Featurization for Indistinguishable Atoms

Christian R. Schwantes

Department of Chemistry, Stanford University, Stanford, CA 94305

E-mail:

What is the overlap integral between k Gaussian density functions, each with the same variance, σ^2, but different means: $\{\mu_i\}_{i=1}^k$? Then overlap is given by:

$$S_k = \int d\mathbf{x} \prod_{i=1}^{k} \left[\frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} \exp\left(-\frac{|\mathbf{x} - \mu_i|^2}{2\sigma^2} \right) \right] \tag{1}$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{dk}{2}}} \int d\mathbf{x} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{k} |\mathbf{x} - \mu_i|^2 \right) \tag{2}$$

The product of many Gaussians is itself a Gaussian function, so we need only complete the square to find the new Gaussian:

$$\begin{aligned} \sum_{i=1}^{k} |\mathbf{x} - \mu_i|^2 &= \sum_{i=1}^{k} \mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mu_i + \mu_i^T\mu_i \\ &= k\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T \left(\sum_{i=1}^{k} \mu_i \right) + \sum_{i=1}^{k} \mu_i^T\mu_i \end{aligned}$$

*To whom correspondence should be addressed

Let $\mathbf{m} = \frac{1}{k}\sum_{i=1}^{k}\mu_i$, then we can complete the square in terms of this vector:

$$\begin{aligned}
\sum_{i=1}^{k}|\mathbf{x}-\mu_i|^2 &= k\left(\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{m} + \frac{\sum_{i=1}^{k}\mu_i^T\mu_i}{k}\right) \\
&= k\left(\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{m} + (-\mathbf{m}^T\mathbf{m}) - (-\mathbf{m}^T\mathbf{m}) + \frac{\sum_{i=1}^{k}\mu_i^T\mu_i}{k}\right) \\
&= k|\mathbf{x}-\mathbf{m}|^2 + k\frac{\sum_{i=1}^{k}\mu_i^T\mu_i}{k} - k\mathbf{m}^T\mathbf{m}
\end{aligned}$$

With, this we can rewrite the integral for calculating S_k:

$$\begin{aligned}
S_k &= \frac{1}{(2\pi\sigma^2)^{\frac{dk}{2}}}\int d\mathbf{x}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{k}|\mathbf{x}-\mu_i|^2\right) && (3)\\
&= \frac{1}{(2\pi\sigma^2)^{\frac{dk}{2}}}\int d\mathbf{x}\exp\left(-\frac{1}{2}\frac{|\mathbf{x}-\mathbf{m}|^2}{\sigma^2/k}\right)\exp\left(-\frac{\sum_{i=1}^{k}\mu_i^T\mu_i - k\mathbf{m}^T\mathbf{m}}{2\sigma^2}\right) && (4)\\
&= \frac{1}{(2\pi\sigma^2)^{\frac{dk}{2}}}(2\pi(\sigma^2/k))^{\frac{d}{2}}\exp\left(-\frac{\sum_{i=1}^{k}\mu_i^T\mu_i - k\mathbf{m}^T\mathbf{m}}{2\sigma^2}\right) && (5)\\
&= (2\pi\sigma^2)^{-\frac{d(k-1)}{2}}k^{-\frac{d}{2}}\exp\left(-\frac{\sum_{i=1}^{k}\mu_i^T\mu_i - k\mathbf{m}^T\mathbf{m}}{2\sigma^2}\right) && (6)\\
&= (2\pi\sigma^2)^{-\frac{d(k-1)}{2}}k^{-\frac{d}{2}}\exp\left(-\frac{\frac{1}{k}\sum_{i=1}^{k}\mu_i^T\mu_i - \mathbf{m}^T\mathbf{m}}{2\sigma^2/k}\right) && (7)\\
& && (8)
\end{aligned}$$

The numerator in the exponential can be written in terms of the variance of the means.

$$\frac{1}{k}\sum_{i=1}^{k}\mu_i^T\mu_i - \mathbf{m}^T\mathbf{m} = \frac{1}{k}\sum_{i=1}^{k}(\mu_i^T\mu_i) - \mathbf{m}^T\mathbf{m} \tag{9}$$

$$= \sum_{j=1}^{d}\left(\frac{1}{k}\sum_{i=1}^{k}(\mu_i)_j(\mu_i)_j\right) - m_j^2 \tag{10}$$

$$= \sum_{j=1}^{d}\mathrm{var}\Big((\mu_i)_j\Big) \tag{11}$$

$$= \sum_{j=1}^{d}\frac{1}{2k^2}\sum_{a=1}^{k}\sum_{b=1}^{k}\Big((\mu_a)_j - (\mu_b)_j\Big)^2 \tag{12}$$

$$= \frac{1}{k^2}\sum_{a=1}^{k}\sum_{b>a}^{k}|\mu_a - \mu_b|^2 \tag{13}$$

And so the overlap integral can be written as:

$$S_k = (2\pi\sigma^2)^{-\frac{d(k-1)}{2}}k^{-\frac{d}{2}}\exp\left(-\frac{\frac{1}{k}\sum_{i=1}^{k}\mu_i^T\mu_i - \mathbf{m}^T\mathbf{m}}{2\sigma^2/k}\right) \tag{14}$$

$$= (2\pi\sigma^2)^{-\frac{d(k-1)}{2}}k^{-\frac{d}{2}}\exp\left(-\frac{\frac{1}{k^2}\sum_{a=1}^{k}\sum_{b>a}^{k}|\mu_a - \mu_b|^2}{2\sigma^2/k}\right) \tag{15}$$

$$= (2\pi\sigma^2)^{-\frac{d(k-1)}{2}}k^{-\frac{d}{2}}\exp\left(-\frac{1}{2k\sigma^2}\sum_{a=1}^{k}\sum_{b>a}^{k}|\mu_a - \mu_b|^2\right) \tag{16}$$

$$\tag{17}$$