dimension-reduction

Shumin An

April 2019

1 Dimension Curse

To get a better understanding of dimension, we may think it in a different perspective. We know that for a n-dimension spherome, its volume is CR^n. And for a hypercube which has 2R of each side is 2^nR^n. So the ratio is:

$$\lim_{n\to 0}\frac{CR^n}{2^nR^n}=0$$

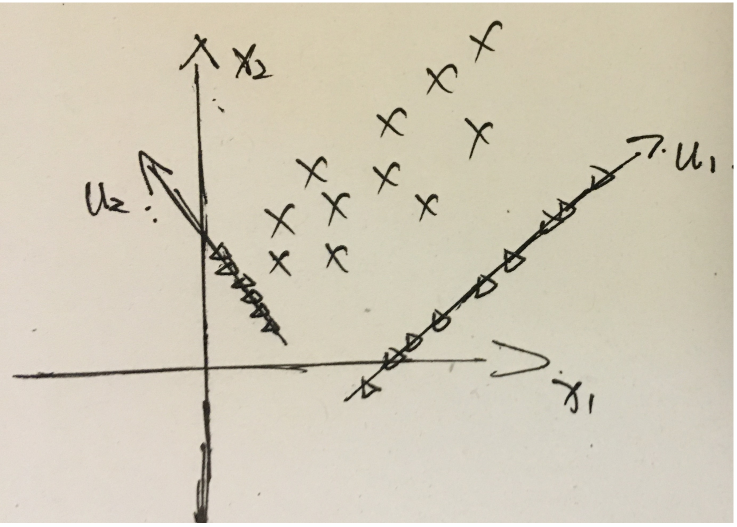
If we have the high-dimensional data, most of the sample will be at the border of the hypercube, which caused difficulty to do classification.

2 PCA

For convenience, we turn the covariance into matrix form:

$$\begin{aligned}
S &= \frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})(x_i-\overline{x})^T \\
&= \frac{1}{N}(x_1-\overline{x},x_2-\overline{x},\cdots,x_N-\overline{x})(x_1-\overline{x},x_2-\overline{x},\cdots,x_N-\overline{x})^T \\
&= \frac{1}{N}(X^T-\frac{1}{N}X^T\mathbb{I}_{N*1}\mathbb{I}_{N*1}^T)(X^T-\frac{1}{N}X^T\mathbb{I}_{N*1}\mathbb{I}_{N*1}^T)^T \\
&= \frac{1}{N}X^T(E_N-\frac{1}{N}\mathbb{I}_{N*1}\mathbb{I}_{1*N})(E_N-\frac{1}{N}\mathbb{I}_{N*1}\mathbb{I}_{1*N})^TX \\
&= \frac{1}{N}X^TH_NH_N^TX \\
&= \frac{1}{N}X^TH_NH_NX=\frac{1}{N}X^THX \qquad (1)
\end{aligned}$$

There's a example that I want to use to explain PCA:



Suppose there's a 2-dimensional feature space, x1 and x2 are features. The label 'x' represents the sample data. PCA's goal is to find a vector μ that when sample data projected on that vector, it has the maximum variance(covered by sample data as long as possible). In general, the variance we mentioned above could be written as:

$$((x_i - \overline{x})^T u_j)^2$$

And the object function:

$$\begin{aligned} J &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{q}((x_i - \overline{x})^T u_j)^2 \\ &= \sum_{j=1}^{q} u_j^T S u_j \ , \ s.t. \ u_j^T u_j = 1 \end{aligned} \tag{2}$$

We solve it with Lagrange Multiplier:

$$\underset{u_j}{argmax}\, L(u_j, \lambda) = \underset{u_j}{argmax}\, u_j^T S u_j + \lambda(1 - u_j^T u_j)$$

So:

$$Su_j = \lambda u_j$$