Data Mining with Spare Grids

Seminar: Computational Aspects of Machine Learning

Sebastian Kreisel
October 1, 2015

Technische Universität München

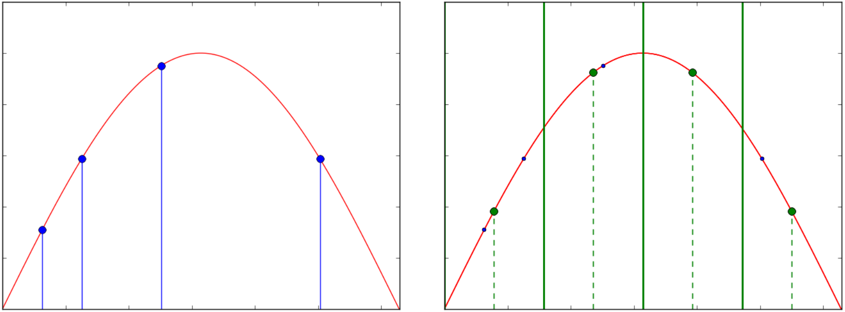
Overview

- Motivation for Sparse Grids
- Sparse Grids: Basics
- Sparse Grids: Machine Learning
- Examples with Data Sets
- Parallelization and Implementation

Motivation for Sparse Grids

Grid based approaches in ML

- Discretizes the space into a grid
- Basis–functions around grid points, not data points



Point based Grid based

Motivation for Sparse Grids

Suitable for

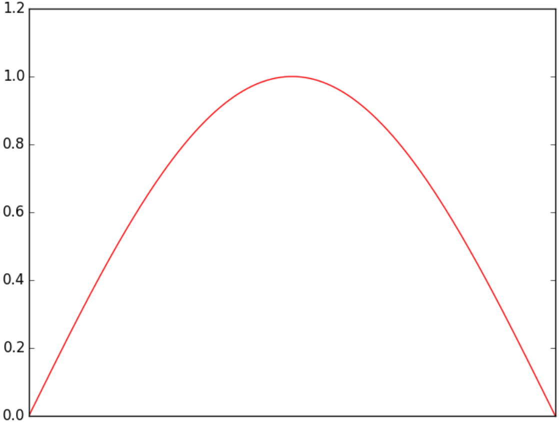
- Big datasets
- Easily/automatically classifiable data
- Medical, seismic, commercial data

Curse of dimensionality

- The volume of a space is exponential in it's dimensions
- The amount of training data required becomes unmanageable
  - because of lacking computational/storage capacities
  - because data aquesition is expensive
- Becomes relevant for $d > 3$
- **Applies to full-grid discretization**

Full Grid Discretization

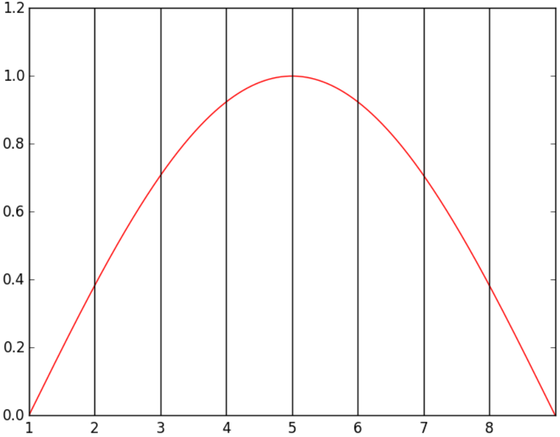
1. A function to interpolate



$$f(x) = sin(\pi \cdot x)$$

Full Grid Discretization

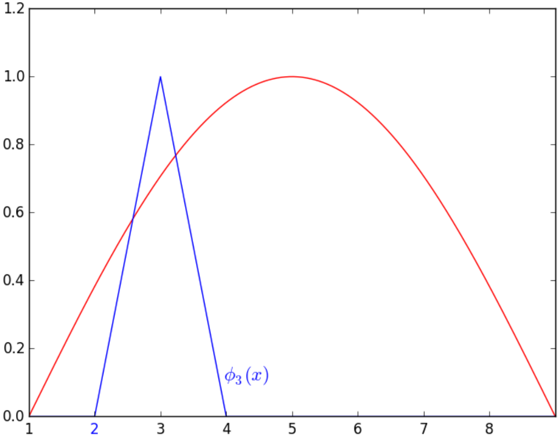
2. A (full, regular) grid



Eight centered gridpoints $i \in \{1..8\}$

Full Grid Discretization

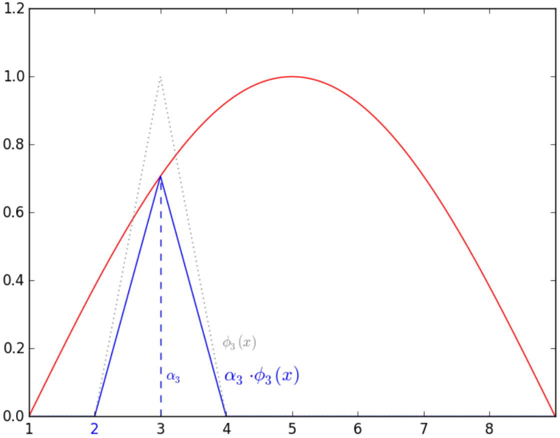
3. A basis function (standard hat function)



$\phi(x-3)$ with $\phi(x) = \max(abs(1-x), 0)$

Full Grid Discretization

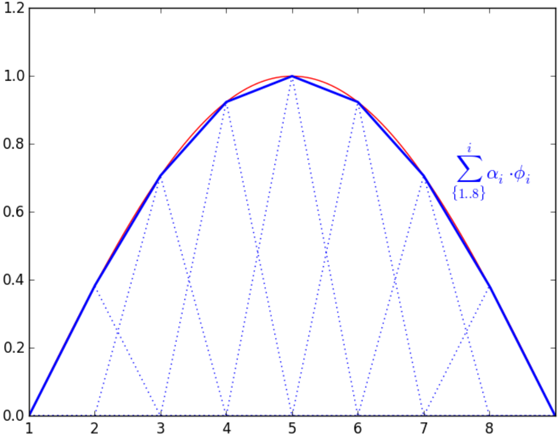
4. A Coefficient (Surplus)



$\alpha_2 \cdot \phi_2(x)$

Full Grid Discretization

Sum over all basis functions



$$f(x) \approx \hat{f}(x) = \sum_i \alpha_i \cdot \phi_i$$

Full Grid Discretization

Full grid interpolation in one dimension

1. A function $f(x)$
2. Gridpoints indexed by $i \in \{1, 2, \dots\}$
3. Basis/Ansatz functions; i.e. hat function $\phi_i(x) = \max(1 - |x|, 0)$
4. Coefficients α_i (hierachcial surplusses)

$$f(x) \approx \hat{f}(x) = \sum_i \alpha_i \phi_i(x)$$

Full Grid Discretization

Level of detail

- Level $l \in \{1, 2, \dots\}$ in addition to index $i \in 2^l$

D dimensions

- For each dimension d: $\hat{f}_d$
- Tensor product over all dimensions $\hat{f}(x) = \prod_j \hat{f}_j(x)$

Sparse Grids – Basics

Hirachial Basis

- Grouping gridpoints into levels $l_1, l_2, \dots$
- For each level a set of gridpoints $I_l = \{\ i \ \mid\ 1 < i < 2^l - 1;\ \ i \text{ odd}\ \}$
  - $l_1 = \{1\}$
  - $l_2 = \{1, 3\}$
  - $l_3 = \{1, 3, 5, 7\}$
  - $\dots$
- For all dimensions seperately

$$\hat{f}_d(x) = \sum_{l,i}\ \alpha_{l,i}\ \cdot\ \phi_{l,i}(x)$$