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):
Introduction

We propose SPARRe appRoXimation Weigted regression (SPARROW), 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 Ω := {1, ..., 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 \( z \):
   \[ f(x) \approx f(z) + (x - z)^T \theta_z + \frac{1}{2}(x - z)^T H_z(x - z) \]
   \( \theta_z \) 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
   \[ \min_{f(z), \theta_z, H_z} \sum_{i \in \Omega} \alpha_i(z) \left[ y_i - f(z) - (x_i - z)^T \theta_z - \frac{1}{2}(x_i - z)^T H_z (x_i - z) \right]^2 \]
   where \( \alpha_i(z) \) is the ith observation weight. We can pose this as
   \[ \min_{\theta_z} \left\| A^T_c [y - X_\Omega \theta_z] \right\|_2^2 \]
   where \( [A^T_c]_i := \alpha_i(z) \) and zero else, \( D_\Omega := \{ f(z), \theta_z, \text{vech}(H_z) \} \), and
   \[ X_\Omega := \begin{bmatrix} 1 (x_i - z)^T \text{vech}(z (x_i - z)) \\ 1 (x_i - z)^T \text{vech}(x_i - z) \end{bmatrix} \]
   The notation vech(\( D \)) is the superoperator of half of the symmetric matrix \( D \).

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

   \[ \hat{f}(z) = \sum_{i \in \Omega} \alpha_i(z) y_i \sum_{i \in \Omega} \alpha_i(z) \]

   Taking only the first column of \( X_\Omega \) gives a locally constant estimate of \( f(z) \):

   \[ \hat{f}(z) = (1^T A_\Omega) A_y = \sum_{i \in \Omega} \alpha_i(z) y_i \]

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

How SPARROW Defines the Observation Weights

We construct a dictionary matrix by concatenating normalized regressors
\[ D := \begin{bmatrix} x_1^T \\ \vdots \\ x_N^T \end{bmatrix} \]

For a given point \( z \), SPARROW finds a solution to \( z \approx Ds \) such that \( s \) has many zero elements by solving the basis pursuit denoising (BPDN) problem
\[ \min_{s} |s|_1 \text{ subject to } \frac{\|z - Ds\|_2^2}{\|z\|_2^2} \leq c^2 \]
where \( c^2 > 0 \). 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 ith observation weight by
\[ \alpha_i(z) = \frac{\left\| z - x_i \right\|^2}{\sum_i \left\| z - x_i \right\|^2} s_i \]
where \( s_i \) is the ith element of \( s, i \in \Omega \).

Experiments and Simulations

The figures below compare locally-constant SPARROW (C-SPAR) and other methods. We use 100 independent trials of 10-fold cross-validation to estimate the mean squared error (MSE). Red lines mark median. Boxes delimit 25 to 75 percentiles. Extrema marked by whiskers, and outliers by pluses.

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

Above we see the ability of local regression methods to model data locally.