Regression with Sparse Approximations of Data

Noorzad, Pardis; Sturm, Bob L.

Published in:
Proceedings of the European Signal Processing Conference

Publication date:
2012

Document Version
Early version, also known as pre-print

Link to publication from Aalborg University

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: August 17, 2024
We propose SPARROW (SPARse appROXimation WEighted regression), a new method for locally polynomial regression function estimation, and an extension of sparse representation classification to the regression setting.

To estimate the regression function at a point, SPARROW uses Taylor polynomial expansion around that point, least-squares optimal parameter estimation, and sparse approximation in terms of a dictionary of regressors and regressands.

Our results show that locally constant SPARROW performs competitively, but the locally linear form, with and without regularization, does not.

What is Local Regression?

Consider we have a dataset of $N$ observations indexed by $\Omega := \{1, \ldots, N\}$:

$$ D := \{(x_i, y_i) : x_i \in \mathbb{R}^M, y_i \in \mathbb{R}, i \in \Omega\} $$

We wish to estimate the regression function $f(x) : \mathbb{R}^M \to \mathbb{R}$ at a point $z \in \mathbb{R}^M$.

1. Approximating this function by a Taylor polynomial about $z$, we have

$$ f(x) \approx f(z) + (x - z)^T \theta z + \frac{1}{2}(x - z)^T H_z (x - z) $$

where $\theta z$ and $H_z$ are the gradient and Hessian of $f(x)$, evaluated at $z$.

2. We can solve for $f(z), \theta z$ and $H_z$ by

$$ \min_{\theta z, H_z} \sum_{i \in \Omega} \alpha(z)_i \parallel y - f(z) - (x_i - z)^T \theta z - \frac{1}{2}(x_i - z)^T H_z (x_i - z) \parallel^2 $$

where $\alpha(z)_i$ is the $i$th observation weight. We can pose this as

$$ \min_{\theta z, H_z} \| [A_{yi}^2[y - X_i \Theta z]] \|_F^2 $$

where $[A_{yi}] := \alpha(z)_i$ and zero else, $\Theta z := \{f(z), \theta z, \text{vec}(H_z)\}_i$, and $X_z : = \begin{bmatrix} 1 (x_i - z)^T \text{vec}^T((x_i - z)(x_i - z)^T) \\ 1 (x_i - z)^T \text{vec}^T((x_i - z)(x_i - z)^T) \end{bmatrix}$. The notation vec(B) is the supervector of half of the symmetric matrix $B$.

3. The first element of the solution $\hat{\theta} z = (X_z^T A_z X_z)^{-1} X_z^T A_z y$ gives the least-squares optimal locally polynomial estimate of $f(z)$

$$ \hat{f}(z) = \hat{\theta} z (X_z^T A_z X_z)^{-1} X_z^T A_z y. $$

Taking only the first column of $X_z$ gives a locally constant estimate of $f(z)$:

$$ \hat{f}(z) = (1^T A_z 1)^{-1} A_z y = \sum_{i \in \Omega} \alpha(z)_i y_i \sum_{i \in \Omega} \alpha(z)_i z $$

Taking the first two columns gives a locally linear estimate of $f(z)$.

We must now define the $N$ observation weights $\{\alpha(z)_i : i \in \Omega\}$.

Weighted $k$-nearest neighbor regression (Wk-NNR) defines the weights by the reciprocal of their Euclidean distance to $z$.

Nadaraya-Watson kernel regression (NWR) method defines the weights using a kernel function, e.g., Gaussian, evaluated with respect to $z$.

SPARROW defines the weights using the sparse approximation of $z$ with respect to the observed points in $D$.

How SPARROW Defines the Observation Weights

We construct a dictionary matrix by concatenating normalized regressors

$$ D := \begin{bmatrix} x_1 \\ |x_1|_2 \\ |x_2|_2 \\ \vdots \\ |x_N|_2 \end{bmatrix}. $$

For a given point $z$, SPARROW finds a solution to $z \approx D s$ such that $s$ has many zero elements by solving the basis pursuit denoising (BPDN) problem

$$ \min_{s \in \mathbb{R}^M} \|s\|_1 \text{ subject to } \frac{\|z - D s\|_2^2}{\|s\|_2^2} \leq c^2 $$

where $c > 0$. Defining $\Sigma$ as a diagonal matrix of the unbiased estimates of the variances observed in the dimensions of the regressors in $D$, SPARROW then defines the $i$th observation weight $\alpha(z)_i$ as

$$ \alpha(z)_i = \frac{\|z - x_i\|_v^2 (\Sigma)^{-1}}{\sum_{i \in \Omega} \|z - x_i\|_v^2 (\Sigma)^{-1} s_i} $$

where $s_i$ is the $i$th element of $s, i \in \Omega$.

Experiments and Simulations

We perform experiments with various datasets to illustrate the competitiveness of the proposed method. The Abalone dataset is a regression problem, and the Bodyfat dataset is a classification problem. The Housing and MPG datasets are used to illustrate the effectiveness of the proposed method in real-world applications.

The figures show the results of the proposed method compared to other methods, such as Nadaraya-Watson kernel regression (NWR) and linear regression (MLR). The results indicate that the proposed method performs competitively with other methods.

Acknowledgments

B. L. Sturm is supported in part by Independent Postdoc Grant 11-105218 from Det Frie Forskningsråd.