THU-70250043, Pattern Recognition (Spring 2018) **Homework: 8**

PCA and Non-linear dimensionality reduction

Lecturer: Changshui Zhang `zcs@mail.tsinghua.edu.cn`

Student: XXX `xxx@mails.tsinghua.edu.cn`

Problem 1

Maximum-variance and Minimum-error approaches to PCA

You have gone through the K-L and PCA algorithms in the class, but don't be confused by the names of this algorithm: Principal component analysis(PCA) is also known as K-L transform, they are completely the same thing.

PCA can be defined as the orthogonal projection of the data onto a lower dimensional linear space, known as the principal subspace, such that *the variance of the projected data is maximized*; It can also be defined as the linear projection that *minimizes the average projection cost*, defined as the mean squared distance between the data points and their projections. We'll consider both approaches in this problem.

Suppose we have a data set of observations $\{\mathbf{x}_n\}$, where $n = 1, ..., N$, $\mathbf{x}_n \in \mathcal{R}^D$. Our goal is to project the data onto a space having dimensionality $M < D$.

Maximum-variance approach

1.1 To begin with, let's consider the projection onto a one-dimensional space($M = 1$). Define the direction of this space using a D-dimensional vector $\mathbf{u}_1$. Prove that the variance of the projected data is given by the following expression: $\mathbf{u}_1^T S \mathbf{u}_1$, where S is called the data covariance matrix.

1.2 We now maximize the projected variance $\mathbf{u}_1^T S \mathbf{u}_1$ with respect to $\mathbf{u}_1$. Clearly, this has to be a constrained maximization to prevent $||\mathbf{u}_1|| \to \infty$, the appropriate constraint comes from the normalization condition $\mathbf{u}_1^T \mathbf{u}_1 = 1$. We can introduce a Lagrange multiplier denoted as λ_1 and solve this maximization problem. Show the details of solving $\mathbf{u}_1$ and the maximum variance.

1.3 For cases $M \geq 2$, we use proof by induction to show the same principle. Suppose the result above holds for some general value of M, show that it consequently holds for dimensionality $M + 1$.

Hint: To do this, remember that we've already selected the M eigen-vectors corresponding to the M largest eigen-values of the data covariance matrix S. We're now tying to maximize the variance on direction $\mathbf{u}_{M+1}$. The maximization should be done subject to the constraints that $\mathbf{u}_{M+1}$ be orthogonal to the existing vectors $\mathbf{u}_1, ..., \mathbf{u}_M$, and also that it be normalized to unit length.

Minimum-error approach

Suppose we have a *complete orthonormal set* of D-dimensional basis vector $\{\mathbf{u}_i\}$ where $i = 1, ..., D$ that satisfy:

$$\mathbf{u}_i^T \mathbf{u}_j = \delta_{ij} \tag{1}$$

Since this basis is *complete*, each data point can be presented as follows:

$$\mathbf{x}_n = \sum_{i=1}^{D} (\mathbf{x}_n^T \mathbf{u}_i) \mathbf{u}_i \tag{2}$$

Our goal is to approximate this data point using a restricted number $M < D$ of variables corresponding to a projection onto a lower-dimensional subspace. Let's make use of the following expression:

$$\tilde{\mathbf{x}}_n = \sum_{i=1}^{M} z_{ni} \mathbf{u}_i + \sum_{i=M+1}^{D} b_i \mathbf{u}_i \tag{3}$$

Where the $\{z_{ni}\}$ depends on the particular data point, whereas the $\{b_i\}$ are constants same for all data points.

Just as the subtitle, we shall minimize the squared distance between the original data point $\mathbf{x}_n$ and its approximation $\tilde{\mathbf{x}}_n$:

$$J = \frac{1}{N} \sum_{n=1}^{N} ||\mathbf{x}_n - \tilde{\mathbf{x}}_n||_2^2 \tag{4}$$

1.4 Prove that the optimal $\{z_{ni}\}$ complies with:

$$z_{ni} = \mathbf{x}_n^T \mathbf{u}_i \tag{5}$$

And the optimal $\{b_i\}$ complies with:

$$b_i = \bar{\mathbf{x}}^T \mathbf{u}_j \tag{6}$$

Where $\bar{\mathbf{x}} = (1/N) \sum_{n=1}^{N} \mathbf{x}_n$.

Problem 2

考虑一个三层 d-k-d(k<d) 神经元实现的一个线性神经网络（无激活函数）。证明当训练自联想模式时（标签即为输入），得到的就是 PCA。

Problem 3

ISOMAP, LLE 对流形的降维

考虑如下的问题并实现 ISOMAP，LLE 等降维方法：

3.1 在三维空间中产生 “N” 形状的流形，使用 ISOMAP 方法降维并作图，给出数据的三维分布图和最佳参数下的降维效果图。

3.2 在三维空间中产生 “3” 形状的流形，使用 LLE 方法降维并作图，给出数据的三维分布图和最佳参数下的降维效果图。

注意：数据在产生过程中可不必严格保证形状，大致符合要求即可。不用在数据的产生上花费过多时间。