Report on Homework 2

CS420, Machine Learning, Shikui Tu, Summer 2018

Zelin Ye 515030910468

1 PCA algorithm

When it comes to principal component, the most common used approach is principal component analysis (PCA) algorithm. Thus, I can use the original PCA algorithm to extract the first principal component of the dataset. The computational details can refer to Alg. 1.

Algorithm 1: Original PCA

Input : The dataset X, a $n \times N$ matrix
Output: The first principal component w
Conduct normalization for X, and make sure the mean of X is 0;
Find the covariance matrix of X, denoted by C:

$$C = XX^T;$$

Calculate the eigenvalues λ and eigenvectors V of X;
Choose the maximal eigenvalue λ_m and corresponding eigenvector v_m;
Calculate the first principal component:

$$w = v_m^T X;$$

return w;

The above algorithm (PCA) is a classical and commonly used method to solve principal components, which has the following characteristics.

Pros:

1. PCA has simple logic and is easy to implement;
2. The orthogonality among principal components chosen by PCA can eliminate the interaction between original data components;
3. PCA belongs to unsupervised learning, and it is not restricted by sample labels.

Cons:

1. PCA treats all samples, namely the collection of eigenvectors, as a whole and neglects the category attributes. Nevertheless, the projection direction it neglects might contain some important separability information;
2. PCA would be time-consuming when facing large amount of data;
3. The actual meanings of principal components extracted by PCA are a little bit ad-hoc and hard to explain.

The original PCA might have good performance when handling linear data, but it would encounter difficulties under non-linear data. In non-linear cases, a variant of PCA called KPCA (kernel principal components analysis) shows its strength.

The innovation of KPCA is that it introduces a non-linear mapping function, mapping the data from original space to high-dimensional space. The details of KPCA can refer to Alg. 2.

Algorithm 2: KPCA

Input : The dataset X, a $n \times N$ matrix
Output: The first principal component w
Conduct normalization for X, and make sure the mean of X is 0;
Find the covariance matrix of X, denoted by C:

$$C = XX^T;$$

Calculate the eigenvalues λ and eigenvectors V of X;
Choose the maximal eigenvalue λ_m and corresponding eigenvector v_m;
Calculate the first principal component:

$$w = v_m^T X;$$

return w;

2 Factor Analysis (FA)

$$\begin{aligned} p(y|x) &= \frac{p(x|y)p(y)}{p(x)} \\ &= \frac{G(x|Ay + \mu, \Sigma_e)G(y|0, \Sigma_y)}{p(x)} \\ &= \frac{G(x|Ay + \mu, \Sigma_e)G(y|0, \Sigma_y)}{G(x|\mu + \mu_e, AA^T\Sigma_y + \Sigma_e)} \end{aligned}$$

where μ_e denotes the mean value of e, generally considered to be 0.

3 Independent Component Analysis (ICA)

ICA aims to decompose the source signal into independent parts. If the source signals are non-Gaussian, the decomposition is unique, or there would be a variety of such decompositions.

Suppose the source signal s consists of two components, conforming to multi-valued normal distribution, namely $s \sim N(0, I)$. Obviously, the probablity density funciton of s is centered on the mean 0, and the projection plane is an ellipse.

Meanwhile, we have $x = As$, where x denotes the actual signals received while A represents a mixing matrix. Then x is also Gaussian, with a mean of 0 and a covariance of $E[xx^T] = E[Ass^TA^T] = AA^T$.

Let C be a orthogonal matrix, and $A^{'} = AR$. If A is replaced by A', then we can get $x^{'} = A^{'}s$. It is easy to find that $x^{'}$ also conforms to normal distribution, with a mean of 0 and a covariance of $E[x^{'}(x^{'})^T] = E[A^{'}ss^T(A^{'})^T] = E[ACss^T(AC)^T] = ACC^TA^T = AA^T$.

Apparently, x and $x^{'}$ conform to the same distribution with different mixing matrices. Then we cannot determine the mixing matrix or the source signals from the received signals. Nevertheless, if x is non-Gaussian (e.g. Uniform Distribution), such case would be effectively avoided. Therefore, maximizing non-Gaussianity should be used as a principle for ICA estimation.

4 Causal discovery algorithms

pass

5 Causal tree reconstruction

pass