Outline

1. Dimensionality reduction
   - Curve of dimensionality

2. Principal component analysis
**Dimensionality reduction**

**Motivation** Given data that are high-dimensional $\mathbf{x} \in \mathbb{R}^D$, we want to find a low-dimensional representation $\mathbf{y} \in \mathbb{R}^M$ such that $M < D$:

- Visualize data and discover intrinsic structures
- Save computational and storage cost
- Robust statistical modeling: curse of dimensionality
Curse of dimensionality

**Intuition** The higher the dimensionality, the more data points we need to train a model.

- To fill a unit-cube in $\mathbb{R}^D$ uniformly with data points, we need $r^D$ where $r$ is the edge length of small cells (i.e., dividing each axis in equal size of $r$.)
  Thus, if data is distributed that way, models such as decision trees need $r^D$ training samples in order to make sure every cell is covered — in case a test sample falls into one of those cells.

- For a unit-ball $x \leq 1$, a large percentage of data live in the shell — between the surface $x = 1$ and the surface $x = 1 - \epsilon$. The percentage is
  
  $$1 - (1 - \epsilon)^D$$

  approaches 1. Thus, most data points in the high-dimensional space are crowded in the shell and are about the same distance from each other.
Curse of dimensionality: more examples

Please check the Wikipedia entry:
Dimensionality reduction

- Linear: the low-dimensional coordinates $y$ is parameterized as
  \[ y = U^T x, \]
  where $U \in \mathbb{R}^{M \times D}$

- Nonlinear: the relationship is through a nonlinear mapping
  \[ y = f(x) \]

In both categories, there are many methods. We will focus on Principal Component Analysis (PCA) — a linear method for dimensionality reduction.
Outline

1. Dimensionality reduction

2. Principal component analysis
**Intuition** Consider the special case $M = 1$, namely, we are transforming $x$ into a scalar via

$$y = u^T x$$

which $u$ is sensible?
Derivation of the first principal components

Please check the hand-written note (PCA.pdf)