Supplemental online material for the paper:

"Multivariate Online Kernel Density Estimation with Gaussian Kernels"

(DOI:10.1016/j.patcog.2011.03.019, journal of Pattern Recognition)

Matej Kristan, Aleš Leonardis, Danijel Skočaj

April 26, 2013

Abstract

This document includes some detailed supplemental derivations used in the bandwidth estimation for the online Kernel Density Estimator which was proposed in the paper "Multivariate Online Kernel Density Estimation with Gaussian Kernels" by authors Matej Kristan, Aleš Leonardis, Danijel Skočaj (submitted to the journal of Pattern Recognition).

1 Functional approximation

In this section we detail the derivation of the approximation to the functional $R(p, \mathbf{F})$ from Section 3 of the above referenced paper. We write the multiple partial derivatives of a d-variate function $g(\mathbf{x})$ as

$$g^{(\mathbf{r})} = \frac{\partial^{|\mathbf{r}|}}{\partial x_1^{r_1} \dots \partial x_d^{r_d}} g(\mathbf{x}), \tag{1}$$

with $\mathbf{r} = (r_1, \dots, r_d)$ a vector of nonnegative integers and $|\mathbf{r}| = \sum_{i=1}^{d} r_i$. According to [1] (page, 98) we can rewrite

$$R(p, \mathbf{F}) = \int \mathrm{tr}^2\{\mathbf{F}\mathcal{G}_p(\mathbf{x})\} d\mathbf{x} \tag{2}$$

into

$$\begin{aligned} R(p, \mathbf{F}) &= \int \mathrm{tr}\{\mathbf{F}\mathcal{G}_{\mathrm{p}}(\mathbf{x})\}\mathrm{tr}\{\mathbf{F}\mathcal{G}_{\mathrm{p}}(\mathbf{x})\} d\mathbf{x} \\ &= \mathrm{vech}^T(\mathbf{F})\mathbf{\Psi}_{\mathcal{G}}\mathrm{vech}(\mathbf{F}), \end{aligned} \tag{3}$$

with $\mathbf{\Psi}_{\mathcal{G}}$ denoting a $\frac{1}{2}d(d+1) \times \frac{1}{2}d(d+1)$ matrix

$$\begin{aligned}\mathbf{\Psi}_{\mathcal{G}} = \int & \mathrm{vech}(2\mathcal{G}_p(\mathbf{x}) - \mathrm{dg}(\mathcal{G}_p(\mathbf{x}))) \times \\ & \mathrm{vech}^T(2\mathcal{G}_p(\mathbf{x}) - \mathrm{dg}(\mathcal{G}_p(\mathbf{x})))d\mathbf{x}, \end{aligned} \tag{4}$$

and where the notation dg denotes the diagonal matrix formed by replacing all off-diagonal entries by zeros. Since $\mathcal{G}_p(\mathbf{x})$ is a matrix of second partial derivatives, each entriy in $\mathbf{\Psi}_{\mathcal{G}}$ can be written in terms of functionals $\psi_{\mathbf{r}}$,

$$\begin{aligned}\psi_{\mathbf{r}} &= \int p^{(\mathbf{r}_1)}(\mathbf{x})p^{(\mathbf{r}_2)}(\mathbf{x})d\mathbf{x} \\ &= \int p^{(\mathbf{r})}(\mathbf{x})p(\mathbf{x})d\mathbf{x} \end{aligned} \tag{5}$$

with $\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2$ and with even $|\mathbf{r_1}|$ and $|\mathbf{r_2}|$. This means that entries in (3) are simply expectations of partial even derivatives of an unknown distribution $p(\mathbf{x})$. We now approximate the unknown distribution by a sample model $p_{\mathrm{s}}(\mathbf{x})$ and approximate its derivative through a kernel density estimate $p_{\mathbf{G}}^{(\mathbf{r})}(\mathbf{x})$, where

$$\begin{aligned}p_{\mathbf{G}}(\mathbf{x}) &= \sum_{i=1}^{N_{\mathrm{g}}} \alpha_{\mathrm{g}i}\phi_{\Sigma_{\mathrm{g}i}}(\mu_{\mathrm{g}i} - \mathbf{x}) \\ p_{\mathrm{s}}(\mathbf{x}) &= \sum_{j=1}^{N_{\mathrm{s}}} \alpha_{\mathrm{s}j}\phi_{\Sigma_{\mathrm{s}j}}(\mu_{\mathrm{s}j} - \mathbf{x}). \end{aligned}$$

This means that the entries $\psi_{\mathbf{r}}$ are approximated by

$$\begin{aligned}\hat{\psi}_{\mathbf{r}} &= \int p_{\mathbf{G}}^{(\mathbf{r})}(\mathbf{x})p_{\mathrm{s}}(\mathbf{x})d\mathbf{x} \\ &= \int p_{\mathbf{G}}^{(\mathbf{r}_1)}(\mathbf{x})p_{s}^{(\mathbf{r}_2)}(\mathbf{x})d\mathbf{x}. \end{aligned} \tag{6}$$

By matching the last line of the (6) with the first line of (5) we get the approximation of $\mathbf{\Psi}_{\mathcal{G}}$,

$$\begin{aligned}\hat{\mathbf{\Psi}}_{\mathcal{G}} = \int & \mathrm{vech}(2\mathcal{G}_{p_{\mathbf{G}}}(x) - \mathrm{dg}(\mathcal{G}_{p_{\mathbf{G}}}(x))) \times \\ & \mathrm{vech}^T(2\mathcal{G}_{ps}(x) - \mathrm{dg}(\mathcal{G}_{ps}(x)))d\mathbf{x}. \end{aligned} \tag{7}$$

Plugging $\hat{\mathbf{\Psi}}_{\mathcal{G}}$ back into (3) yields the following approximation of $R(p, \mathbf{F})$:

$$\hat{R}(p, \mathbf{F}, \mathbf{G}) = \int \mathrm{tr}\{\mathbf{F}\mathcal{G}_{p_{\mathbf{G}}}(\mathbf{x})\}\mathrm{tr}\{\mathbf{F}\mathcal{G}_{p_s}(\mathbf{x})\}d\mathbf{x}. \tag{8}$$

2 Closed-form functional calculation

To derive a closed-form solution to

$$\hat{R}(p, \mathbf{F}, \mathbf{G}) = \int \operatorname{tr}\{\mathbf{F}\mathcal{G}_{p_{\mathbf{G}}}(\mathbf{x})\}\operatorname{tr}\{\mathbf{F}\mathcal{G}_{p_s}(\mathbf{x})\} \tag{9}$$

which is based only on matrix algebra, we follow closely the derivation of a similar integral which was studied in the Appendix of M.P. Wand's paper [2]. As in that paper, we will require some established results:

$$\mathcal{G}_{\phi_{\mathbf{\Sigma}}(\cdot-\mu)} = \phi_{\mathbf{\Sigma}}(\mathbf{x})\{\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)(\mathbf{x}-\mu)^T - \mathbf{I}\}\mathbf{\Sigma}^{-1}, \tag{10}$$

$$\begin{gathered}\phi_{\mathbf{\Sigma}_i}(\mathbf{x}-\mu_i)\phi_{\mathbf{\Sigma}_j}(\mathbf{x}-\mu_j) = \\ \phi_{\mathbf{\Sigma}_i+\mathbf{\Sigma}_j}(\mu_i-\mu_j)\phi_{\mathbf{\Sigma}_i(\mathbf{\Sigma}_i+\mathbf{\Sigma}_j)^{-1}\mathbf{\Sigma}_j}(\mathbf{x}-\mu^*)\end{gathered} \tag{11}$$

where

$$\mu^* = \mathbf{\Sigma}_j(\mathbf{\Sigma}_i+\mathbf{\Sigma}_j)^{-1}\mu_i + \mathbf{\Sigma}_i(\mathbf{\Sigma}_i+\mathbf{\Sigma}_j)^{-1}\mu_j, \tag{12}$$

and

$$\begin{gathered}\operatorname{Cov}(\mathbf{X}^T\mathbf{A}\mathbf{X}, (\mathbf{X}-\mathbf{c})^T\mathbf{B}(\mathbf{X}-\mathbf{c})) = \\ 2tr[\mathbf{A}\mathbf{\Sigma}\mathbf{B}\{\mathbf{\Sigma}+2(\mu-\mathbf{c})\mu^T\}],\end{gathered} \tag{13}$$

where $\mathbf{X}$ is random vector distributed in $\phi_{\mathbf{\Sigma}}(\mu-\mathbf{x})$, $\mathbf{A}$ and $\mathbf{B}$ are $d\times d$ symmetric constant matrices and $\mathbf{c}$ is a $d\times 1$ constant vector. We start by expanding the integral

$$\begin{gathered}\int \operatorname{tr}\{\mathbf{F}\mathcal{G}_{p_{\mathbf{G}}}(\mathbf{x})\}\operatorname{tr}\{\mathbf{F}\mathcal{G}_{p_s}(\mathbf{x})\} = \\ \sum_{i=1}^{N_\mathrm{g}}\sum_{j=1}^{N_\mathrm{s}} \alpha_{\mathrm{g}i}\alpha_{\mathrm{s}j}\phi_{\mathbf{\Sigma}_{\mathrm{g}i}+\mathbf{\Sigma}_{\mathrm{s}j}}(\mu_{\mathrm{g}i}-\mu_{\mathrm{s}j}) \\ \times E[\operatorname{tr}\{\mathbf{F}\mathbf{\Sigma}_{\mathrm{g}i}^{-1}[(Y-\mu_{\mathrm{g}i})(Y-\mu_{\mathrm{g}i})^T\mathbf{\Sigma}_{\mathrm{g}i}^{-1}-\mathbf{I}]\} \\ \times\operatorname{tr}\{\mathbf{F}\mathbf{\Sigma}_{\mathrm{s}j}^{-1}[(Y-\mu_{\mathrm{s}j})(Y-\mu_{\mathrm{s}j})^T\mathbf{\Sigma}_{\mathrm{s}j}^{-1}-\mathbf{I}]\}],\end{gathered} \tag{14}$$

where $\mathbf{I}$ is an identity matrix and $\mathbf{Y}$ a random vector distributed in $\phi_{\mathbf{\Sigma}_{\mathrm{g}i}(\mathbf{\Sigma}_{\mathrm{g}i}+\mathbf{\Sigma}_{\mathrm{s}j})^{-1}\mathbf{\Sigma}_{\mathrm{s}j}}(\mathbf{x}-\mu_{ij}^*)$, with

$$\mu_{ij}^* = \mathbf{\Sigma}_{\mathrm{s}j}(\mathbf{\Sigma}_{\mathrm{g}i}+\mathbf{\Sigma}_{\mathrm{s}j})^{-1}\mu_{\mathrm{g}i} + \mathbf{\Sigma}_{\mathrm{g}i}(\mathbf{\Sigma}_{\mathrm{g}i}+\mathbf{\Sigma}_{\mathrm{s}j})^{-1}\mu_{\mathrm{s}j}. \tag{15}$$

Since $E(\mathbf{UV}) = \operatorname{Cov}(\mathbf{U},\mathbf{V}) + E(\mathbf{U})E(\mathbf{V})$ for two random variables $\mathbf{U}$ and $\mathbf{V}$ the expectation in (14) can be written as

$$\begin{gathered}\operatorname{Cov}\{(Y-\mu_{\mathrm{g}i})^T\mathbf{\Sigma}_{\mathrm{g}i}^{-1}\mathbf{F}\mathbf{\Sigma}_{\mathrm{g}i}^{-1}(Y-\mu_{\mathrm{g}i}), \\ (Y-\mu_{\mathrm{s}j})^T\mathbf{\Sigma}_{\mathrm{s}j}^{-1}\mathbf{F}\mathbf{\Sigma}_{\mathrm{s}j}^{-1}(Y-\mu_{\mathrm{s}j}) \\ +\operatorname{tr}\{\mathbf{F}\mathbf{\Sigma}_{\mathrm{g}i}^{-1}(E[(Y-\mu_{\mathrm{g}i})(Y-\mu_{\mathrm{g}i})^T]\mathbf{\Sigma}_{\mathrm{g}i}^{-1}-\mathbf{I})\} \\ \times\operatorname{tr}\{\mathbf{F}\mathbf{\Sigma}_{\mathrm{s}j}^{-1}(E[(Y-\mu_{\mathrm{s}j})(Y-\mu_{\mathrm{s}j})^T]\mathbf{\Sigma}_{\mathrm{s}j}^{-1}-\mathbf{I})\}.\end{gathered} \tag{16}$$

Since $\mu^*_{ij} - \mu_{\rm si} = \mathbf{\Sigma}_{\rm g\mathit{i}}(\mathbf{\Sigma}_{\rm g\mathit{i}} + \mathbf{\Sigma}_{\rm s\mathit{j}})^{-1}(\mu_{\rm s\mathit{j}} - \mu_{\rm g\mathit{i}})$, (13) and matrix algebra can be used to show that the covariance term is

$$2\mathrm{tr}\{\mathbf{F}(\mathbf{\Sigma}_{\mathrm{g}i} + \mathbf{\Sigma}_{\mathrm{s}j})^{-1}\mathbf{F}(\mathbf{\Sigma}_{\mathrm{g}i} + \mathbf{\Sigma}_{\mathrm{s}j})^{-1} \times \\ [\mathbf{I} - 2(\mu_{\mathrm{g}i} - \mu_{\mathrm{s}j})(\mu_{\mathrm{g}i} - \mu_{\mathrm{s}j})^T(\mathbf{\Sigma}_{\mathrm{g}i} + \mathbf{\Sigma}_{\mathrm{s}j})]\}.$$

Using $E[(Y-\mu_{\mathrm{g}i})(Y-\mu_{\mathrm{g}i})^T] = \mathbf{\Sigma}_{\mathrm{g}i}(\mathbf{\Sigma}_{\mathrm{g}i}+\mathbf{\Sigma}_{\mathrm{s}j})^{-1}\mathbf{\Sigma}_{\mathrm{s}j} + (\mu^*_{ij}-\mu_{\mathrm{g}i})(\mu^*_{ij}-\mu_{\mathrm{g}i})^T$ it can be shown that each of the factor in the second term is equal to

$$-\mathrm{tr}\{\mathbf{F}(\mathbf{\Sigma}_{\mathrm{g}i} + \mathbf{\Sigma}_{\mathrm{s}j})^{-1}[\mathbf{I} - (\mu_{\mathrm{g}i} - \mu_{\mathrm{s}j})(\mu_{\mathrm{g}i} - \mu_{\mathrm{s}j})^T(\mathbf{\Sigma}_{\mathrm{g}i} + \mathbf{\Sigma}_{\mathrm{s}j})^{-1}]\}. \quad (17)$$

Combining these with (14) and applying the definitions

$$\begin{aligned} \mathbf{A}_{ij} &= (\mathbf{\Sigma}_{\mathrm{gi}} + \mathbf{\Sigma}_{\mathrm{sj}})^{-1}, \\ \mathbf{B}_{ij} &= \mathbf{A}_{ij}\{\mathbf{I} - 2(\mu_{\mathrm{g}i} - \mu_{\mathrm{s}j})(\mu_{\mathrm{g}i} - \mu_{\mathrm{s}j})^T\mathbf{A}_{ij}\}, \\ \mathbf{C}_{ij} &= \mathbf{A}_{ij}\{\mathbf{I} - (\mu_{\mathrm{g}i} - \mu_{\mathrm{s}j})(\mu_{\mathrm{g}i} - \mu_{\mathrm{s}j})^T\mathbf{A}_{ij}\} \end{aligned} \quad (18)$$

leads to

$$\hat{R}(p, \mathbf{F}, \mathbf{G}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j\phi_{\mathbf{\Sigma}_{\mathrm{g}i}+\mathbf{\Sigma}_{\mathrm{s}j}}(\mu_{\mathrm{g}i} - \mu_{\mathrm{s}j}) \times \\ [2\mathrm{tr}(\mathbf{F}\mathbf{A}_{ij}\mathbf{F}\mathbf{B}_{ij}) + \mathrm{tr}^2(\mathbf{F}\mathbf{C}_{ij})]. \quad (19)$$

Finally, using the definitions

$$\mathbf{A}_{ij} = (\mathbf{\Sigma}_{\mathrm{gi}} + \mathbf{\Sigma}_{\mathrm{sj}})^{-1}\ ,\ \Delta_{ij} = \mu_i - \mu_j\ ,\ m_{ij} = \Delta_{ij}^T\mathbf{A}_{ij}\Delta_{ij}, \quad (20)$$

we can simplify (19) into

$$\hat{R}(p, \mathbf{F}, \mathbf{G}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j\phi_{\mathbf{A}_{ij}^{-1}}(\Delta_{ij}) \times \\ [2\mathrm{tr}(\mathbf{F}^2\mathbf{A}_{ij}^2)[1 - 2m_{ij}] + \mathrm{tr}^2(\mathbf{F}\mathbf{A}_{ij})[1 - m_{ij}]^2]. \quad (21)$$

References

[1] M. P. Wand, M. C. Jones, Kernel Smoothing, Chapman & Hall/CRC, 1995.

[2] M. P. Wand, Error analysis for general multivariate kernel estimators, Nonparametric Statistics 2 (1992) 1–15.