CS420 Machine Learning Homework 2

Xiaoyu Shi, 516030910567

1 PCA algorithm

PCA:

Data: $X = \{x_1, x_2, \ldots, x_N\}$, $x \in \mathbf{R}^{n\times 1}$
for i *from* 1 *to* n **do**
    Zero-normalize(X);
end
$C = \frac{1}{N} XX^T$;
$evalue = eigenvalues(C)$;
$evector = eigenvectors(C)$;
sort(evalue);
#descending order
Find the corresponding evector e of evalue[1];
$output = e^T X$

Algorithm 1: PCA algorithm

SVD:

Data: $X = \{x_1, x_2, \ldots, x_N\}$, $x \in \mathbf{R}^{n\times 1}$
$X = U \sum V^T$;
$\sum V^T = U^T X$;
$output = \sum V^T[1]$

Algorithm 2: SVD algorithm

Pros and cons:

- PCA have to calculate covariance matrix, but SVD does not have to, so when the dataset is huge, the amount of calculation for PCA can be large, and svd iterative calculation can be faster.
- For sparse matrix, SVD is more suitable for them.
- PCA can only calculate the principal component on one direction, while SVD can get that on another direction.
- PCA is simpler and easier to complete.

2 Factor Analysis

According to Bayes rule, we can get $f_{Y|X}(y|x) = \frac{G(x|Ay+\mu,\sum_e)G(y|0,\sum_y)}{p(Ay+\mu+e)}$.

Since $X+Y\ N(E[x]+E[Y], var(X)+var(Y)+2conv(X,Y))$, we can get $f_{Y|X}(y|x) = \frac{G(x|Ay+\mu,\sum_e)G(y|0,\sum_y)}{G(x|\mu+\mu_e,A\sum_y A^T+\sum_e)}$.

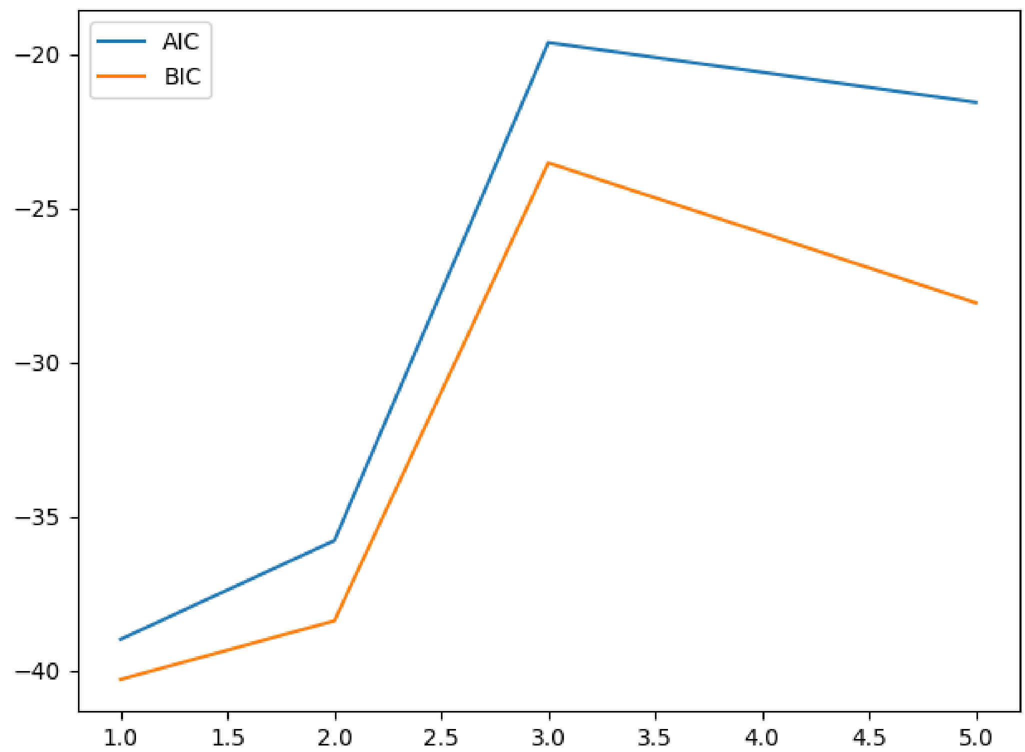
Usually, μ_e is considered as 0.

3 Independent Component Analysis

The central limit theorem shows that for mixed signals, the probability density is close to the Gaussian distribution of the probability distribution of any one source signal; in turn, the non-Gaussian of the maximized signal is consistent with the statistical independence of the maximized signal. ICA is a method of blind source seperation task, so non-Guassianity is used as a principle for ICA estimation.

4 Dimensionality Reduction by FA

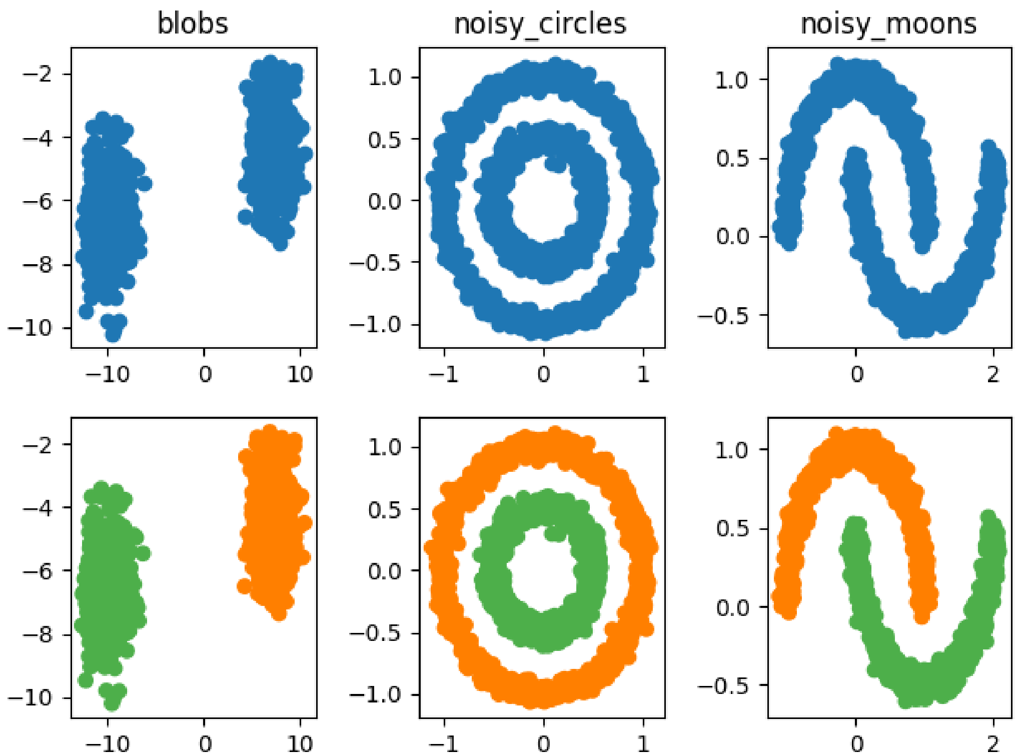
We set $N = 100, dim(y) = m = 3, \mu = 0, \sigma^2 = 0.1, n = 10$ and randomly generate the dataset. Then we adopt the FA model and set the M to 5. The result is as follows.



Both BIC and AIC gives the hightest score when $M = 3$.

5 Spectral Clustering

It works when the data is distributed as blobs, circles or moons. Some instances of them are showed as follows.



the main reason leading to the failure of the spectral clustering on noisy data is that the block structure of the affinity matrix is destroyed by noise. It can be looked like figure 5.