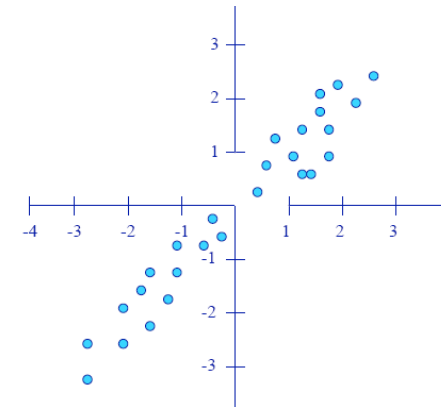


Principal Component Analysis (PCA)

- PCA is a way of expressing a high dimensional data set via an alternative set of dimensions that are
 - Orthogonal to each other
 - Easy to rank by how “representative” they are
 - Not always easy to interpret
- It is useful because it allows visualization of the data in the most representative dimensions
- It may (or not) provide useful predictions for clustering and phenotype prediction

Principal Component Analysis (PCA)

- Identify direction of greatest variation.
- Dimensions can be either genes or samples



Principal Component Analysis (PCA)

- Rotate data so that the new coordinates are based on correlation structure of the data
- Select a subset of the new coordinates with high variability, and use for data visualization, summarization and clustering
 - e.g., clustering tumor samples using PCA on expression profiles

Principal Component Analysis (PCA)

- Re-express data using coordinates based on principal components (PCs)
- Dimension reduction is a result of ignoring anything beyond the first k PCs

