Supplementary material for the paper: Robust Nonparametric Regression with Metric-Space valued Output

Matthias Hein
Department of Computer Science, Saarland University
Campus E1 1, 66123 Saarbrücken, Germany
hein@cs.uni-sb.de

1 Missing Proof from Section 2

Lemma 1 *Let N be a complete metric space such that $d(x,y) < \infty$ for all $x, y \in N$ and every closed and bounded set is compact. If Γ is (α, s)-bounded and $R'_\Gamma(x,q) < \infty$ for some $q \in N$, then*

- $R'_\Gamma(x,p) < \infty$ *for all* $p \in N$,
- $R'_\Gamma(x,\cdot)$ *is continuous on* N,
- *The set of minimizers* $Q^* = \underset{q \in N}{\arg\min}\, R'_\Gamma(x,q)$ *exists and is compact.*

Proof: As Γ is montonically increasing and convex, we have for any $p, y \in N$,

$$\Gamma\big(d_N(p,y)\big) \leq \Gamma\big(d_N(p,q) + d_N(q,y)\big) \leq \frac{1}{2}[\Gamma\big(2d_N(p,q)\big) + \Gamma\big(2d_N(q,y)\big)],$$

Moreover, since Γ is (α, s)-bounded we have,

$$\Gamma(2x) \leq a\Gamma(x)\,\mathbb{1}_{x \geq s} + \Gamma(2s)\,\mathbb{1}_{x<s}\,.$$

Taking expectations with respect to $Y|X = x$ we get,

$$R'_\Gamma(x,p) \leq \Gamma(2s) + \frac{a}{2}\Gamma\big(d_N(p,q)\big) + \frac{a}{2}R'_\Gamma(x,q).$$

Next, we show continuity of $R'_\Gamma(x,.)$. Using Lemma 2 we get,

$$\begin{aligned}|R'_\Gamma(x,p) - R'_\Gamma(x,q)| &= |\mathbb{E}[\Gamma\big(d_N(p,Y)\big) - \Gamma\big(d_N(q,Y)\big)]| \\ &\leq d(p,q)\,|\mathbb{E}[\max\{\Gamma'\big(d_N(p,Y)\big), \Gamma'\big(d_N(q,Y)\big)\}]|.\end{aligned}$$

Now, for $x \geq s$ we have $\Gamma'(x) \leq \frac{\Gamma(2x)-\Gamma(x)}{x} \leq (a-1)\frac{\Gamma(x)}{s}$ and for $x < s$, $\Gamma'(x) \leq \Gamma'(s)$. Thus

$$\mathbb{E}[\Gamma'\big(d_N(p,Y)\big)] \leq \frac{(a-1)}{s}\mathbb{E}[\Gamma\big(d_N(p,Y)\big)] + \Gamma'(s),$$

which shows using $\max\{a,b\} \leq a+b$ the continuity of $R'_\Gamma(x,\cdot)$.

Finally, we consider the set $S_\varepsilon = \{q \in N \,|\, R'_\Gamma(x,q) \leq \inf_{p \in N} R'_\Gamma(x,p) + \varepsilon\}$ which is closed since $R'(x,\cdot)$ is continuous. Moreover, let $q_1, q_2 \in S_\varepsilon$, then

$$\Gamma\big(d_N(q_1,q_2)\big) \leq \Gamma(2s) + \frac{a}{2}\Gamma\big(d_N(q_1,y)\big) + \frac{a}{2}\Gamma\big(d_N(q_2,y)\big) \leq \Gamma(2s) + \frac{a}{2}R'_\Gamma(x,q_1) + \frac{a}{2}R'_\Gamma(x,q_2).$$

For $x \geq s$ we have shown above $x \leq (a-1)\frac{\Gamma(x)}{\Gamma(x')} \leq \frac{\Gamma(x)}{\Gamma'(s)}$ and thus either $d_N(q_1,q_2) \leq s$ or

$$d_N(q_1,q_2) \leq (a-1)\frac{\Gamma(2s) + \frac{a}{2}R'_\Gamma(x,q_1) + \frac{a}{2}R'_\Gamma(x,q_2)}{\Gamma'(s)},$$

which shows that the set S_ε is bounded and thus compact. It is non-empty since $R'_\Gamma(x,\cdot)$ is continuous. The set of minimizers $Q^* = \cap_{\varepsilon>0} S_\varepsilon$ is compact and non-empty as it is the intersection of a nested sequence of non-empty, compact sets. $\square$

2 Missing Proofs from Section 5 and 7

The supplementary material contains the proofs which due to space constraints could not be included into the paper. For convenience we restate here Assumptions (A1) from the paper.

Assumptions (A1):

- $(X_i, Y_i)_{i=1}^l$ is an i.i.d. sample of P on $M \times N$,
- M and N are compact manifolds,
- The data-generating measure P on $M \times N$ is absolutely continuous with respect to the natural volume element,
- The marginal density on M fulfills: $p(x) \geq p_{\min}, \forall\, x \in M$,
- The density $p(y,\cdot)$ is continuous on M for all $y \in N$,
- The kernel fulfills: $a\, 1_{s \leq r_1} \leq k(s) \leq b\, e^{-\gamma\, s^2}$ and $\int_{\mathbb{R}^m} \|x\|\; k(\|x\|)\, dx < \infty$,
- The loss $\Gamma : \mathbb{R}_+ \to \mathbb{R}_+$ is (α, s)-bounded.

This proposition collects results from [1].

Proposition 1 *Let M be a compact m-dimensional Riemannian manifold. Then, there exists $r_0 > 0$ and $S_1, S_2 > 0$ such that for all $x \in M$ all balls $B(x,r)$ with radius $r \leq r_0$ it holds,*

$$S_1\, r^m \;\leq\; \mathrm{vol}\big(B(x,r)\big) \;\leq\; S_2\, r^m.$$

Moreover, the cardinality K of a δ-covering of M is upper bounded as, $K \leq \frac{\mathrm{vol}(N)}{S_1}\left(\frac{2}{\delta}\right)^m$.

Proposition 2 *Let the assumptions A1 hold, then if f is continuous we get for any $x \in M \backslash \partial M$,*

$$\lim_{h \to 0} \int_M k_h(d_M(x,z)) f(z)\, dV(z) \;=\; C_x f(x),$$

where $C_x = \lim_{h\to 0} \int_M k_h(d_M(x,z))\, dV(z) > 0$. If moreover f is Lipschitz continuous with Lipschitz constant L, then there exists a $h_0 > 0$ such that for all $h < h_0(x)$,

$$\int_M k_h(d_M(x,z)) f(z)\, dV(z) = C_x\, f(x) + O(h).$$

Proof: We denote by $\mathrm{inj}(M)$ the injectivity radius of M. As f is continuous for any $\varepsilon > 0$, $\exists \delta$ such that $d(x,z) < \delta$ implies $|f(x) - f(z)| < \varepsilon$. Suppose that ε is chosen small enough, so that $\delta < \mathrm{inj}(M)$,

$$\begin{aligned}
&\int_M k_h(d_M(x,z))\big(f(z) - f(x)\big)\, dV(z)\\
\leq\, &\varepsilon \int_{B(x,\delta)} k_h(d_M(x,z))\, dV(z) + 2\, \|f\|_\infty \int_{M \backslash B(x,\delta)} k_h(d_M(x,z))\, dV(z)\\
\leq\, &\varepsilon \int_{B(x,\delta)} k_h(\|y\|) dy + \|f\|_\infty \frac{\mathrm{vol}(M)}{h^m} b e^{-\gamma \frac{\delta^2}{h^2}},
\end{aligned}$$

where we have introduced in the last step normal coordinates centered at x on $B(x,\delta)$ so that $d_M(x,z) = \|y\|$. Note, that the second term is independent of ε and for each $\delta > 0$ converges to zero as $h \to 0$. Next, we note that the volume element on the ball $B(x, \delta)$ can be upper bounded as, $dV(y) = \sqrt{\det g}\big|_y \, dy \leq C \, dy$. Thus,

$$\int_{B(x,\delta)} k_h(\|y\|) dV(y) \leq C \int_{B(x,\delta)} k_h(\|y\|) dy = C \int_{B(x,\frac{\delta}{h})} k(\|y'\|) dy' \leq C \int_{\mathbb{R}^m} k(\|y'\|) dy',$$

where we made the substitution $y' = \frac{y}{h}$. Note that thus the upper bound on the first term is independent of h and both terms can be made arbitrarily small. Finally, using Proposition 1, we get

$$\int_M k_h(d_M(x,z)) \, dV(z) \geq \frac{a}{h^m} \int_M 1_{d_M(x,z) \leq h \, r_1} \, dV(z) = \frac{a}{h^m} \, \mathrm{vol}(B(x, hr_1)) \geq a \, S_1 \, r_1^m,$$

so that $C_x = \lim_{h \to 0} \int_M k_h(d_M(x,z)) \, dV(z) > 0$.

For Lipschitz continuous function f choose $\delta = \mathrm{inj}(M)$. The second term on $M \backslash B(x, \delta)$ can be treated as above noting that $\frac{1}{h^m} e^{-\gamma \frac{\delta^2}{h^2}} \leq C_2 h$ for sufficiently small h. Moreover,

$$\int_{B(x,\delta)} k_h(d_M(x,z)) |f(z) - f(x)| \, dV(z) \leq \; L \int_{B(x,\delta)} k_h(d_M(x,z)) d_M(x,z) \, dV(z)$$

$$\leq L \, C_1 \int_{B(0,\delta)} k_h(\|y\|) \, \|y\| \; dy = \; h \, C_1 \, L \int_{B(0,\frac{\delta}{h})} k(\|y'\|) \, \|y'\| \; dy' \leq h \, C_1 \, L \int_{\mathbb{R}^m} k(\|y'\|) \, \|y'\| \; dy',$$

where we again used normal coordinates y centered at x and the coordinate transformation $y' = \frac{y}{h}$. Moreover, $\int_{\mathbb{R}^m} k(\|y'\|) \, \|y'\| \; dy' < \infty$ by assumption on the kernel function. $\square$

Lemma 2 *Let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be convex, differentiable and monotonically increasing. Then*

$$\min\{\phi'(x), \phi'(y)\} |y - x| \; \leq \; |\phi(y) - \phi(x)| \; \leq \; \max\{\phi'(x), \phi'(y)\} |y - x|.$$

Proof: Using the first order condition of a convex function and $\phi(x) \leq \phi(y)$ for $x \leq y$,

$$\begin{aligned} \phi(y) - \phi(x) &\geq \phi'(x)(y - x) \quad &&\Rightarrow \phi(x) - \phi(y) \leq \phi'(x)(x - y), \\ \phi(x) - \phi(y) &\geq \phi'(y)(x - y) \quad &&\Rightarrow \phi(y) - \phi(x) \leq \phi'(y)(y - x). \end{aligned}$$

The left part yields the lower bound and the right part the upper bound. $\square$

References

[1] M. Hein. Uniform convergence of adaptive graph-based regularization. In G. Lugosi and H. Simon, editors, *Proc. of the 19th Conf. on Learning Theory (COLT)*, pages 50–64, Berlin, 2006. Springer.