CoDE - A simple example

Without loss of generality, assume that we observe bimodal data $\mathbb{X} = (x_1, x_2)$ with unimodal Gaussian data modalities. Therefore, there are two expert distributions $q(z|x_1)$ and $q(z|x_2)$, and one unknown consensus distribution $q(z|x_1, x_2)$, so we can drop the k superscript in the formulae below. Furthermore, assume that the unknown parameter is $\theta = 8$, the expert estimates and their uncertainty are $\mu_1 = 4$, $\mu_2 = 8$, $\sigma_1^2 = 3$, and $\sigma_2^2 = 1$, and that the estimates have correlation $\rho = 0.6$. Therefore, we have that

$$\boldsymbol{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} 4 \\ 8 \end{bmatrix} = \begin{bmatrix} \mu_1^1 \\ \mu_2^1 \end{bmatrix}, \quad \boldsymbol{e} = \begin{bmatrix} 4 - \theta \\ 8 - \theta \end{bmatrix} = \begin{bmatrix} e_1^1 \\ e_2^1 \end{bmatrix}$$

and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 3 & 0.6 \cdot \sqrt{3} \cdot \sqrt{1} \\ 0.6 \cdot \sqrt{3} \cdot \sqrt{1} & 1 \end{bmatrix}.$$

Then we can calculate the consensus distribution using Lemma 2 as follows:

$$\mathcal{A} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} \\ \alpha_{2,1} & \alpha_{2,2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \alpha_{1,1} + \alpha_{2,1} + \alpha_{1,2} + \alpha_{2,2},$$

$$\mathcal{B} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} \\ \alpha_{2,1} & \alpha_{2,2} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = (\alpha_{1,1} + \alpha_{2,1})\mu_1 + (\alpha_{1,2} + \alpha_{2,2})\mu_2,$$

and $q(z|x_1, x_2) \sim \mathcal{N}\left(\frac{(\alpha_{1,1}+\alpha_{2,1})\mu_1 + (\alpha_{1,2}+\alpha_{2,2})\mu_2}{\alpha_{1,1}+\alpha_{1,2}+\alpha_{2,1}+\alpha_{2,2}}, \frac{1}{\alpha_{1,1}+\alpha_{1,2}+\alpha_{2,1}+\alpha_{2,2}}\right)$, where $\boldsymbol{\Sigma}^{-1} = \alpha_{i,j}$.

CoE subsumes PoE: Now, let us assume that $\rho = 0$. Hence,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}, \quad \boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2} \begin{bmatrix} \sigma_2^2 & 0 \\ 0 & \sigma_1^2 \end{bmatrix} = \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix},$$

$\mathcal{A} = \tau_{1,1} + \tau_{2,2}$, $\mathcal{B} = \tau_{1,1}\mu_1 + \tau_{2,2}\mu_2$, and the consensus distribution is $q(z|x_1, x_2) \sim \mathcal{N}\left(\frac{\tau_{1,1}\mu_1 + \tau_{2,2}\mu_2}{\tau_{1,1}+\tau_{2,2}}, \frac{1}{\tau_{1,1}+\tau_{2,2}}\right)$, where $\tau_i = 1/\sigma_i^2$, like in [**?**]. Therefore, for $\rho = 0$ the CoDE parameters are simply PoE parameters.

As shown by [**?**], we can write the mean of the consensus distribution as

$$\mu_{CoDE} = \omega_1 \mu_1 + \omega_2 \mu_2,$$

where $\omega_1 = \frac{(\sigma_2^2 - \rho\sigma_1\sigma_2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$ and $\omega_2 = \frac{(\sigma_1^2 - \rho\sigma_1\sigma_2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$, to show that the posterior parameter μ_{CoDE} is a weighted average of the expert parameters μ_1 and μ_2. Note that the weights can be negative depending on the value of ρ with respect to σ_1 and σ_2. In the above example, for $\rho = 0.6$, the posterior mean is $\mu_{CoDE} = -0.02 * 4 + 1.02 * 8$ and for $\rho = 0$ the mean becomes $\mu_{CoDE} = 0.25 * 4 + 0.75 * 8$. In both cases, the weights sum up to 1. However, when the correlation between expert estimates is taken into account and the correlation is relatively high, the posterior mean leans even more towards the more accurate (with less variance)

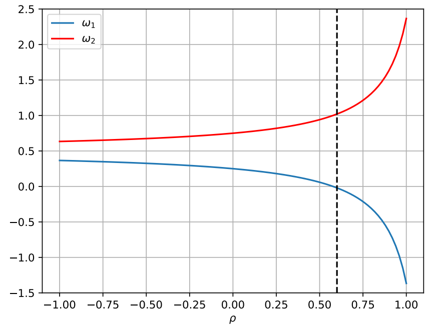

Figure 1: The consensus parameter μ_{CoDE} can be estimated as $\mu_{CoDE} = \omega_1\mu_1 + \omega_2\mu_2$, where the weights ω_1 and ω_2 are functions of the correlation parameter ρ.

estimate. Fig. 1 shows both weights as a function of ρ and, only for ρ greater than 0.5, the weight for the less accurate estimate becomes negative. On the other hand, for $\rho = 0$ the weights are always positive and μ_{CoDE} is a convex function of the expert mean parameters. Similarly, we can write the variance of the consensus distribution as

$$\sigma^2_{CoDE} = \frac{(1-\rho^2)\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

Note that when the dependence between experts is neglected, i.e. $\rho = 0$, the variance of the consensus distribution is higher than it should be, as the denominator increases.

Two expert distributions and 2D joint representation $\boldsymbol{z}$: In this case, $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$ and the matrices for the consensus distributions are

$$\boldsymbol{u} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1^1 \\ \mu_2^1 \\ \mu_1^2 \\ \mu_2^2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}^1 \\ \boldsymbol{\mu}^2 \end{bmatrix}, \quad \boldsymbol{e} = \begin{bmatrix} e_1^1 = \mu_1^1 - \theta_1 \\ e_2^1 = \mu_2^1 - \theta_1 \\ e_1^2 = \mu_1^2 - \theta_2 \\ e_2^2 = \mu_2^2 - \theta_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{bmatrix},$$

where

$$\boldsymbol{\Sigma}_d = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}$$

for $d = 1, 2$ is the covariance matrix of the d-th dimension, and $\mathbf{0}$ is a 2x2 zero matrix. Therefore,

$$\mathcal{A} = \boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & 0 & 0 \\ \alpha_{2,1} & \alpha_{2,2} & 0 & 0 \\ 0 & 0 & \alpha_{1,1} & \alpha_{1,2} \\ 0 & 0 & \alpha_{2,1} & \alpha_{2,2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

and

$$\mathcal{B} = \boldsymbol{u}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & 0 & 0 \\ \alpha_{2,1} & \alpha_{2,2} & 0 & 0 \\ 0 & 0 & \alpha_{1,1} & \alpha_{1,2} \\ 0 & 0 & \alpha_{2,1} & \alpha_{2,2} \end{bmatrix} \begin{bmatrix} \mu_1^1 \\ \mu_2^1 \\ \mu_1^2 \\ \mu_2^2 \end{bmatrix}$$

where $\boldsymbol{\Sigma}_d^{-1} = \alpha_{i,j}$ for $d = 1, 2$.