Bias Reduction and Metric Learning for Nearest-Neighbor Estimation of Kullback-Leibler Divergence (Supplementary Material)

Yung-Kyun Noh[1], Masashi Sugiyama[2], Song Liu[2], Marthinus C. du Plessis[2], Frank Chongwoo Park[3], and Daniel D. Lee[4]
[1]KAIST, Korea; [2]Tokyo Institute of Technology, Japan;
[3]Seoul National University, Korea; [4]University of Pennsylvania, USA

Appendix A. Derivation of Eq.(11)

Eq.(11) is derived from the calculation of average for $p(\mathbf{x}_{\mathrm{NN}})$ on the shell S of the hypersphere:

$$\int_{\{S:||\mathbf{x}_{\mathrm{NN}}-\mathbf{x}||=d\}} p(\mathbf{x}_{\mathrm{NN}})\ dV \tag{1}$$

$$\simeq \int p(\mathbf{x})\ dV + \int \nabla p(\mathbf{x})^\top \mathbf{w}\ dV(\mathbf{w}) + \frac{1}{2}\int \mathbf{w}^\top \nabla\nabla p(\mathbf{x})\mathbf{w}\ dV(\mathbf{w}) \tag{2}$$

with $\mathbf{w} = \mathbf{x}_{\mathrm{NN}} - \mathbf{x}$. Each term can be written as

$$\int p(\mathbf{x})\ dV = p(\mathbf{x})\int dV, \tag{3}$$

$$\int \nabla p(\mathbf{x})^\top \mathbf{w}\ dV \tag{4}$$
$$= \int_{-d}^{d} \nabla p(\mathbf{x})^\top \frac{\alpha\nabla p}{||\nabla p||} V_S(\alpha)\ d\alpha,$$

$$\frac{1}{2}\int \mathbf{w}^\top \nabla\nabla p(\mathbf{x})\mathbf{w}\ dV \tag{5}$$
$$= \frac{1}{2}\int\cdots\int_{\sqrt{\beta_1^2+\ldots+\beta_D^2}=d} (\beta_1\mathbf{u}_1+\ldots+\beta_D\mathbf{u}_D)^\top\nabla\nabla p(\mathbf{x}) (\beta_1\mathbf{u}_1+\ldots+\beta_D\mathbf{u}_D)\ d\beta_1\ldots d\beta_D,$$

with a volume constant $\gamma = \pi^{D/2}/\Gamma\left(\frac{D}{2}+1\right)$, a vector $\mathbf{w} = \mathbf{x}_{\mathrm{NN}} - \mathbf{x} = \beta_1\mathbf{u}_1 + \ldots + \beta_D\mathbf{u}_D$ with eigenvector bases of $\nabla\nabla p(\mathbf{x})$, $\mathbf{u}_1, \ldots, \mathbf{u}_D$, and corresponding constants to obtain $\mathbf{w}$, $\beta_1, \ldots, \beta_D$. Here, $V_S(\alpha)$ is the infinitesimal volume of hypersphere of radius $r = \sqrt{d^2-\alpha^2}$.

The zero-th order expansion, Eq.(3), is the $p(\mathbf{x})$ multiplied by the volume of the hypersphere shell with radius d. Thus, Eq.(3) is $p(\mathbf{x})D\gamma d^{D-1}\ dd$. The first order expansion, Eq.(4), is zero because of the symmetry. In order to calculate the 2nd expansion, Eq.(5), we need to calculate several integrations first: for $\beta = l\sin\theta$ and $d\beta = l\cos\theta d\theta$,

$$\int_{-l}^{l} d\beta (l^2-\beta^2)^m \tag{6}$$
$$= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} d\theta\ l\cos\theta(l^2\cos^2\theta)^m$$
$$= l^{2m+1}\pi^{1/2}\frac{\Gamma(m+1)}{\Gamma(m+\frac{3}{2})},$$

$$\int_{-l}^{l} d\beta\ \beta(l^2-\beta^2)^m \tag{7}$$
$$= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} d\theta\ l\cos\theta\ l\sin\theta(l^2\cos^2\theta)^{\frac{D}{2}-2} = 0,$$

$$\int_{-l}^{l} d\beta\ \beta^2(l^2-\beta^2)^m \tag{8}$$
$$= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} d\theta\ l\cos\theta\ l^2\sin^2\theta(l^2\cos^2\theta)^m$$
$$= \frac{l^{2m+3}}{2m+3}\pi^{1/2}\frac{\Gamma(m+1)}{\Gamma(m+\frac{3}{2})}.$$

We also note that

$$\int_0^{\frac{\pi}{2}} \cos^{k-2}\theta d\theta = \frac{k}{k-1}\int_0^{\frac{\pi}{2}}\cos^k\theta d\theta. \tag{9}$$

If we reformulate Eq.(5),

$$\int_{-d}^{d} d\beta_1 \int_{-\sqrt{d^2-\beta_1^2}}^{\sqrt{d^2-\beta_1^2}} d\beta_2 \cdots \int_{-\sqrt{d^2-\beta_1^2-\ldots-\beta_{D-1}^2}}^{\sqrt{d^2-\beta_2^2-\ldots-\beta_{D-1}^2}} d\beta_D \tag{10}$$
$$(\beta_1\mathbf{u}_1+\ldots+\beta_D\mathbf{u}_D)^\top\nabla\nabla p(\mathbf{x})(\beta_1\mathbf{u}_1+\ldots+\beta_D\mathbf{u}_D)$$
$$= \int_{-d}^{d} d\beta_1 \int_{-\sqrt{d^2-\beta_1^2}}^{\sqrt{d^2-\beta_1^2}} d\beta_2 \cdots \int_{-\sqrt{d^2-\beta_1^2-\ldots-\beta_{D-1}^2}}^{\sqrt{d^2-\beta_2^2-\ldots-\beta_{D-1}^2}} d\beta_D$$
$$\beta_D^2\lambda_D + \beta_D(\ldots) + (\beta_1\mathbf{u}_1+\ldots+\beta_{D-1}\mathbf{u}_{D-1})^\top \nabla\nabla p(\mathbf{x})(\beta_1\mathbf{u}_1+\ldots+\beta_{D-1}\mathbf{u}_{D-1}), \tag{11}$$

where λ_i is the eigenvalue of $\nabla\nabla p$ which corresponds to $\mathbf{u}_i$. Here, the second term $\beta_D(\ldots)$ vanishes according to Eq.(7), and we can use Eq.(6) and (8) to calculate the integration recursively. Now, Eq.(5) becomes

$$Eq.(5) = \frac{d^2}{2}\nabla^2 p(\mathbf{x}) \cdot D\gamma d^{D-1}\ dd. \quad (12)$$

If we divide Eq.(12) with the volume of the shell, $D\gamma d^{D-1}\ dd$, we get the deviation

$$\frac{\|\mathbf{x}_{\mathrm{NN}} - \mathbf{x}\|^2}{2D}\nabla^2 p(\mathbf{x}). \quad (13)$$

Appendix B. Derivation of metric matrix A in Eq.(28)

The solution of the semidefinite program in Eq.(24) is not unique. Here, we explain how we chose the analytic solution in Eq.(24). We first reformulate the semidefinite program using the eigenvalues of B: $b_1, \ldots, b_D$. The optimal solution A shares the eigenvectors with B from the form of the equation Eq.(24). Then the optimization problem changes to the problem of finding eigenvalues of A, and the optimization can be formulated as

$$\min_{\mathbf{x}} \left(\sum_i \frac{b_i}{x_i}\right)^2 \quad \text{where} \quad \prod_i x_i = 1,\ x_i \geq \epsilon > 0, \quad (14)$$

where the eigenvalues can be obtained from $\lim_{\epsilon\to 0}\mathbf{x}$. For simpler analysis, we use the change of variable $u_i = \log x_i$ to obtain a new objective function:

$$\begin{aligned} L &= \left(\sum_i b_i e^{-u_i}\right)^2 + C\left(\sum_i u_i\right) && (15)\\ &\quad - \sum_i \gamma_i\left(u_i - \log\epsilon\right) && (16)\end{aligned}$$

with Lagrangian multipliers C and γ_i for $i = 1, \ldots, D$. The derivative of this objective function with respect to u_j is zero:

$$\begin{aligned} \frac{\partial L}{\partial u_j} &= 2\left(\sum_i b_i e^{-u_i}\right)\left(-b_j e^{-u_j}\right) + C - \gamma_j && (17)\\ &= 0, && (18)\end{aligned}$$

and the following equation is satisfied for all i:

$$C = \gamma_j + 2\left(\sum_i b_i e^{-u_i}\right) b_j e^{-u_j} \quad (19)$$

Here, we consider the KKT condition for γ_i, where for each i, it satisfies either

$$\gamma_i = 0 \quad \text{or} \quad u_i = \log\epsilon \quad (20)$$

Here, we can find a solution with $\gamma_i = 0$ for all i, and we do not have to consider the second condition. If $\gamma_i = 0$ for all i,

$$C = 2\left(\sum_i b_i e^{-u_i}\right) b_j e^{-u_j} \quad \text{for all } i \quad (21)$$

there are two possible equivalencies for this equation,

$$\text{Condition 1:} \qquad \sum_i b_i e^{-u_i} = 0 \quad (22)$$

$$\text{Condition 2:} \qquad b_j e^{-u_j} = b_k e^{-u_k} \quad (23)$$

Condition 1 is equivalent to the equation $\sum_i \frac{b_i}{x_i} = 0$, and unless all b_i have the same sign, there are many possible solutions. On the other hand, in order for the condition 2 to be satisfied, either $b_i > 0$ for all i or $b_i < 0$ for all i. In this case, we have unique solutions,

$$\begin{aligned} x_i &= \frac{b_i}{\left(\prod_i b_i\right)^{\frac{1}{D}}} \quad \text{if } b_i > 0 \text{ for all } i && (24)\\ x_i &= -\frac{b_i}{\left(\prod_i b_i\right)^{\frac{1}{D}}} \quad \text{if } b_i < 0 \text{ for all } i. && (25)\end{aligned}$$

To obtain a metric that satisfies the boundary condition between the two situations where Condition 1 and Condition 2 are satisfied, we can set

$$\begin{aligned} x_i &= C_+\frac{b_i}{\left(\prod_i b_i\right)^{\frac{1}{D}}} \quad \text{if } b_i > 0 && (26)\\ x_i &= -C_-\frac{b_i}{\left(\prod_i b_i\right)^{\frac{1}{D}}} \quad \text{if } b_i < 0, && (27)\end{aligned}$$

and find C_+ and C_- that satisfies $\sum_i \frac{b_i}{x_i} = 0$: from $\sum_{i\in+}\frac{\left(\prod_i b_i\right)^{\frac{1}{D}}}{C+} - \sum_{i\in-}\frac{\left(\prod_i b_i\right)^{\frac{1}{D}}}{C-} = 0$, we can choose $C_+ = d_+$ and $C_- = d_-$ where d_+ and d_- are the number of positive eigenvalues and negative eigenvalues of B, respectively. This completes the derivation of the matrix A.