

Mathematical derivation of LocalGaussian computation for Manifold Parzen, and errata

Pascal Vincent

Département d'Informatique et Recherche Opérationnelle
Université de Montréal

P.O. Box 6128, Downtown Branch, Montreal, H3C 3J7, Qc, Canada

vincentp@iro.umontreal.ca

Technical Report 1259

Département d'Informatique et Recherche Opérationnelle

March 16, 2005

Abstract

The aim of this report is to correct an inconsistency in the mathematical formulas that appeared in our previously published *Manifold Parzen* article [1], regarding the computation of the density of “oriented” high dimensional Gaussian “pancakes” for which we store only the first d leading eigenvectors and eigenvalues (rather than a full $n \times n$ covariance matrix, or its inverse). We give a detailed derivation leading to the correct formulas.

1 Detailed mathematical derivation of LocalGaussian evaluation

We consider, in \mathbb{R}^n , the multivariate Gaussian density $\mathcal{N}_{\mu,C}$ parameterized by mean vector $\mu \in \mathbb{R}^n$ and $n \times n$ covariance matrix C . The density at any point $x \in \mathbb{R}^n$ is given by

$$\mathcal{N}_{\mu,C}(x) = \frac{1}{\sqrt{(2\pi)^n |C|}} e^{-\frac{1}{2}(x-\mu)'C^{-1}(x-\mu)} \quad (1)$$

where $|C|$ is the determinant of C .

Let $\tilde{x} = x - \mu$

Let $C = VDV'$ the eigen-decomposition of C where the columns of V are the orthonormal eigenvectors and D is a diagonal matrix with the eigenvalues sorted in decreasing order $(\lambda_1, \dots, \lambda_n)$.

We replace D with \tilde{D} which is a diagonal matrix containing modified eigenvalues $\tilde{\lambda}_{1..n}$ such that **all eigenvalues after $\tilde{\lambda}_d$ are given the same value σ^2** , i.e.

$$\tilde{\lambda}_{1..n} = (\tilde{\lambda}_1, \dots, \tilde{\lambda}_d, \sigma^2, \dots, \sigma^2)$$

Using $\tilde{C} = V\tilde{D}V'$ instead of C in equation 1, we get:

$$\begin{aligned} \tilde{\mathcal{N}}_{\mu,C}(x) &= \frac{1}{\sqrt{(2\pi)^n |\tilde{C}|}} e^{-\frac{1}{2} \tilde{x}' \tilde{C}^{-1} \tilde{x}} \\ &= e^{-\frac{1}{2} \log((2\pi)^n |\tilde{C}|)} e^{-\frac{1}{2} \tilde{x}' \tilde{C}^{-1} \tilde{x}} \\ &= e^{-0.5(n \log(2\pi) + \log(|\tilde{C}|) + \tilde{x}' \tilde{C}^{-1} \tilde{x})} \end{aligned} \tag{2}$$

$$\text{In other words} \quad \tilde{\mathcal{N}}_{\mu,C}(x) = e^{-0.5(\mathbf{r} + \mathbf{q})} \tag{3}$$

$$\text{with} \quad \mathbf{r} = n \log(2\pi) + \log(|\tilde{C}|) \tag{4}$$

$$\text{and} \quad \mathbf{q} = \tilde{x}' \tilde{C}^{-1} \tilde{x} \tag{5}$$

Moreover, since V is an orthonormal basis, we have

$$\begin{aligned} \|V' \tilde{x}\|^2 &= \|\tilde{x}\|^2 \\ \sum_{i=1}^n (V'_i \tilde{x})^2 &= \|\tilde{x}\|^2 \end{aligned} \tag{6}$$

where $V'_i \tilde{x}$ is the usual dot product between the i^{th} eigen-vector and centered input \tilde{x} .

$$\begin{aligned} \text{Proof :} \quad \|V' \tilde{x}\|^2 &= (V' \tilde{x})'(V' \tilde{x}) \\ &= \tilde{x}' V V' \tilde{x} \\ &= \tilde{x}' I \tilde{x} \\ &= \|\tilde{x}\|^2 \end{aligned}$$

In addition, having the above eigendecomposition,

$$\begin{aligned} \text{we have} \quad |\tilde{C}| &= \prod_{i=1}^n \tilde{\lambda}_i \\ \text{thus} \quad \log(|\tilde{C}|) &= \sum_{i=1}^n \log(\tilde{\lambda}_i) \end{aligned}$$

$$\log(|\tilde{C}|) = (n-d)\log(\sigma^2) + \sum_{i=1}^d \log(\tilde{\lambda}_i) \quad (7)$$

Replacing 7 in 4, we get

$$\mathbf{r} = n \log(2\pi) + (n-d)\log(\sigma^2) + \sum_{i=1}^d \log(\tilde{\lambda}_i) \quad (8)$$

In addition we have

$$\begin{aligned} \mathbf{q} &= \tilde{x}' \tilde{C}^{-1} \tilde{x} \\ &= \tilde{x}' (V \tilde{D} V')^{-1} \tilde{x} \\ &= \tilde{x}' V \tilde{D}^{-1} V' \tilde{x} \\ &= \sum_{i=1}^n \frac{1}{\tilde{\lambda}_i} (V'_i \tilde{x})^2 \\ &= \left(\sum_{i=1}^n \left(\frac{1}{\tilde{\lambda}_i} - \frac{1}{\sigma^2} \right) (V'_i \tilde{x})^2 \right) + \frac{1}{\sigma^2} \sum_{i=1}^n (V'_i \tilde{x})^2 \end{aligned}$$

Since $\tilde{\lambda}_i = \sigma^2$ for all $i > d$, $(\frac{1}{\tilde{\lambda}_i} - \frac{1}{\sigma^2}) = 0$ for $i > d$. As a consequence the first sum can be replaced by a sum from 1 to d (instead of from 1 to n). Also from equation 6 the second sum can be replaced by $\|\tilde{x}\|^2$. This yields:

$$\mathbf{q} = \frac{1}{\sigma^2} \|\tilde{x}\|^2 + \sum_{i=1}^d \left(\frac{1}{\tilde{\lambda}_i} - \frac{1}{\sigma^2} \right) (V'_i \tilde{x})^2 \quad (9)$$

We have thus eliminated the need to store and compute with eigenvectors $V_{d+1} \dots V_n$.

2 Summing up: the correct formulas

To sum this all up, we can compute the density as follows:

$$\tilde{\mathcal{N}}_{\mu, C}(x) = e^{-0.5(\mathbf{r} + \mathbf{q})} \quad (10)$$

$$\text{with } \mathbf{r} = n \log(2\pi) + (n-d)\log(\sigma^2) + \sum_{i=1}^d \log(\tilde{\lambda}_i) \quad (11)$$

$$\text{and } \mathbf{q} = \frac{1}{\sigma^2} \|\tilde{x}\|^2 + \sum_{i=1}^d \left(\frac{1}{\tilde{\lambda}_i} - \frac{1}{\sigma^2} \right) (V'_i \tilde{x})^2 \quad (12)$$

3 Errata for Manifold Parzen article

Typo in the pseudo-code for `Mparzen::train` step 4) should be $\lambda_{ij} = \sigma^2 + \frac{s_j^2}{k}$ instead of $\sigma^2 + \frac{s_j^2}{l}$.

Also, in our initial experiments, we actually considered two possible choices for $\tilde{\lambda}_i$ (for $i \leq d$) and σ^2 :

a) $(\tilde{\lambda}_1, \dots, \tilde{\lambda}_d) = (\lambda_1, \dots, \lambda_d)$ and $\sigma^2 = \lambda_{d+1}$

which leads to:

$$\begin{aligned} \mathbf{r} &= n \log(2\pi) + (n-d) \log(\sigma^2) + \sum_{i=1}^d \log(\lambda_i) \\ \mathbf{q} &= \frac{1}{\sigma^2} \|\tilde{x}\|^2 + \sum_{i=1}^d \left(\frac{1}{\lambda_i} - \frac{1}{\sigma^2} \right) (V_i' \tilde{x})^2 \end{aligned}$$

b) σ^2 is a user specified value and $(\tilde{\lambda}_1, \dots, \tilde{\lambda}_d) = (\lambda_1 + \sigma^2, \dots, \lambda_d + \sigma^2)$

which leads to:

$$\begin{aligned} \mathbf{r} &= n \log(2\pi) + (n-d) \log(\sigma^2) + \sum_{i=1}^d \log(\lambda_i + \sigma^2) \\ \mathbf{q} &= \frac{1}{\sigma^2} \|\tilde{x}\|^2 + \sum_{i=1}^d \left(\frac{1}{\lambda_i + \sigma^2} - \frac{1}{\sigma^2} \right) (V_i' \tilde{x})^2 \end{aligned}$$

We mentioned only scenario **b)** in the Manifold Parzen article [1] (due to space constraints). But somehow these two slightly different versions got mixed up in the write-up, leading to the somewhat inconsistent formulas that appear in the article (taking \mathbf{r} from **b)** and \mathbf{q} from **a)**). In addition, we mistakenly wrote $d \log(2\pi)$ instead of $n \log(2\pi)$.

However, after verification, the actual *code* used to perform the experiments reported in the article (implementing scenario **b)**) appears correct.

References

- [1] Pascal Vincent and Yoshua Bengio. Manifold parzen windows. In S. Thrun, S. Becker and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 825–832, Cambridge, MA, 2003. MIT Press.