Asymptotics of Over-Parameterized Polynomial Models

Nathan Wycoff

March 30, 2019

1 Notation

N	Number of obs
d	Degree of polynomial
$\mathbf{x}$	Locations of observations (1D problem).
$\mathbf{A}(\mathbf{x})$	polynomial terms evaluated at each $\mathbf{X}$ matrix, an $N \times D$ matrix.
$\mathbf{y}$	Observations

2 Interpolation with Minimal Derivative Norm

Norm minimization with respect to the coordinates, which is often the way things are done in machine learning, is somewhat arbitrary for our case as they depend on the basis chosen. It happens to be that the norm of the Legendre basis coordinates of a polynomial also gives something like the ℓ_2 norm of that polynomial (ask anthony).

So we are confronted with the following optimization problem:

$$\min_{\mathbf{c}} ||\mathbf{c}||_2$$
$$S.T.$$
$$\mathbf{A}(\mathbf{x})\mathbf{c} = \mathbf{y}$$

A simple application of Lagrangian duality reveals that the optimum is given by $\mathbf{c} = (\mathbf{D}^\top\mathbf{D})^\dagger\mathbf{A}^\top(\mathbf{A}(\mathbf{D}^\top\mathbf{D})^\dagger\mathbf{A}^\top)^\dagger\mathbf{y}$.