Sequential Event Prediction with Association Rules

Cynthia Rudin
Ben Letham
Ansaf Salleb-Aouissi
Gene Kogan
David Madigan
Welcome to Our Fruit Stand

CUT RIGHT NOW
Each day we recommend a few items that will be great for delivery tomorrow.
LEARN ABOUT OUR DAILY QUALITY RATINGS — CLICK HERE

- **Sweet 2 Eat Peaches** $2.99/lb
- **Hass Avocados, Ready-to-Eat Pack** $4.49/2pk
- **Golden Pineapple** $3.99/ea
- **Green Seedless Grapes** $3.49/lb (was $3.99)
- **Kiwi** $0.69/ea 5 for $3.00

OUR FRESH FRUIT

- Apples
- Bananas
- Citrus
- Grapes
- Pears
- Melons
- Berries & Cherries
- Stone Fruit
- Tropical & Specialty
- Cut Fruit & Fruit Salad
- Fresh-Squeezed Juice
- Avocados
- Dried Fruit & Nuts
SWEET 2 EAT PEACHES

FARM FRESH $2.99 /lb
Med

Organic $3.49 /lb
Med

About | Nutrition
You have to respect a fruit with the power to change the seasons. The sweet, full flavor of a ripe peach insists that it’s summer, even if it’s February and you’re stranded in Greenland. Everything about the eating experience — the enticing smell, the juicy flesh, even the downy fuzz — makes you understand what “peachy” really means.

Plan ahead! Our peaches arrive firm. They’ll soften in 3-5 days at room temperature, or you can speed up the process to 2-3 days by storing them in our complimentary fruit ripening bag. When your peaches yield to gentle pressure, they’re ready.

Peach Season: Summer

Estimated 0.67 lb $1.99
YOUR CART

Please review the items in your cart before going to Checkout.

Welcome to FreshDirect

UPDATE CART

FRUIT

Sweet 2 Eat Peaches (Farm Fresh, Med) ($2.99/lb) $1.99*

Estimated Subtotal: $1.99*

Order Subtotal: $1.99*

ESTIMATED TOTAL: $1.99*

Enter promotion code: 

*NEW Have a FreshDirect Gift Card? Enter it here.

T = Taxable Item  S = Special Price  D = State Bottle Deposit

FRESHDIRECT FAVORITES

Horizon Organic 2% Milk 1/2 gallon, 64oz $4.29/ea

Cal-Organic Organic Baby Carrots 1 lb $2.49/ea

Boneless Skinless Chicken Breast, Raised Without Antibiotics 7-8 pc/pk, No Antibiotics Used, Family Pack $4.39/lb

Crispy Chicken Fingers 10 oz $5.99/3pcs

Blueberries Farm Fresh, 1 pint $2.99/pkg (was $4.99)

PEAK SEASON PRODUCE

DON'T-MISS DEALS

These are just a few of our most-loved products.
<table>
<thead>
<tr>
<th>Item</th>
<th>Quantity</th>
<th>Price</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earthbound Farm Organic Baby Spinach</td>
<td>1</td>
<td>$3.99</td>
<td>Remove</td>
</tr>
<tr>
<td>Grimmway Baby Carrots</td>
<td>2</td>
<td>$3.50</td>
<td>Remove</td>
</tr>
<tr>
<td>Kirby Cucumber</td>
<td>2</td>
<td>$0.85</td>
<td>Remove</td>
</tr>
<tr>
<td>Red Bell Peppers</td>
<td>1</td>
<td>$1.49</td>
<td>Remove</td>
</tr>
<tr>
<td>Savoy Cabbage</td>
<td>1</td>
<td>$3.90</td>
<td>Remove</td>
</tr>
<tr>
<td>White Cauliflower</td>
<td>1</td>
<td>$3.90</td>
<td>Remove</td>
</tr>
</tbody>
</table>

**Estimated Subtotal:** $55.00**

**Tax:** $0.71

**State Bottle Deposit:** $0.60

**Order Subtotal:** $55.86**

**Total Taxi:** $0.71

**State Bottle Deposit:** $0.60

**Delivery Charge:** FREE with DeliveryPass

**Fuel Surcharge (waived):** $0.00

**ESTIMATED TOTAL:** $57.19**

*NEW Have a FreshDirect Gift Card? Enter it here.

**YOUR FAVORITES**

- Gerber Graduates for Toddlers
- Cheese Pasta
- Chicken & Vegetables

**RECOMMENDED FOR YOU**

- Gerber Graduates for Toddlers
- Light in Tomato Sauce with Carrots
- Peas & Carrots
- Vegetarian Split Pea Soup

**DON'T-MISS DEALS**

- Danzoni Light 2% Frozen Raspberry Yogurt
- Any 4 flavors for $3.40
- Any 2 flavors for $5.40

Continue Checkout:
I’ve ordered chicken and sesame 15 times, and I’ve ordered lemon 13 out of 15 of those times.

Why isn’t it recommending me lemon?
Top 10 algorithms in data mining

Xindong Wu · Vipin Kumar · J. Ross Quinlan · Joydeep Ghosh · Qiang Yang · Hiroshi Motoda · Geoffrey J. McLachlan · Angus Ng · Bing Liu · Philip S. Yu · Zhi-Hua Zhou · Michael Steinbach · David J. Hand · Dan Steinberg

Received: 9 July 2007 / Revised: 28 September 2007 / Accepted: 8 October 2007
Published online: 4 December 2007
© Springer-Verlag London Limited 2007

Abstract This paper presents the top 10 data mining algorithms identified by the IEEE International Conference on Data Mining (ICDM) in December 2006: C4.5, k-Means, SVM, Apriori, EM, PageRank, AdaBoost, kNN, Naive Bayes, and CART. These top 10 algorithms are among the most influential data mining algorithms in the research community. With each algorithm, we provide a description of the algorithm, discuss the impact of the algorithm, and review current and further research on the algorithm. These 10 algorithms cover classification,

R. Agrawal; T. Imielinski; A. Swami: Mining Association Rules Between Sets of Items in Large Databases", SIGMOD Conference 1993: 207-216
I’ve ordered chicken and sesame 15 times, and I’ve ordered lemon 13 out of 15 of those times.

Sup (chicken ∪ sesame) = 15
Sup (chicken ∪ sesame ∪ lemon) = 13
I’ve ordered chicken and sesame 15 times, and I’ve ordered lemon 13 out of 15 of those times.

\[
\text{Sup (chicken } \cup \text{ sesame)} = 15 \\
\text{Sup (chicken } \cup \text{ sesame } \cup \text{ lemon)} = 13 \\
\text{Conf ((chicken } \cup \text{ sesame) } \Rightarrow \text{ lemon)} = \frac{13}{15}
\]
I’ve ordered chicken and sesame 15 times, and I’ve ordered lemon 13 out of 15 of those times.

\[
\text{Sup (chicken } \cup \text{ sesame}) = 15
\]

\[
\text{Conf ((chicken } \cup \text{ sesame) } \Rightarrow \text{ lemon}) = 13/15
\]
**“Max Confidence, Min Support” Algorithm**

**Step 1.** Find all rules $a \Rightarrow b$, where $a$ is in an “allowed” set $A$, and $\text{Sup}(a) \geq \theta$.

**Step 2.** Rank rules in descending order of $\text{Conf}(a \Rightarrow b)$, recommend the right hand sides of the top ranked rules.

<table>
<thead>
<tr>
<th>Support($a$)</th>
<th>Confidence($a \Rightarrow b$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>13/15 = 0.867</td>
</tr>
<tr>
<td>25</td>
<td>20/25 = 0.8</td>
</tr>
<tr>
<td>17</td>
<td>12/17 = 0.706</td>
</tr>
<tr>
<td>50</td>
<td>34/50 = 0.68</td>
</tr>
</tbody>
</table>
About Max Conf, Min Support Algorithm

- Min support threshold allows generalization
About Max Conf, Min Support Algorithm

- Min support threshold allows generalization

- ...But excludes nuggets (low support, high conf)
About Max Conf, Min Support Algorithm

• Min support threshold allows generalization

• ...But excludes nuggets (low support, high conf)

• Not thrilled with ranking rules by confidence, ignoring the support.

Conf=.99, Support=10000 vs. Conf=1, Support=10
Try “Adjusted Confidence” Instead

AdjustedConf(a → b) = Sup(a ∪ b)/(Sup(a)+K)

Bayesian estimate of the confidence
(AdjustedConf=Conf when K=0)
**“Adjusted Confidence” Algorithm**

**Step 1.** Find all rules \( a \Rightarrow b \), where \( a \) is in an “allowed” set \( A \).

**Step 2.** Rank rules in descending order of \( \text{AdjustedConf}(a \Rightarrow b) \), recommend the right hand sides of the top ranked rules.

<table>
<thead>
<tr>
<th>Support((a))</th>
<th>( \text{AdjConf}(a \Rightarrow b), K=5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>( 20/(25+5)=.67 )</td>
</tr>
<tr>
<td>15</td>
<td>( 13/(15+5)=.65 )</td>
</tr>
<tr>
<td>50</td>
<td>( 34/(50+5)=.62 )</td>
</tr>
<tr>
<td>17</td>
<td>( 12/(17+5)=.55 )</td>
</tr>
</tbody>
</table>
Try “Adjusted Confidence” Instead

AdjustedConf(a → b) = Sup(a U b)/(Sup(a)+K)

• Nuggets can be used

• Among rules with similar confidence, prefers rules with higher support

• K encourages larger support, allows to generalize

Conf=.99, Support=10000 vs. Conf=1, Support=10
This work:

• Develops a framework for using association rules in supervised learning
  – small sample bounds, not just large sample
This work:

• Develops a framework for using association rules in supervised learning
  
  – small sample bounds, not just large sample

• Formalizes the problem of sequential event prediction ("the online grocery store problem")
This work:

• Develops a framework for using association rules in supervised learning
  – small sample bounds, not just large sample

• Formalizes the problem of sequential event prediction ("the online grocery store problem")

• Tries to determine what the important quantities in the learning process are:
  \[ m, |A|, p_{\text{min}}, K \text{ or } \theta \]

- size of training set
- size of allowed set
- probability of least probable items or itemsets
- min supp thresh
- adj. conf parameter
Max Conf, Min Supp Algorithm has stronger guarantees than Adjusted Conf Algorithm (Bias-Variance tradeoff)
Outline of paper (long version)

• Define learning problems

• Large sample bounds
  – VC bound (for classification only)
  – pointwise stability bounds

• Small sample bounds
  – uniform stability bound for max conf, min sup
  – pointwise stability bounds for adjusted conf

• Experiments
Outline of paper (long version)

• Define learning problems
  seq. pred. only

• Large sample bounds
  – VC bound (for classification only)
  – pointwise stability bounds (soon)

• Small sample bounds
  – uniform stability bound for max conf, min sup
  – pointwise stability bounds for adjusted conf

• Experiments
$X = \{\text{set of all items}\}$

$z \in 2^X \times \Pi$  \hspace{1cm} \text{Ordered basket of items}

$z_{.,t} = t^{th}$ item in basket $z$

$T_z = |z|$  \hspace{1cm} \text{Number of items in basket}

$S = \{z_i\}_{i=1 \ldots m}$  \hspace{1cm} \text{Training set, m random baskets}

$f_{S,K}(a,b) = \frac{\text{Sup}(a \cup b)}{\text{Sup}(a) + K}$  \hspace{1cm} \text{Adjusted Confidence of}  \

\hspace{1cm} a \rightarrow b$
Highest-scoring-correct rule \( a^+_{SztK} \rightarrow z_{..,t+1} \)

\[
a^+_{SztK} \in \arg\max_{a \in \{z_{..,1} \ldots z_{..,t}\}, a \in A} f_{S,K}(a, z_{..,t+1})
\]
Highest-scoring-correct rule $a_{SztK}^+ \rightarrow z_{.,t+1}$

$$a_{SztK}^+ \in \arg \max_{a \in \{z_{.,1} \ldots z_{.,t}\}, a \in A} f_{S,K}(a, z_{.,t+1})$$

Highest-scoring-incorrect rule $a_{SztK}^- \rightarrow b_{SztK}^-$

$$[a_{SztK}^-, b_{SztK}^-] \in \arg \max_{a \in \{z_{.,1} \ldots z_{.,t}\}, a \in A, b \in X \backslash \{z_{.,1} \ldots z_{.,t+1}\}} f_{S,K}(a, b)$$

This rule can’t recommend $z_{t+1}$
Highest-scoring-correct rule $a_{SztK}^+ \rightarrow z_{.,t+1}$

$$a_{SztK}^+ \in \arg\max_{a \in \{z_{.,1} \ldots z_{.,t}\}, a \in A} f_{S,K}(a, z_{.,t+1})$$

Highest-scoring-incorrect rule $a_{SztK}^- \rightarrow b_{SztK}^-$

$$[a_{SztK}^-, b_{SztK}^-] \in \arg\max_{a \in \{z_{.,1} \ldots z_{.,t}\}, a \in A, b \in X \setminus \{z_{.,1} \ldots z_{.,t+1}\}} f_{S,K}(a, b)$$

0-1 Loss function

$$l_{0-1,Kr}(f_{S,K}, z) := \frac{1}{T_z} \sum_{t=0}^{T_z-1} \mathbb{1}[f_{S,K_r}(a_{SztK}^+, z_{.,t+1}) \leq f_{S,K_r}(a_{SztK}^-, b_{SztK}^-)]$$

Lose a point if score of incorrect rule > score of correct rule

Loss Function has separate $K$ than algorithm
Highest-scoring-correct rule \( a_{SztK}^+ \to z_{.,t+1} \)

\[
a_{SztK}^+ \in \arg\max_{a \in \{z_{.,1} \ldots z_{.,t}\}, a \in A} f_{S,K}(a, z_{.,t+1})
\]

Highest-scoring-incorrect rule \( a_{SztK}^- \to b_{SztK}^- \)

\[
[a_{SztK}^-, b_{SztK}^-] \in \arg\max_{a \in \{z_{.,1} \ldots z_{.,t}\}, a \in A}
\quad b \in X \setminus \{z_{.,1} \ldots z_{.,t+1}\}
\]

0-1 Loss function

\[
l_{0-1,K_r}(f_{S,K}, z) := \frac{1}{T_z} \sum_{t=0}^{T_z-1} \mathbb{1}[f_{S,K_r}(a_{SztK}^+, z_{.,t+1}) \leq f_{S,K_r}(a_{SztK}^-, b_{SztK}^-)]
\]

True Error

We’ll upper bound TrueErr

\[
\text{TrueErr}(f_{S,K}, K_r) := \mathbb{E}_{z \sim D} l_{0-1,K_r}(f_{S,K}, z)
\]
0-1 Loss function

\[ l_{0-1, K_r}(f_{S,K}, z) := \frac{1}{T_z} \sum_{t=0}^{T_z-1} 1[f_{S,K}(a_{SztK}^{+}, z_{t+1}) \leq f_{S,K}(a_{SztK}^{-}, b_{SztK}^{-})] \]
0-1 Loss function

\[ l_{0-1,K_r}(f_{S,K}, z) := \frac{1}{T_z} \sum_{t=0}^{T_z-1} 1[f_{S,K_r}(a^{+}_{SztK}, z_{t+1}) \leq f_{S,K_r}(a^{-}_{SztK}, b^{-}_{SztK})] \]
0-1 Loss function

\[ l_{0-1,K_r}(f_{S,K},z) := \frac{1}{T_z} \sum_{t=0}^{T_z-1} \mathbf{1}[f_{S,K_r}(a_{SztK}^+,z.t+1) \leq f_{S,K_r}(a_{SztK}^-,b_{SztK}^-)] \]

Continuous loss function

\[ l_{\gamma,K_r}(f_{S,K},z) := \frac{1}{T_z} \sum_{t=0}^{T_z-1} c_{\gamma}(f_{S,K}(a_{SztK}^+,z.t+1) - f_{S,K}(a_{SztK}^-,b_{SztK}^-)) \]

Uses differences in adjusted confidence

\[ c_{\gamma}(x) \]

1

1 - x/\gamma

0
Continuous loss function

\[ l_{\gamma, K_r}(f_S, K, z) := \frac{1}{T_z} \sum_{t=0}^{T_z-1} c_\gamma(f_{S, K}(a_{SztK}^+, z_{t+1}) - f_{S, K}(a_{SztK}^-, b_{SztK}^-)) \]

Empirical Error

\[ \text{EmpErr}_{\gamma}(f_{S, K}, K_r) := \frac{1}{m} \sum_{i=1}^{m} l_{\gamma, K_r}(f_{S, K}, z_i) \]
Outline of paper (long version)

• Define learning problems ✔

• Large sample bounds
  – VC bound (for classification only)
  – pointwise stability bounds ← seq. pred. only

• Small sample bounds
  – uniform stability bound for max conf, min sup ← seq.
  – pointwise stability bounds for adjusted conf ← pred.
  only

Stability bounds use tools of:
  Bousquet & Elisseeff 2002, Devroye & Wagner 1979
Theorem (*Large sample, Adj. Conf., Seq. Event Pred.*)

For set of rules $A$, $K \geq 0$, $K_r \geq 0$, with prob. $1 - \delta$ (w.r.t $S \sim D^m$),

$$\text{TrueErr}(f_{S,K}, K_r) \leq \text{EmpErr}_\gamma(f_{S,K}, K_r) + \sqrt{\frac{1}{\delta} \left[ \frac{1}{2m} + 6\beta \right]}$$
Theorem (*Large sample, Adj. Conf., Seq. Event Pred.*)

For set of rules $A$, $K \geq 0$, $K_r \geq 0$, with prob. $1 - \delta$ (w.r.t $S \sim D^m$),

$$\text{TrueErr}(f_{S,K},K_r) \leq \text{EmpErr}_\gamma(f_{S,K},K_r) + \sqrt{\frac{1}{\delta} \left[ \frac{1}{2m} + 6\beta \right]}$$

where $\beta = \frac{2|\mathcal{A}|}{\gamma} \left[ \frac{1}{(m-1)p_{\min A} + K} + \frac{|K - K_r|}{m + K} \right] + O\left( \frac{1}{m^2} \right)$

and where $\mathcal{A} = \{a \in A : P_Z(a \subseteq z) > 0\}$

and $p_{\min A} := \min_{a \in \mathcal{A}} P_{Z \sim D}(a \subseteq z)$

probable itemsets

probability of least probable itemset
where $\beta = \frac{2|\mathcal{A}|}{\gamma} \left[ \frac{1}{(m-1)p_{\min A} + K} + \frac{|K - K_r|}{m + K} \right] + O\left(\frac{1}{m^2}\right)$
where \( \beta = \frac{2|\mathcal{A}|}{\gamma} \left[ \frac{1}{(m - 1)p_{\text{min}_A} + K} + \frac{|K - K_r|}{(m - 1)p_{\text{min}_A} + K_r} \cdot \frac{m}{m + K} \right] + O\left( \frac{1}{m^2} \right) \)

and where \( \mathcal{A} = \{ a \in A : P_z(a \subseteq z) > 0 \} \quad \text{probable itemsets} \)

\[ p_{\text{min}_A} := \min_{a \in \mathcal{A}} P_{z \sim D}(a \subseteq z) \quad \text{probability of least probable itemset} \]
where $\beta = \frac{2|\mathcal{A}|}{\gamma} \left[ \frac{1}{(m-1)p_{\text{min}}A + K} + \frac{|K - K_r|}{m + K} \right] + O\left(\frac{1}{m^2}\right)$

Vanishes when $K = K_r$

$K$ assists
when $K=K_r=0...$
Corollary (*Large sample, Adj. Conf., Seq. Event Pred.*)

With prob. $1 - \delta$ (w.r.t $S \sim D^m$),

$$\text{TrueErr}(f_{S,K}, 0) \leq \text{EmpErr}_\gamma(f_{S,K}, 0) + \sqrt{\frac{1}{\delta} \left[ \frac{1}{2m} + \frac{12|A|}{\gamma(m-1)p_{\min A}} + O\left(\frac{1}{m^2}\right) \right]}$$
Outline of paper (long version)

• Define learning problems ✔

• Large sample bounds
  – VC bound (for classification only)
  – pointwise stability bounds ✔

• Small sample bounds
  – uniform stability bound for max conf, min sup
  – pointwise stability bounds for adjusted conf ➡️

  seq. pred.

  only
Theorem *(Small sample, Adj. Conf., Seq. Event Pred.)*

For set of rules $A$, $K \geq 0$, $K_r \geq 0$, with prob. $1 - \delta$ (w.r.t $S \sim D^m$),

\[ \text{TrueErr}^*_\gamma(f_{S,K},K_r) \leq \text{EmpErr}^*_\gamma(f_{S,K},K_r) + \sqrt{\frac{1}{\delta} \left[ \frac{1}{2m} + 6\beta \right]} \]

where $\beta = \frac{2}{\gamma K} \left( 1 - \frac{(m-1)p_{\min}}{m+K} \right)$

\[ + \frac{2}{\gamma} |K_r - K| \mathbb{E}_{\zeta \in \text{Bin}(m-1,p_{\min})} \left[ \frac{1}{K \left( \frac{\zeta}{m+K-\zeta-1} \right)} + K_r \left( \frac{m}{m+K} + \frac{1}{K} \left( 1 - \frac{\zeta}{m+K} \right) \right) \right] \]

and where $Q = \{ x \in X : P_{z \sim D}(x \in z) > 0 \}$

$p_{\min} := \min_{x \in Q} P_{z \sim D}(x \in z)$
Theorem (Small sample, Adj. Conf., Seq. Event Pred.)

For set of rules $A$, $K \geq 0$, $K_r \geq 0$, with prob. $1 - \delta$ (w.r.t $S \sim D^m$),

$$\text{TrueErr}^*_\gamma(f_{S,K},K_r) \leq \text{EmpErr}^*_\gamma(f_{S,K},K_r) + \sqrt{\frac{1}{\delta} \left[ \frac{1}{2m} + 6\beta \right]}$$

where $\beta = \frac{2}{\gamma K} \left( 1 - \frac{(m-1)p_{\min}}{m+K} \right)$

$$+ \frac{2}{\gamma} |K_r - K| E_{\zeta \in \text{Bin}(m-1,p_{\min})} \left[ \frac{1}{K \left( \frac{\zeta}{m+K - \zeta - 1} \right) + K_r} \left( \frac{m}{m+K} + \frac{1}{K} \left( 1 - \frac{\zeta}{m+K} \right) \right) \right]$$

and where $Q = \left\{ x \in X : P_{z \sim D}(x \in z) > 0 \right\}$ probable items

$p_{\min} := \min_{x \in Q} P_{z \sim D}(x \in z)$ probability of least probable item
where $\beta = \frac{2}{\gamma K} \left( 1 - \frac{(m-1)p_{\text{min}}}{m+K} \right)$

$$+ \frac{2}{\gamma} |K_r - K| \mathbb{E}_{\zeta \in \text{Bin}(m-1, p_{\text{min}})} \left[ \frac{1}{K} \left( \frac{\zeta}{m+K-\zeta-1} \right) + K_r \left( \frac{m}{m+K} + \frac{1}{K} \left( 1 - \frac{\zeta}{m+K} \right) \right) \right]$$

$K$ assists

Vanishes when $K=K_r$

Does not depend on $|A|$
Assuming $m \gg K \gg 0$, $\zeta \approx mp_{\text{min}}$, 

$$\beta \approx \beta_{\text{Approx}} := \frac{2}{\gamma K} \left(1 - \frac{(m-1)p_{\text{min}}}{m + K}\right) + \frac{2}{\gamma} \left|K_r - K\right| \frac{1}{K \frac{p_{\text{min}}}{1 - p_{\text{min}}} + K_r}$$

$\beta$ and $\beta_{\text{Approx}}$
Outline of paper (long version)

• Define learning problems ✔

• Large sample bounds
  – VC bound (for classification only)
  – pointwise stability bounds ✔

• Small sample bounds
  – uniform stability bound for max conf, min sup
  – pointwise stability bounds for adjusted conf ✔
Summary of Bounds

• Large sample bounds and small sample bounds
• Adjusted confidence provides a weaker support threshold, allowing “nuggets” while still being able to generalize
• Highlighted quantities that may be important for learning with rules

\[ |A|, K \text{ or } \theta, p_{\text{min}A}, p_{\text{min}} \text{ or } p_{\text{min},y} \]
Current Work

• Bayesian hierarchical modeling for association rules and sequential events
  – predicting symptoms of medical patients in a clinical trial
  – (with Tyler McCormick and David Madigan)

• Supervised ranking for sequential event prediction
  – (with Ben Letham and David Madigan)

• New approaches to mining for rules
  – (with Allison Chang and Dimitris Bertsimas)
Thank you!
Outline of paper (long version)

• Define learning problems (soon)

• Large sample bounds
  – VC bound (for classification)
  – pointwise stability bounds

• Small sample bounds
  – uniform stability bound for max conf, min sup
  – pointwise stability bounds for adjusted conf
Outline of paper (long version)

• Define learning problems

• Large sample bounds
  – VC bound (for classification) \[ |A| \]
  – pointwise stability bounds \[ |A|, K, p_{\text{min}A} \]

• Small sample bounds
  – uniform stability bound for max conf, min sup \[ \theta \]
  – pointwise stability bounds for adjusted conf
    \[ K, \quad p_{\text{min}} \text{ or } p_{\text{min},y} \]
Uniform Bound for Max-Score Classifiers

\[ g : A \times \{-1,1\} \rightarrow \mathbb{R} \quad \text{scoring function for } a \rightarrow y \]

\[ G := \{g\} \quad \text{set of all scoring functions} \]

\[ f_g(x) = \arg \max_{y \in \{-1,1\}} \max_{a \in A, a \subseteq x} g(a, y) \quad \text{decision function} \]

\[ \mathcal{F}_{\text{maxscore}} := \{f_g : g \in G\} \quad \text{all decision functions} \]
Uniform Bound for Max-Score Classifiers

\[ g : A \times \{-1,1\} \rightarrow R \quad \text{scoring function for } a \rightarrow y \]

\[ G := \{g\} \quad \text{set of all scoring functions} \]

\[ f_g(x) = \arg \max_{y \in \{-1,1\}} \max_{a \in A, a \subseteq x} g(a, y) \quad \text{decision function} \]

\[ F_{\text{maxscore}} := \{f_g : g \in G\} \quad \text{all decision functions} \]

**Theorem:** \( VC \dim(F_{\text{maxscore}}) = |A| \)