Report on Homework 2

CS420, Machine Learning, Shikui Tu, Summer 2018

Zelin Ye 515030910468

1 PCA algorithm

Algorithm 1: A variant of k-means

Input : The number of clusters K

Output: $\pi_k, \mu_k, \Sigma_k, (k = 1, 2, ..., K)$

Initialize the means μ_k, covariances Σ_k, mixing coefficients π_k and threshold $Thres$;

Evaluating the initial value of the log likelihood;

while *the convergence criterion of parameters or log likelihood is not satisfied* **do**

E step. Evaluate the responsibilities with the current parameter values:

$$\omega \leftarrow \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum\limits_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)} \quad , \quad \gamma(z_{nk}) \leftarrow \begin{cases} \omega & \omega > Thres \\ 0 & \omega \leq Thres \end{cases}$$

$$z_n \leftarrow \frac{e^{z_{n_i}}}{\sum\limits_{j=1}^{K} e^{z_{n_j}}}$$

M step. Re-estimate the parameters with the current responsibilities:

$$N_k \leftarrow \sum_{n=1}^{N} \gamma(z_{nk}) \quad , \quad \mu_k^{new} \leftarrow \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) x_n$$

$$\Sigma_k^{new} \leftarrow \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k^{new})(x_n - \mu_k^{new})^T$$

$$\pi_k^{new} \leftarrow \frac{N_k}{N}$$

Evaluate the log likelihood:

$$\ln p(X|\mu, \Sigma, \pi) \leftarrow \sum_{k=1}^{K} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \right\}.$$

return $\pi_k, \mu_k, \Sigma_k, (k = 1, 2, ..., K)$;

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