EE E6887 (Statistical Pattern Recognition)
Solutions for homework 8

P.1 Use the fact that the sum of samples from two Gaussian is again a Gaussian to show why independent component analysis can not isolate sources perfectly if two or more components are Gaussian.

Answer:

Assume $\mathbf{x}$ is a vector of independent Gaussian random variables. A linear mixture of $\mathbf{x}$ gives $\mathbf{y}$:

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

where $\mathbf{y}$ is also a vector of independent Gaussian random variables. The solution of ICA decomposition gives

$$\widehat{\mathbf{x}} = \mathbf{B}\mathbf{y}$$

subject to the condition that $E[\widehat{\mathbf{x}}\widehat{\mathbf{x}}^T] = I$. The condition is due to the fact that $\widehat{\mathbf{x}}$ is supposed to a vector of independent Gaussian random variables. Independent Gaussian random variables have zero cross-correlation (i.e. the off-diagonal components of $E[\widehat{\mathbf{x}}\widehat{\mathbf{x}}^T]$ is zero). The equal diagonal components of $E[\widehat{\mathbf{x}}\widehat{\mathbf{x}}^T]$ is due to the fact that the scale of the random variables cannot be recovered and hence their variance are set equal.

In this case, we would see that $\widehat{\mathbf{z}} = \mathbf{R}\widehat{\mathbf{x}}$ are also an ICA solution as $\widehat{\mathbf{z}}$ also satisfies the condition:

$$E[\widehat{\mathbf{z}}\widehat{\mathbf{z}}^T] = E[\mathbf{R}\widehat{\mathbf{x}}\widehat{\mathbf{x}}^T\mathbf{R}^T] = \mathbf{R}E[\widehat{\mathbf{x}}\widehat{\mathbf{x}}^T]\mathbf{R}^T = I$$

where $\mathbf{R}$ is the rotation matrix which has the property that $\mathbf{R}\mathbf{R}^T = \mathbf{R}\mathbf{R}^{-1} = I$

This means the ICA solution in this case has a rotation ambiguity.

P.2 Let cluster $\mathcal{D}_i$ contain n_i samples, and let d_{ij} be some measure of the distance between two cluster $\mathcal{D}_i$ and $\mathcal{D}_j$. In general, one might expect that if $\mathcal{D}_i$ and $\mathcal{D}_j$ are merged to form a new cluster $\mathcal{D}_k$, then the distance from $\mathcal{D}_k$ to some other cluster $\mathcal{D}_h$ is not simply related to d_{hi} and d_{hj}. However, consider the equation

$$d_{hk} = \alpha_i d_{hi} + \alpha_j d_{hj} + \beta d_{ij} + \gamma |d_{hi} - d_{hj}|$$

Show that the following choices for the coefficients α_i, α_j, β and γ lead to the distance functions indicated.

(a) d_{min}: $\alpha_i = \alpha_j = 0.5$, $\beta = 0$, $\gamma = -0.5$

(b) d_{max}: $\alpha_i = \alpha_j = 0.5$, $\beta = 0$, $\gamma = 0.5$

Answer:

(a) With the parameter setting, we have

$$d_{hk} = 0.5 d_{hi} + 0.5 d_{hj} - 0.5 |d_{hi} - d_{hj}|$$

If $d_{hi} \leq h_{hj}$,

$$d_{hk} = 0.5 d_{hi} + 0.5 d_{hj} + 0.5 (d_{hi} - d_{hj}) = d_{hi}$$

If $d_{hi} > h_{hj}$

$$d_{hk} = 0.5 d_{hi} + 0.5 d_{hj} - 0.5 (d_{hi} - d_{hj}) = d_{hj}$$

So we can see that this parameter setting leads to d_{min}:

$$d_{hk} = \min\{d_{hi}, d_{hj}\}$$

(b) With the parameter setting, we have

$$d_{hk} = 0.5 d_{hi} + 0.5 d_{hj} + 0.5 |d_{hi} - d_{hj}|$$

If $d_{hi} \leq h_{hj}$,

$$d_{hk} = 0.5 d_{hi} + 0.5 d_{hj} - 0.5 (d_{hi} - d_{hj}) = d_{hj}$$

If $d_{hi} > h_{hj}$

$$d_{hk} = 0.5 d_{hi} + 0.5 d_{hj} + 0.5 (d_{hi} - d_{hj}) = d_{hi}$$

So we can see that this parameter setting leads to d_{max}:

$$d_{hk} = \max\{d_{hi}, d_{hj}\}$$