The Lovász-Bregman Divergence and Connections to Rank Aggregation, Clustering, and Web Ranking

Rishabh Iyer     Jeff Bilmes

University of Washington, Seattle

UAI-2013
Outline

1. Ranking and Machine Learning
2. The Lovász-Bregman divergences
3. Properties of the Lovász-Bregman
4. Applications
5. Summary
Combining Scores and Rankings

Occur in a number of Machine Learning applications:


Occur in a number of Machine Learning applications:

Combining Classifiers (Lebanon & Lafferty, 2002)
Combining Scores and Rankings

Occur in a number of Machine Learning applications:

Combining Classifiers (Lebanon & Lafferty, 2002)

Aggregating Preferences (Murphy & Martin, 2003)
Combining Scores and Rankings

Occur in a number of Machine Learning applications:

1. Combining Classifiers (Lebanon & Lafferty, 2002)
2. Aggregating Preferences (Murphy & Martin, 2003)
3. Web Ranking (Liu, 2009)
Denote $\sigma$ as a permutation of $\{1, 2, \cdots, n\}$ such that $\sigma(i)$ denotes the item at rank $i$ and $\sigma^{-1}(i)$ as the rank of item $i$. 

\[ \sigma(1) \quad \sigma(2) \quad \sigma(3) \quad \sigma(4) \quad \sigma(5) \quad \sigma(6) \quad \sigma(7) \quad \sigma(8) \]
Combining Scores and Rankings

- Denote $\sigma$ as a permutation of $\{1, 2, \cdots, n\}$ such that $\sigma(i)$ denotes the item at rank $i$ and $\sigma^{-1}(i)$ as the rank of item $i$.

- Denote $\{\sigma_1, \sigma_2, \ldots, \sigma_k\}$ as a set of $k$ permutations.
Combining Scores and Rankings

- Denote $\sigma$ as a permutation of $\{1, 2, \cdots, n\}$ such that $\sigma(i)$ denotes the item at rank $i$ and $\sigma^{-1}(i)$ as the rank of item $i$.

- Denote $\{\sigma_1, \sigma_2, \ldots, \sigma_k\}$ as a set of $k$ permutations.

- Some important problems concerning rankings:
Combining Scores and Rankings

- Denote $\sigma$ as a permutation of $\{1, 2, \cdots, n\}$ such that $\sigma(i)$ denotes the item at rank $i$ and $\sigma^{-1}(i)$ as the rank of item $i$.

- Denote $\{\sigma_1, \sigma_2, \ldots, \sigma_k\}$ as a set of $k$ permutations.

- Some important problems concerning rankings:
  - **Combining Permutations**: Given permutations $\sigma_1, \sigma_2, \cdots, \sigma_k$, find a representative $\sigma$, which is “close“ to $\sigma_1, \sigma_2, \cdots, \sigma_k$. 
Combining Scores and Rankings

Denote $\sigma$ as a permutation of \{1, 2, \cdots, n\} such that $\sigma(i)$ denotes the item at rank $i$ and $\sigma^{-1}(i)$ as the rank of item $i$.

Denote $\{\sigma_1, \sigma_2, \ldots, \sigma_k\}$ as a set of $k$ permutations.

Some important problems concerning rankings:

1. **Combining Permutations:** Given permutations $\sigma_1, \sigma_2, \cdots, \sigma_k$, find a representative $\sigma$, which is “close“ to $\sigma_1, \sigma_2, \cdots, \sigma_k$.

2. **Combining Scores:** Given a set of score vectors $x_1, x_2, \cdots, x_k$, find a representative $\sigma$, which is “close“ to $x_1, x_2, \cdots, x_k$. 
Combining Scores and Rankings

- Denote $\sigma$ as a permutation of $\{1, 2, \cdots, n\}$ such that $\sigma(i)$ denotes the item at rank $i$ and $\sigma^{-1}(i)$ as the rank of item $i$.

- Denote $\{\sigma_1, \sigma_2, \ldots, \sigma_k\}$ as a set of $k$ permutations.

- Some important problems concerning rankings:
  1. **Combining Permutations:** Given permutations $\sigma_1, \sigma_2, \cdots, \sigma_k$, find a representative $\sigma$, which is “close“ to $\sigma_1, \sigma_2, \cdots, \sigma_k$.
  2. **Combining Scores:** Given a set of score vectors $x_1, x_2, \cdots, x_k$, find a representative $\sigma$, which is “close“ to $x_1, x_2, \cdots, x_k$.
  3. **Clustering:** Cluster the set of permutations $\sigma_1, \sigma_2, \cdots, \sigma_k$ (or equivalently score vectors $x_1, x_2, \cdots, x_k$).
Rank aggregation

- Combine a set of rankings $\sigma_1, \sigma_2, \cdots, \sigma_k$. 
Combine a set of rankings $\sigma_1, \sigma_2, \cdots, \sigma_k$. 
Rank aggregation

- Combine a set of rankings $\sigma_1, \sigma_2, \cdots, \sigma_k$.

- Often done using permutation based distance metrics.
Permutation based Distance Metrics $d(\sigma, \pi)$

- Metric on the space of permutations.
Permutation based Distance Metrics $d(\sigma, \pi)$

- Metric on the space of permutations.
- Kendall $\tau$,

$$d_T(\sigma, \pi) = \sum_{i,j, i<j} I(\sigma^{-1}\pi(i) > \sigma^{-1}\pi(j))$$

and Spearman’s footrule:

$$d_S(\sigma, \pi) = \sum_{i=1}^{n} |\sigma^{-1}(i) - \pi^{-1}(i)|$$
Permutation based Distance Metrics $d(\sigma, \pi)$

- Metric on the space of permutations.
- Kendall $\tau$,
  \[
d_T(\sigma, \pi) = \sum_{i,j,i<j} I(\sigma^{-1}\pi(i) > \sigma^{-1}\pi(j))
  \]
  and Spearman’s footrule:
  \[
d_S(\sigma, \pi) = \sum_{i=1}^{n} |\sigma^{-1}(i) - \pi^{-1}(i)|
  \]
- Invariance with respect to re-orderings – i.e $d(\pi\sigma, \pi\tau) = d(\sigma, \tau)$. 

Iyer & Bilmes, 2013
Permutation based Distance Metrics $d(\sigma, \pi)$

- Metric on the space of permutations.
- Kendall $\tau$,
  \[
d_T(\sigma, \pi) = \sum_{i,j, i<j} I(\sigma^{-1} \pi(i) > \sigma^{-1} \pi(j))
  \]
  and Spearman’s footrule:
  \[
d_S(\sigma, \pi) = \sum_{i=1}^n |\sigma^{-1}(i) - \pi^{-1}(i)|
  \]
- Invariance with respect to re-orderings – i.e. $d(\pi\sigma, \pi\tau) = d(\sigma, \tau)$.
- Given a set of permutations $\sigma_1, \sigma_2, \cdots, \sigma_k$, find a permutation $\sigma$:
  \[
  \sigma = \arg\min_{\pi} \sum_{i=1}^k d(\sigma_i, \pi) \tag{1}
  \]
Score Aggregation

- What if one has scores instead of just the orderings? For example,
Score Aggregation

What if one has scores instead of just the orderings? For example,

1. Combining Classifiers: probability distribution
Score Aggregation

What if one has scores instead of just the orderings? For example,

1. Combining Classifiers: probability distribution
2. Web ranking: Feature functions
Score Aggregation

What if one has scores instead of just the orderings? For example,
1. Combining Classifiers: probability distribution
2. Web ranking: Feature functions

Need to combine score vectors $x_1, x_2, \cdots, x_k$ and find a representative ordering $\sigma$. 
What if one has scores instead of just the orderings? For example,

1. Combining Classifiers: probability distribution
2. Web ranking: Feature functions

Need to combine score vectors $x_1, x_2, \ldots, x_k$ and find a representative ordering $\sigma$.
A natural formulation of this problem is through a score & permutation based divergence.
A natural formulation of this problem is through a score & permutation based divergence. 

Represents distortion between a score $x$ and an ordering $\sigma$. 
Score & permutation based divergence $d(x||\sigma)$

- A natural formulation of this problem is through a score & permutation based divergence.
- Represents distortion between a score $x$ and an ordering $\sigma$.
- Additional notion of ‘confidence‘ of the ordering.
A natural formulation of this problem is through a score & permutation based divergence.

Represents distortion between a score $x$ and an ordering $\sigma$.

Additional notion of ‘confidence‘ of the ordering.

Given a set of scores $x_1, x_2, \cdots, x_k$, find a permutation $\sigma$:

$$\sigma = \arg\min_{\pi} \sum_{i=1}^{k} d(x_i || \pi) \quad (2)$$
This Talk!

Lovász-Bregman Divergences

Web Ranking

Lovász-Bregman Divergences

Rank Aggregation
Given a differentiable convex function \( \phi \), define (Bregman, 1967):

\[
d_{\phi}(x, y) = \phi(x) - \phi(y) - \langle \nabla \phi(y), x - y \rangle.
\]
Bregman Divergences

- Given a differentiable convex function $\phi$, define (Bregman, 1967):

$$d_\phi(x, y) = \phi(x) - \phi(y) - \langle \nabla \phi(y), x - y \rangle.$$  

- Occur naturally in many machine learning applications:
Given a differentiable convex function $\phi$, define (Bregman, 1967):

$$d_\phi(x, y) = \phi(x) - \phi(y) - \langle \nabla \phi(y), x - y \rangle.$$
Submodular functions: special class of set functions.

\[ f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B \] (3)
Submodular Set functions

- Submodular functions: special class of set functions.

\[ f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B \]  

\( f(A \cup v) - f(A) \geq f(B \cup v) - f(B) \text{, if } A \subseteq B \) (3)
Submodular functions: special class of set functions.

\[ f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B \]  \hspace{1cm} (3)

Gain = 1
Submodular functions: special class of set functions.

\[ f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B \]  

Gain = 1

Gain = 0
Submodular Set functions

- Submodular functions: special class of set functions.

\[ f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B \]  

(3)

- Gain = 1

- Gain = 0

- Admit a natural convex extension, called the Lovász extension!
Lovász extension of a submodular function (Lovász, 1983)

- Given a vector $y$, define permutation $\sigma_y$ that “sorts” $y$, in that: $y[\sigma_y(1)] \geq \cdots \geq y[\sigma_y(n)]$. 

Iyer & Bilmes, 2013
Lovász extension of a submodular function (Lovász, 1983)

- Given a vector $y$, define permutation $\sigma_y$ that “sorts” $y$, in that: $y[\sigma_y(1)] \geq \cdots \geq y[\sigma_y(n)]$.
- Also, define cumulative unions $\Sigma_k = \{\sigma(1), \sigma(2), \ldots, \sigma(k)\}$:

$$\Sigma_1 \subseteq \Sigma_2 \subseteq \Sigma_3$$
Given a vector $y$, define permutation $\sigma_y$ that “sorts” $y$, in that:

$$y[\sigma_y(1)] \geq \cdots \geq y[\sigma_y(n)].$$

Also, define cumulative unions $\Sigma_k = \{\sigma(1), \sigma(2), \ldots, \sigma(k)\}$:

The Lovász Extension:

$$\hat{f}(y) = \langle y, h^f_{\sigma_y} \rangle \quad (4)$$

where:

$$h^f_{\sigma_y}(\sigma_y(k)) = f(\Sigma_k) - f(\Sigma_{k-1}), \forall k \quad (5)$$
Lovász extension of a submodular function (Lovász, 1983)

- Given a vector $y$, define permutation $\sigma_y$ that “sorts” $y$, in that: $y[\sigma_y(1)] \geq \cdots \geq y[\sigma_y(n)]$.
- Also, define cumulative unions $\Sigma_k = \{\sigma(1), \sigma(2), \ldots, \sigma(k)\}$:

\[
\begin{array}{cccccccc}
& \sigma(1) & \sigma(2) & \sigma(3) & \sigma(4) & \sigma(5) & \sigma(6) & \sigma(7) & \sigma(8) \\
\Sigma_1 & & & & & & & & \\
\Sigma_2 & \Sigma_1 & & & & & & & \\
\Sigma_3 & \Sigma_2 & & & & & & & \\
\end{array}
\]

- The Lovász Extension:

\[
\hat{f}(y) = \langle y, h_{\sigma_y}^f \rangle
\]

where:

\[
h_{\sigma_y}^f(\sigma_y(k)) = f(\Sigma_k) - f(\Sigma_{k-1}), \forall k
\]

- If the point $y$ is totally ordered, $\hat{f}$ has a unique subgradient at $y$. 
Given a vector \( y \), define permutation \( \sigma_y \) that “sorts” \( y \), in that:
\[
y[\sigma_y(1)] \geq \cdots \geq y[\sigma_y(n)].
\]
Also, define cumulative unions \( \Sigma_k = \{\sigma(1), \sigma(2), \ldots, \sigma(k)\} \):

\[
\begin{array}{c}
\sigma(1) \\
\downarrow \\
\sum_1 \\
\downarrow \\
\sum_2 \\
\downarrow \\
\sum_3
\end{array}
\]

The Lovász Extension:

\[
\hat{f}(y) = \langle y, h^f_{\sigma_y} \rangle
\]

where:

\[
h^f_{\sigma_y}(\sigma_y(k)) = f(\Sigma_k) - f(\Sigma_{k-1}), \forall k
\]

If the point \( y \) is totally ordered, \( \hat{f} \) has a unique subgradient at \( y \).
Moreover, the subgradient \( h^f_{\sigma_y} \) depends only on \( \sigma_y \).
The Lovász-Bregman divergence

- Defined via the generalized Bregman divergences (Kiwiel, 1997)
The Lovász-Bregman divergence

- Defined via the generalized Bregman divergences (Kiwiel, 1997)
- A natural expression for the Lovász-Bregman when $y$ is totally ordered:

$$d_{\hat{f}}(x, y) = \hat{f}(x) - \langle h_{\sigma_y}^f, x \rangle$$  (6)
The Lovász-Bregman divergence

- Defined via the generalized Bregman divergences (Kiwiel, 1997)
- A natural expression for the Lovász-Bregman when \( y \) is totally ordered:

\[
d_{\hat{f}}(x, y) = \hat{f}(x) - \langle h_{\sigma_y}^f, x \rangle = \langle x, h_{\sigma_x}^f - h_{\sigma_y}^f \rangle
\]  

(6)
The Lovász-Bregman divergence

- Defined via the generalized Bregman divergences (Kiwiel, 1997)
- A natural expression for the Lovász-Bregman when \( y \) is totally ordered:
  \[
  d_{\hat{f}}(x, y) = \hat{f}(x) - \langle h^f_{\sigma_y}, x \rangle = \langle x, h^f_{\sigma_x} - h^f_{\sigma_y} \rangle
  \]

- \( d_{\hat{f}}(x, y) \) depends on \( y \) only via its permutation \( \sigma_y \).
Lovász-Bregman is a score based permutation based divergence!

$$d_{\hat{f}}(x||\sigma) = \langle x, h_{\sigma x}^f - h_{\sigma}^f \rangle$$  \hspace{1cm} (7)
Lovász-Bregman is a score based permutation based divergence!

\[ d_{\hat{f}}(x|\sigma) = \langle x, h^f_{\sigma_x} - h^f_{\sigma} \rangle \]  

\underline{Lemma}

\[ d_{\hat{f}}(x|\sigma) = 0 \text{ if and only if } \sigma_x = \sigma. \]
Lovász-Bregman is a score based permutation based divergence!

\[ d_{\hat{f}}(x||\sigma) = \langle x, h^f_{\sigma^x} - h^f_{\sigma} \rangle \]  

**Lemma**

\[ d_{\hat{f}}(x||\sigma) = 0 \text{ if and only if } \sigma^x = \sigma. \]

- Akin to the permutation metrics, except for additional dependence on valuations.
Examples of Lovász-Bregman

- **Cut functions:** \( f(X) = \sum_{i \in X, j \in V \setminus X} d_{ij}, \)

\[
d_{\hat{f}}(x, y) = \sum_{i < j} d_{ij} |x_i - x_j| I(\sigma^{-1}_x \sigma(i) > \sigma^{-1}_x \sigma(j)) \tag{8}
\]
Examples of Lovász-Bregman

- **Cut functions:** \( f(X) = \sum_{i \in X, j \in V \setminus X} d_{ij} \),

\[
d_{\hat{f}}(x, y) = \sum_{i < j} d_{ij} |x_i - x_j| I(\sigma_x^{-1}\sigma(i) > \sigma_x^{-1}\sigma(j)) \tag{8}
\]

- Akin to the Kendall \( \tau \).
Examples of Lovász-Bregman

- **Cut functions:**
  \[ f(X) = \sum_{i \in X, j \in V \setminus X} d_{ij}, \]

  \[ d_{\hat{f}}(x, y) = \sum_{i < j} d_{ij} |x_i - x_j| I(\sigma_x^{-1}(i) > \sigma_x^{-1}(j)) \] (8)

- Akin to the Kendall \( \tau \).
- Setting \( d_{ij} = 1/|x_i - x_j| \), \( d_{\hat{f}}(x||\sigma) = d_T(\sigma_x, \sigma) \).
Examples of Lovász-Bregman

- **Cut functions:** \( f(X) = \sum_{i \in X, j \in V \setminus X} d_{ij}, \)

  \[ d_{\hat{f}}(x, y) = \sum_{i < j} d_{ij} |x_i - x_j| I(\sigma_x^{-1}(i) > \sigma_x^{-1}(j)) \]  

  - Akin to the Kendall \( \tau \).
  - Setting \( d_{ij} = 1/|x_i - x_j|, \) \( d_{\hat{f}}(x||\sigma) = d_T(\sigma_x, \sigma). \)

- **Concave over Cardinality:** \( f(X) = g(|X|), \)

  \[ d_{\hat{f}}(x, y) = \sum_{i=1}^{n} x(\sigma_x(i))\delta_g(i) - \sum_{i=1}^{k} x(\sigma(i))\delta_g(i) \]
Examples of Lovász-Bregman

- **Cut functions:** \( f(X) = \sum_{i \in X} \sum_{j \in V \setminus X} d_{ij} \),

\[
d_{\hat{f}}(x, y) = \sum_{i < j} d_{ij} |x_i - x_j| I(\sigma_x^{-1} \sigma(i) > \sigma_x^{-1} \sigma(j)) \tag{8}
\]

- Akin to the Kendall \( \tau \).
- Setting \( d_{ij} = 1/|x_i - x_j| \), \( d_{\hat{f}}(x||\sigma) = d_T(\sigma_x, \sigma) \).

- **Concave over Cardinality:** \( f(X) = g(|X|) \),

\[
d_{\hat{f}}(x, y) = \sum_{i=1}^{n} x(\sigma_x(i)) \delta_g(i) - \sum_{i=1}^{k} x(\sigma(i)) \delta_g(i) \tag{9}
\]

- Setting \( f(X) = \min\{|X|, k\} \),

\[
d_{\hat{f}}(x, y) = \sum_{i=1}^{k} x(\sigma_x(i)) - \sum_{i=1}^{k} x(\sigma(i)).
\]
Lovász-Bregman as Ranking Measures

Subsume commonly used loss measures in web ranking (see paper for details).
Subsume commonly used loss measures in web ranking (see paper for details).

- **The Normalized Discounted Cumulative Gain (NDCG)**
Subsume commonly used loss measures in web ranking (see paper for details).

- **The Normalized Discounted Cumulative Gain (NDCG)** (Järvelin & Kekäläinen 2002, Ravikumar et al, 2011). A special instance of the LB divergence corresponding to concave over cardinality functions!
Lovász-Bregman as Ranking Measures

Subsume commonly used loss measures in web ranking (see paper for details).

- **The Normalized Discounted Cumulative Gain (NDCG)** (Järvelin & Kekäläinen 2002, Ravikumar et al, 2011). A special instance of the LB divergence corresponding to concave over cardinality functions!

- **Area Under Curve (AUC)** (Fawcett, 2006).
Subsume commonly used loss measures in web ranking (see paper for details).

- **The Normalized Discounted Cumulative Gain (NDCG)** (Järvelin & Kekäläinen 2002, Ravikumar et al, 2011). A special instance of the LB divergence corresponding to concave over cardinality functions!

- **Area Under Curve (AUC)** (Fawcett, 2006). A special instance of the LB divergence corresponding to cut functions!
Properties of the Lovász-Bregman

- **Convexity:** The Lovász-Bregman $d_{\hat{T}}(x||\sigma)$ is convex in $x$ for a given $\sigma$. 
Properties of the Lovász-Bregman

- **Convexity:** The Lovász-Bregman $d_\hat{f}(x||\sigma)$ is convex in $x$ for a given $\sigma$.

- **Invariance over relabellings:** Given a submodular function depending only on cardinality, $d_\hat{f}(\tau x||\tau \sigma) = d_\hat{f}(x||\sigma)$.
Properties of the Lovász-Bregman

- **Convexity:** The Lovász-Bregman $d_\hat{f}(x||\sigma)$ is convex in $x$ for a given $\sigma$.

- **Invariance over relabellings:** Given a submodular function depending only on cardinality, $d_\hat{f}(\tau x||\tau \sigma) = d_\hat{f}(x||\sigma)$.

- **Dependence on values and not just orderings:** Low confidence in the ordering of $x \Rightarrow d_\hat{f}(x||\sigma)$ small for every permutation $\sigma$.

The Lovász-Bregman divergence (left) and Kendall $\tau \ d_T(\sigma_x, \sigma)$ (right)
● **Priority for higher rankings:** Greater penalty to misorderings of $\sigma_x$ and $\sigma$ higher up in the rankings.
• **Priority for higher rankings:** Greater penalty to misorderings of $\sigma_X$ and $\sigma$ higher up in the rankings.

• **Extension to partial rankings:** Natural interpretations for $d_\hat{f}(x||\sigma)$ when $\sigma$ given as a top $k$ list or a partial ordering.
Priority for higher rankings: Greater penalty to misorderings of \( \sigma_x \) and \( \sigma \) higher up in the rankings.

Extension to partial rankings: Natural interpretations for \( d_\tilde{f}(x \| \sigma) \) when \( \sigma \) given as a top \( k \) list or a partial ordering.

Lovász Mallows model: Forms of Mallows model and Generalized Mallows model:
- **Priority for higher rankings:** Greater penalty to misorderings of \( \sigma_x \) and \( \sigma \) higher up in the rankings.

- **Extension to partial rankings:** Natural interpretations for \( d_{\hat{f}}(x||\sigma) \) when \( \sigma \) given as a top \( k \) list or a partial ordering.

- **Lovász Mallows model:** Forms of Mallows model and Generalized Mallows model:

\[
  p(x|\theta, \sigma) = \frac{\exp(-\theta d_{\hat{f}}(x||\sigma))}{Z(\theta, \sigma)}, \quad p(\sigma|\Theta, \mathcal{X}) = \frac{\exp(-\sum_{i=1}^{n} \theta_i d_{\hat{f}}(x_i||\sigma))}{Z(\Theta, \mathcal{X})}
\]
Priority for higher rankings: Greater penalty to misorderings of $\sigma_x$ and $\sigma$ higher up in the rankings.

Extension to partial rankings: Natural interpretations for $d_f(x||\sigma)$ when $\sigma$ given as a top $k$ list or a partial ordering.

Lovász Mallows model: Forms of Mallows model and Generalized Mallows model:

$$p(x|\theta, \sigma) = \frac{\exp(-\theta d_f(x||\sigma))}{Z(\theta, \sigma)}, \quad p(\sigma|\Theta, \mathcal{X}) = \frac{\exp(-\sum_{i=1}^{n} \theta_i d_f(x_i||\sigma))}{Z(\Theta, \mathcal{X})}$$

We shall see interesting connections to web ranking!
Combining Permutations

- Combine permutations: $\sigma_1, \sigma_2, \cdots, \sigma_n$. 

\[ \sigma = \arg \min_{\pi} \sum_{i=1}^{n} d(\sigma_i, \pi) \]

NP hard for most permutation based metrics!

Combining Scores

Often we have a collection of scores $\{x_1, x_2, \cdots, x_n\}$:

\[ \sigma = \arg \min_{\pi} \sum_{i=1}^{n} \hat{d}(x_i | | \pi) \]

Can be solved in closed form!

\[ \sigma = \sigma_{\mu} \text{, where } \mu = \frac{1}{n} \sum_{i=1}^{n} w_i x_i \]
Rank Aggregation v.s Score Aggregation

Combining Permutations

- Combine permutations: $\sigma_1, \sigma_2, \cdots, \sigma_n$.
- $\sigma = \arg\min_{\pi} \sum_{i=1}^{n} d(\sigma_i, \pi)$
Rank Aggregation v.s Score Aggregation

**Combining Permutations**

- Combine permutations: \( \sigma_1, \sigma_2, \cdots, \sigma_n \).
- \( \sigma = \arg\min_{\pi} \sum_{i=1}^{n} d(\sigma_i, \pi) \)
- NP hard for most permutation based metrics!
Combining Permutations

- Combine permutations: \( \sigma_1, \sigma_2, \ldots, \sigma_n \).
- \( \sigma = \arg\min_\pi \sum_{i=1}^{n} d(\sigma_i, \pi) \)
- NP hard for most permutation based metrics!

Combining Scores

- Often we have a collection of scores \( \{x_1, x_2, \ldots, x_n\} \):
Rank Aggregation v.s Score Aggregation

Combining Permutations
- Combine permutations: $\sigma_1, \sigma_2, \cdots, \sigma_n$.
- $\sigma = \arg\min_{\pi} \sum_{i=1}^{n} d(\sigma_i, \pi)$
- NP hard for most permutation based metrics!

Combining Scores
- Often we have a collection of scores $\{x_1, x_2, \cdots, x_n\}$:
- $\sigma = \arg\min_{\pi} \sum_{i=1}^{n} \hat{d}_f(x_i || \pi)$
## Combining Permutations

- Combine permutations: $\sigma_1, \sigma_2, \cdots, \sigma_n$.
- $\sigma = \arg\min_{\pi} \sum_{i=1}^{n} d(\sigma_i, \pi)$
- NP hard for most permutation based metrics!

## Combining Scores

- Often we have a collection of scores $\{x_1, x_2, \cdots, x_n\}$:
- $\sigma = \arg\min_{\pi} \sum_{i=1}^{n} d_{\hat{f}}(x_i \mid \pi)$
- Can be solved in closed form!
Combining Permutations

- Combine permutations: $\sigma_1, \sigma_2, \cdots, \sigma_n$.
- $\sigma = \arg\min_\pi \sum_{i=1}^n d(\sigma_i, \pi)$
- NP hard for most permutation based metrics!

Combining Scores

- Often we have a collection of scores $\{x_1, x_2, \cdots, x_n\}$:
- $\sigma = \arg\min_\pi \sum_{i=1}^n d_\hat{\ell}(x_i \| \pi)$
- Can be solved in closed form!
- $\sigma = \sigma_\mu$, where $\mu = \frac{1}{n} \sum_{i=1}^n w_i x_i$
A new view of web ranking

\[
D = \{d_1, d_2, \ldots, d_N\},
\]

Documents

Features

\[
\sigma = \arg\min_{\pi} \Psi(D, \pi) = \arg\min_{\pi} \sum_{i=1}^{M} w_i \hat{f}(d_i | \sigma)
\]

\[
\sigma = \sigma_{\mu}, \mu = \sum_{i=1}^{M} w_i d_i \quad \text{and} \quad \mu(j) = \langle w, d_j \rangle.
\]

Functions of this form used in the past (Yue et al 2007, Chakrabarti et al 2008).
A new view of web ranking

\[ D = \{ d_1, d_2, \ldots, d_N \}, \quad d_j \in \mathbb{R}^M. \]
A new view of web ranking

- $\mathcal{D} = \{d_1, d_2, \cdots, d_N\}$, $d_j \in \mathbb{R}^M$.
- Feature vectors: $d^1, d^2, \cdots, d^M$ and hence $d^i(j) = d_j(i)$. 
A new view of web ranking

- $\mathcal{D} = \{d_1, d_2, \cdots, d_N\}, d_j \in \mathbb{R}^M$.
- Feature vectors: $d^1, d^2, \cdots, d^M$ and hence $d^i(j) = d_j(i)$.

**Inference Problem:**

$$\sigma = \arg\min_{\pi} \Psi(\mathcal{D}, \pi) = \arg\min_{\pi} \sum_{i=1}^{M} w^i d^i_{\hat{f}}(d^i||\sigma)$$
A new view of web ranking

- \( \mathcal{D} = \{d_1, d_2, \ldots, d_N\}, d_j \in \mathbb{R}^M. \)
- Feature vectors: \( d^1, d^2, \ldots, d^M \) and hence \( d^i(j) = d_j(i) \).

Inference Problem: \( \sigma = \arg\min_\pi \Psi(\mathcal{D}, \pi) = \arg\min_\pi \sum_{i=1}^M w^i d^i_f (d^i || \sigma) \)

\( \sigma = \sigma_\mu, \mu = \sum_{i=1}^M w_i d^i \) and \( \mu(j) = \langle w, d_j \rangle. \)
A new view of web ranking

- $\mathcal{D} = \{d_1, d_2, \cdots, d_N\}$, $d_j \in \mathbb{R}^M$.
- Feature vectors: $d^1, d^2, \cdots, d^M$ and hence $d^i(j) = d_j(i)$.

Inference Problem: $\sigma = \arg\min_{\pi} \Psi(\mathcal{D}, \pi) = \arg\min_{\pi} \sum_{i=1}^{M} w^i d_i^\pi (d^i || \sigma)$

- $\sigma = \sigma_\mu$, $\mu = \sum_{i=1}^{M} w_i d^i$ and $\mu(j) = \langle w, d_j \rangle$.
- Functions of this form used in the past (Yue et al 2007, Chakrabarti et al 2008).
Conditional Models for ranking

- Conditional probability models for ranking:

\[
p(\sigma|\Theta, \mathcal{D}) \propto \exp(-\Psi(\mathcal{D}, \sigma)) \propto \exp\left(-\sum_{i=1}^{M} w^i d_f^i(d^i||\sigma)\right) \tag{10}
\]
Conditional Models for ranking

- Conditional probability models for ranking:

  \[ p(\sigma|\Theta, D) \propto \exp(-\Psi(D, \sigma)) \propto \exp\left(-\sum_{i=1}^{M} w^i d^i_f(d^i||\sigma)\right) \quad (10) \]

- This is exactly the Mallow’s model corresponding to Lovász-Bregman divergence!
Conditional Models for ranking

- Conditional probability models for ranking:

\[ p(\sigma|\Theta, D) \propto \exp(-\Psi(D, \sigma)) \propto \exp(-\sum_{i=1}^{M} w^i d_i^f(d^i||\sigma)) \]  \hspace{1cm} (10)

- This is exactly the Mallow’s model corresponding to Lovász-Bregman divergence!

- These models have been used in past work (Dubey et al, 2009).
Score based clustering

- K-means style clustering algorithm for clustering ordered vectors.
Score based clustering

- K-means style clustering algorithm for clustering ordered vectors.
- Each step in the k-means is easy!
Score based clustering

- K-means style clustering algorithm for clustering ordered vectors.
- Each step in the k-means is easy!
- Some clustering visualizations:

![Clustering based on orderings in 2 and 3 Dimensions.](image)
Summary

- Rank aggregation and permutation based metrics.
- Lovász-Bregman divergence as score & permutation divergence.
- Properties of the Lovász-Bregman divergence.
- Interesting connections to web ranking and rank aggregation.
Thank You

Questions