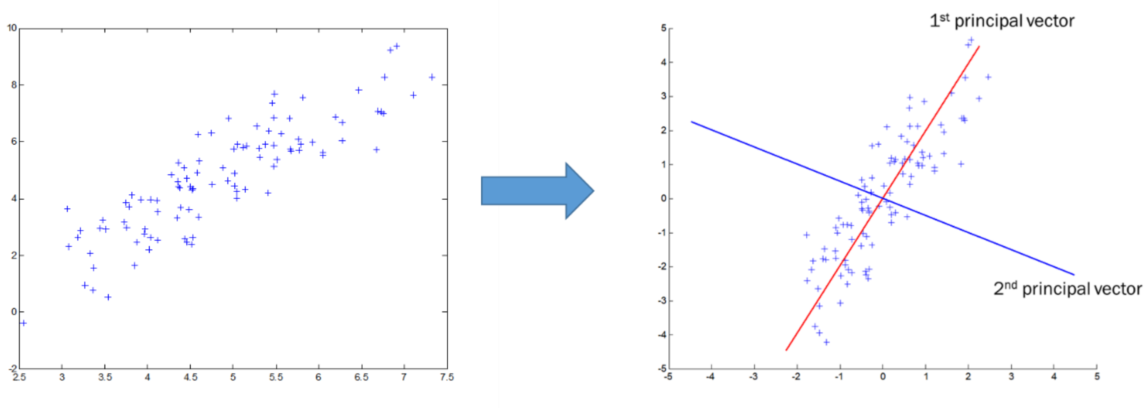
- Combining features using a linear or non-linear transformatio
- Selecting a subset of features (i.e., feature selection).

Linear combinations are particularly attractive because they are simple to compute and analytically tractable.



Principal Component Analysis (PCA)

It is an unsupervised dimensionality reduction technique which reduces the dimensionality of a data set by finding a new set of variables, smaller than the original set of variables by retaining most of the sample's information.



Algorithm

1. *X ← Prepare N x D data*
2. *X subtract mean x from each row in X => N x D*
3. *Σ ← covariance matrix of X => D x D*
4. *Find eigenvectors and eigenvalues of Σ => D vectors*
5. *Coeff ← the K eigenvectors with largest eigenvalues => D x K*
6. *Matrix multiplication of X (N x D) with Coeff (D x K) to obtain PCs (N x K)*