Andrew Nystrom, Research Scientist at Contata Solutions
John Hughes, Professor of Computer Science at Brown University

Efficient Calculation of Second Degree Polynomial Features on Sparse Matrices

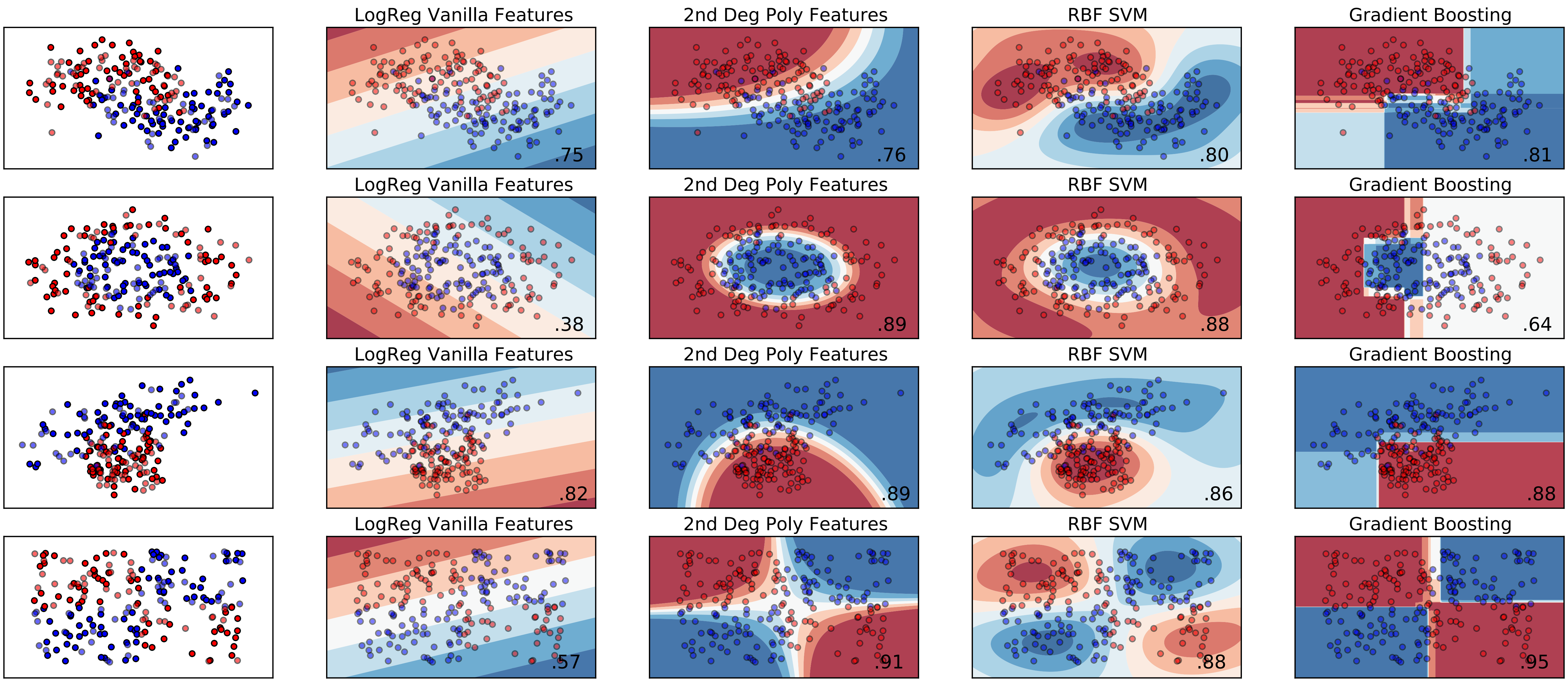
Handling Nonlinearity

- Classify in kernel space, e.g. rbf kernel
- Nonlinear Classifiers, e.g. gradient boosting
- Apply an input transformation

Polynomial Features

- Products of combinations of features
  - $\{x_i \cdot x_j : i,j \in \{0,1,\ldots,D-1\} \wedge i \leq j\}$
- [a, b, c] has polynomial features [a^2, ab, ac, b^2, bc, c^2]
- D features, $(D^2+D)/2$ 2nd degree polynomial features

What they can do (graphically)



Calculation

- Naive Calculation
  - Walk through all pairs, take products
  - Zero whenever either inputs are zero
  - Unnecessary calculations when matrix is sparse
- Smarter Calculation
  - Take products of non-zero elements
  - Problem: knowing output column index given input indices

Column Output Location

$\{x_i \cdot x_j : i, j \in \{0, 1, ..., D-1\} \wedge i \leq j\}$

$\{(i, j) : i, j \in \{0, 1, ..., D-1\} \wedge i \leq j\} \rightarrow \{0, 1, ..., \frac{D^2+D}{2} - 1\}$

$$
\begin{array}{ccc}
 & & j \\
 & \vec{x} & \begin{array}{ccccc} x_0 & x_1 & x_2 & x_3 & x_4 \end{array} \\
i & \begin{array}{c} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} & \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ - & 5 & 6 & 7 & 8 \\ - & - & 9 & 10 & 11 \\ - & - & - & 12 & 13 \\ - & - & - & - & 14 \end{pmatrix}
\end{array}
$$

$$\text{polynomial-index}(i, j | D) = \frac{2Di - i^2 + 2j - 3i - 2}{2} + i + 1$$

Algorithm Comparison

DENSE INTERACTION(A)

N = row count of A
D = column count of A
B = Matrix of size $N \times \frac{D^2-D}{2}$
for *row* in A
  for $\imath = 0$ **to** $D-1$
    for $\jmath = i+1$ **to** D
      $B[i,j] = row[i] \cdot row[j]$

O(D²), Θ(D²)

SPARSE POLYNOMIALS(A)

$\mathrm{map}(a,b) = \frac{2Da-a^2+2b-3a-2}{2} + a + 1$
N = row count of A
D = column count of A
B = Compressed Sparse Row Matrix of size $N \times \frac{D^2+D}{2}$
for *row* in A
  N_{zc} = nonzero columns of *row*
  for $i = 0$ **to** $|N_{zc}|$
    for $j = i$ **to** $|N_{zc}|$
      $k = \mathrm{map}(i,j)$
      r = index of *row*
      $B[r,k] = row[i] \cdot row[j]$

$O(D^2)$, $\Theta(d^2D^2)$
where d is density

Average complexities decrease quadratically w.r.t. density!

Real Usage

dataset	instances	features	polynomial features	density	dense space	sparse space	dense time	sparse time	performance boost
20 Newsgroups	11,314	130,107	8,463,980,778	0.12					
Connect Four	67,557	126	8,001	0.11					