Toward Manifold-Adaptive Learning

Amir massoud Farahmand **Csaba Szepesvári**
Department of Computing Science, University of Alberta
Edmonton, Canada
{amir, szepesva}@cs.ualberta.ca

Jean-Yves Audibert
CERTIS - Ecole des Ponts
19, rue Alfred Nobel - Cité Descartes, 77455 Marne-la-Vallée France
audibert@cermics.enpc.fr

1 Introduction

Inputs coming from high-dimensional spaces are common in many real-world problems such as a robot control with visual inputs. Yet learning in such cases is in general difficult, a fact often referred to as the "curse of dimensionality". In particular, in regression or classification, in order to achieve a certain accuracy algorithms are known to require exponentially many samples in the dimension of the inputs in the worst-case [1]. The exponential dependence on the input dimension forces us to develop methods that are efficient in exploiting regularities of the data. Classically, smoothness is the best known example of such a regularity. In this abstract we outline two methods for two problems that are efficient in exploiting when the data points lie on a low dimensional submanifold of the input space.

Specifically, we consider the case when the data points lie on a manifold M of dimension d, which is embedded in the higher-dimensional input space with dimension D (i.e. $d \leq D$). A method is called *manifold-adaptive* if its sample complexity can be bounded by a quantity whose exponent depends only on d and not on D. Thus a manifold-adaptive method may enjoy a considerably better sample complexity whenever $d \ll D$. Although there are many learning methods that are designed to be manifold adaptive (or manifold friendly), they more often than not lack a rigorous proof of this property (one exception is the recent work of Scott and Nowak on dyadic decision trees in a classification context, cf. [2]).

The first method, proposed by us earlier in [3], concerns the problem of estimating the dimension of a manifold based on points sampled from it. The second method is the classical k-nearest neighbor regressor. We find it intriguing that this method was not specifically designed to be manifold-adaptivity, yet it is relatively simple to prove that it possesses this property.

2 Manifold-Adaptive Dimension Estimation

Let the observed i.i.d. samples $X_1, \ldots, X_n \in \mathbb{R}^D$ come from a distribution supported on the manifold M. Define $\eta(x, r)$ by the relation $\mathbb{P}\left(X_i \in B(x, r)\right) = \eta(x, r) r^d$, where $B(x, r) \in \mathbb{R}^D$ is a ball of radius r around the point $x \in M$ in the Euclidean space $\mathbb{R}^D$. Consider the local estimates of the dimension defined

$$\hat{d}(x) = \frac{\ln 2}{\ln(\hat{r}^{(k)}(x)/\hat{r}^{(\lceil k/2 \rceil)}(x))}, \tag{1}$$

where $\hat{r}^{(k)}(x)$ denotes the distance of point x to its kth nearest neighbor in the data $X_1, \ldots, X_n$. Aggregate the local estimates by averaging or voting:

$$\hat{d}_{\text{average}} = \left[\frac{\sum_{i=1}^n (\hat{d}(X_i) \wedge D)}{n} \right], \qquad \hat{d}_{\text{vote}} = \underset{d' \in \mathbb{N}^+}{\arg\max} \ \sum_{i=1}^n \mathbb{I}_{\{\hat{d}(X_i) = d'\}}. \tag{2}$$

Here $a \wedge b = \min(a, b)$ and $[x]$ denotes the rounded value of x. Having the following regularity assumptions about the manifold and the sampling distribution, we get Theorem 1.

Assumption 1 *(1) For any point x in the support M of the distribution generating the data, the function $\eta(x,r) = r^{-d}\mathbb{P}[X \in \mathcal{B}_{\mathbb{R}^D}(x,r)]$ defined for $r > 0$ is bounded away from zero and locally Lipschitz to the extent that for some positive real numbers $\eta_{\min}$, $\tilde{r}$ and L, for any $(x, r, r') \in M \times (0,\tilde{r}) \times (0,\tilde{r})$, we have $\eta(x,r) \geq \eta_{\min}$ and $|\eta(x,r') - \eta(x,r)| \leq L|r' - r|$. (2) $\exists \tilde{r} > 0$ satisfying for any $x \in M$, $M \cap B(x,\tilde{r}) \subset \left\{y \in \mathbb{R}^D : \text{angle}(y - x, T_x M) < \pi/12\right\}$, where $T_x M$ is the tangent space of M at $x \in M$.*

Theorem 1 *Under Assumption 1, the following bounds hold:*

$$\log\left(\mathbb{P}\left(\hat{d}_{\text{vote}} \neq d\right)\right) \leq -\frac{c' n}{(k n_d)^2} + O(1); \qquad \log\left(\mathbb{P}\left(\hat{d}_{\text{average}} \neq d\right)\right) \leq -\frac{c'' n}{(D n_d)^2} + O(1). \quad (3)$$

Here c' and c'' are positive constants that do not depend on k, n, d and D and where n_d is the minimal number of cones in $\mathbb{R}^d$ centered at the origin and having angles $\pi/12$ such that $\mathbb{R}^d$ is covered by the union of the cones. n_d is upper bounded by c^d for some universal constant $c < 8$.

This theorem shows that the difficulty of dimension estimation depends mainly on d and not D. (Though the rate for the averaging method depends on D, the dependence is only polynomial.) We also prove that Assumption 1 can be substantially relaxed [4].

3 Manifold-Adaptive k-NN Regression

The k-NN estimator is probably the simplest nonparametric regression method. Hence, we think that it is interesting that this simple method can be made manifold adaptive relatively easily. The main insight of the proof of this fact is that the expected distance of a random point to its nearest neighbor depends only on the dimension of the manifold and not on the dimension of the input. More precisely, we have the following, improved version of Lemma 6.4 of [1] (cf. [4]):

Lemma 1 *Assume that M is a d-dimensional bounded manifold. Let $X, X_1, \ldots, X_n$ be i.i.d. samples from M, $X_{(1)}(X)$ be the nearest neighbor of X among $(X_1, \ldots, X_n)$. Then, assuming $d \geq 3$, we have*

$$E\{\|X_{(1)}(X) - X\|^2\} \leq \frac{c}{n^{2/d}}.$$

By replacing this lemma wherever Lemma 6.4 is used in Theorem 6.2 of [1], we get the following manifold-adaptive convergence rate for a k-NN estimator.

Theorem 2 *Assume that M is a d-dimensional bounded manifold. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. samples such that $X_i \in M$, $Y_i \in \mathbb{R}$, $\text{Var}\,[Y_1] < +\infty$, the regressor $m(x) = E[Y|X = x]$ is C-Lipschitz, $d \geq 3$. Let m_n be the k_n-NN estimate where k_n is an appropriate integer that depends on the estimated dimension of the manifold. Then $\mathbb{E}\left[\|m_n - m\|^2\right] \leq O(n^{-2/(d+2)})$.*

We expect that with the techniques developed here results similar to Theorem 2 can be shown to hold for a larger class of regressor, including regressors that are able to adapt to other regularities of the data, such as the smoothness of the regressor.

References

[1] L. Györfi, M. Kohler, A. Kryzak, and H. Walk. *A distribution - Free Theory of Nonparametric Regression*. Springer, 2002.

[2] C. Scott and R. Nowak. Minimax-optimal classification with dyadic decision trees. *IEEE Transactions on Information Theory*, 52:1335–1353, 2006.

[3] A. M. Farahmand, Cs. Szepesvári, and J.-Y. Audibert. Manifold-adaptive dimension estimation. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 265–272, New York, NY, USA, 2007. ACM Press.

[4] A. M. Farahmand, Cs. Szepesvári, and J.-Y. Audibert. Nearest neighborhood methods for manifold-adaptive dimension estimation and regression. *submitted*, 2007.