

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

Rishabh Iyer    Jeff Bilmes

University of Washington, Seattle

UAI-2013



# Outline

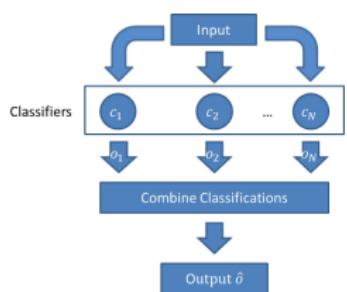
- 1 Ranking and Machine Learning
- 2 The Lovász-Bregman divergences
- 3 Properties of the Lovász-Bregman
- 4 Applications
- 5 Summary

# Combining Scores and Rankings

Occur in a number of Machine Learning applications:

# Combining Scores and Rankings

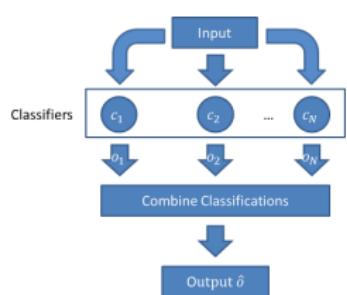
Occur in a number of Machine Learning applications:



Combining Classifiers  
(Lebanon & Lafferty, 2002)

# Combining Scores and Rankings

Occur in a number of Machine Learning applications:



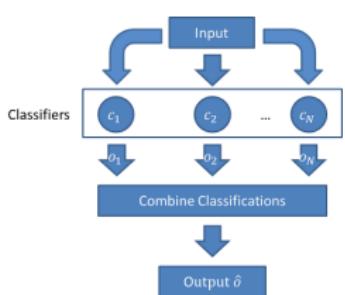
Combining Classifiers  
(Lebanon & Lafferty, 2002)



Aggregating Preferences  
(Murphy & Martin,  
2003)

# Combining Scores and Rankings

Occur in a number of Machine Learning applications:



Combining Classifiers  
(Lebanon & Lafferty, 2002)



Aggregating Preferences  
(Murphy & Martin,  
2003)

Google UAI/2013

Web Images Maps Shopping Videos More Search tools

About 11,200,000 results (0.47 seconds)

**UAI 2013 - Association for Uncertainty in Artificial Intelligence**  
[uai.org/uai2013/](http://uai.org/uai2013/)

The Conference on Uncertainty in Artificial Intelligence (UAI) is the premier international conference in research related to representation, inference, learning ...

You've visited this page many times. Last visit: 6/29/13

**Accepted Papers**  
UAI 2013 ACCEPTED PAPERS: Oral Presentations, ID 196 ...

**Workshops**  
Big Data and Complex Models - A UAI Application Workshop

**Scholarships**  
UAI 2013 Scholarships: The tutorials will be held on Thursday, July ...

**Program Committee**  
Program Committee: Ryan Adams, Harvard, Arndt Arentz ...

**More results from uai.org »**

**UAI 2013 Call for Papers | International Society for Bayesian Analysis**  
[bayes.org/~Forums/BayesPublicForums/Conferences/](http://bayes.org/~Forums/BayesPublicForums/Conferences/)

Oct 8, 2012 - CALL FOR PAPERS: 29th Uncertainty in Artificial Intelligence (UAI) Conference, July 11-15th 2013 <http://uai.cs.washington.edu/>, Key Dates ...

**UAI 2013 - 29th Conference on Uncertainty in Artificial Intelligence**  
[www.uai.cs.washington.edu/](http://uai.cs.washington.edu/)

Jul 12 - Jul 16 - Bellevue, Washington, USA

**UAI 2013 - 29th Conference on Uncertainty in Artificial Intelligence**

Web Ranking (Liu, 2009)

# Combining Scores and Rankings

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

$$\sigma(1) \quad \sigma(2) \quad \sigma(3) \quad \sigma(4) \quad \sigma(5) \quad \sigma(6) \quad \sigma(7) \quad \sigma(8)$$

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

# Combining Scores and Rankings

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

$$\sigma(1) \quad \sigma(2) \quad \sigma(3) \quad \sigma(4) \quad \sigma(5) \quad \sigma(6) \quad \sigma(7) \quad \sigma(8)$$

- Denote  $\{\sigma_1, \sigma_2, \dots, \sigma_k\}$  as a set of  $k$  permutations.
- Some important problems concerning rankings:

# Combining Scores and Rankings

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

$$\sigma(1) \quad \sigma(2) \quad \sigma(3) \quad \sigma(4) \quad \sigma(5) \quad \sigma(6) \quad \sigma(7) \quad \sigma(8)$$

- Denote  $\{\sigma_1, \sigma_2, \dots, \sigma_k\}$  as a set of  $k$  permutations.
- Some important problems concerning rankings:

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

# Combining Scores and Rankings

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

$$\sigma(1) \quad \sigma(2) \quad \sigma(3) \quad \sigma(4) \quad \sigma(5) \quad \sigma(6) \quad \sigma(7) \quad \sigma(8)$$

- Denote  $\{\sigma_1, \sigma_2, \dots, \sigma_k\}$  as a set of  $k$  permutations.
- Some important problems concerning rankings:
  - Combining Permutations:** Given permutations  $\sigma_1, \sigma_2, \dots, \sigma_k$ , find a representative  $\sigma$ , which is “close” to  $\sigma_1, \sigma_2, \dots, \sigma_k$ .
  - Combining Scores:** Given a set of score vectors  $x_1, x_2, \dots, x_k$ , find a representative  $\sigma$ , which is “close” to  $x_1, x_2, \dots, x_k$ .

# Combining Scores and Rankings

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

$$\sigma(1) \quad \sigma(2) \quad \sigma(3) \quad \sigma(4) \quad \sigma(5) \quad \sigma(6) \quad \sigma(7) \quad \sigma(8)$$

- Denote  $\{\sigma_1, \sigma_2, \dots, \sigma_k\}$  as a set of  $k$  permutations.
- Some important problems concerning rankings:

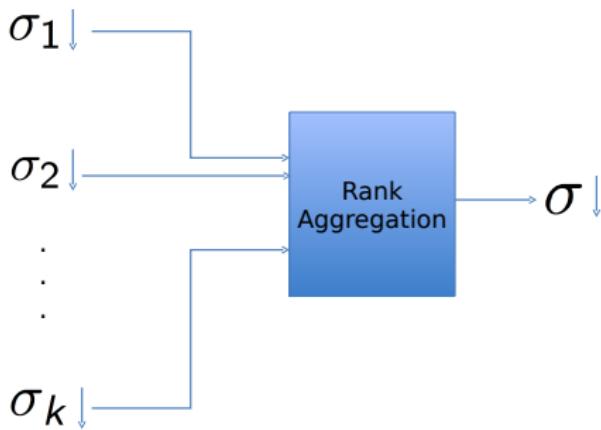
- 1 Combining Permutations:** Given permutations  $\sigma_1, \sigma_2, \dots, \sigma_k$ , find a representative  $\sigma$ , which is “close” to  $\sigma_1, \sigma_2, \dots, \sigma_k$ .
- 2 Combining Scores:** Given a set of score vectors  $x_1, x_2, \dots, x_k$ , find a representative  $\sigma$ , which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

# Rank aggregation

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

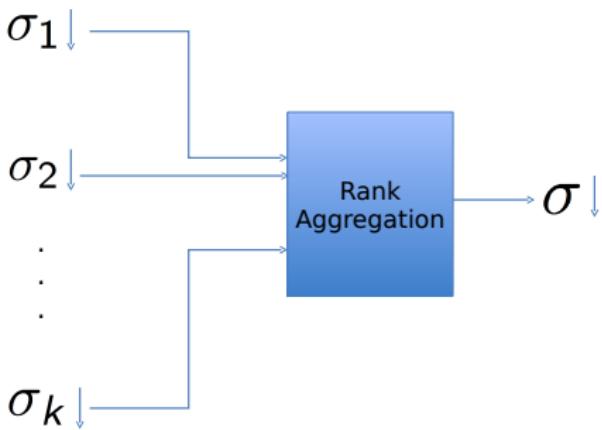
# Rank aggregation

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



# Rank aggregation

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



- 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  $d(\pi\sigma, \pi\tau) = d(\sigma, \tau)$ .

# 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  $\sigma_1, \sigma_2, \dots, \sigma_k$ , find a permutation  $\sigma$ :

$$\sigma = \operatorname{argmin}_{\pi} \sum_{i=1}^k d(\sigma_i, \pi) \quad (1)$$

# 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

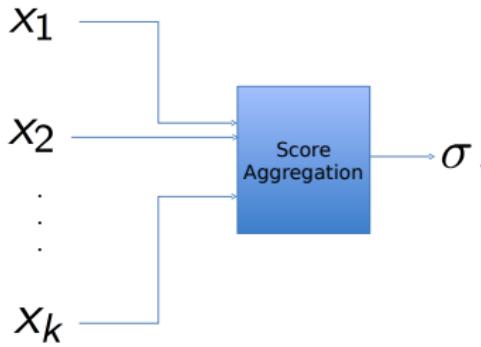
- What if one has scores instead of just the orderings? For example,
  - ➊ Combining Classifiers: probability distribution
  - ➋ Web ranking: Feature functions

# 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  $x_1, x_2, \dots, x_k$  and find a representative ordering  $\sigma$ .

# 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  $x_1, x_2, \dots, x_k$  and find a representative ordering  $\sigma$ .



# Score & permutation based divergence $d(x||\sigma)$

- A natural formulation of this problem is through a score & permutation based divergence.

# 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  $\sigma$ .

# 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.

# 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.
- Given a set of scores  $x_1, x_2, \dots, x_k$ , find a permutation  $\sigma$ :

$$\sigma = \operatorname{argmin}_{\pi} \sum_{i=1}^k d(x_i || \pi) \quad (2)$$

# This Talk!



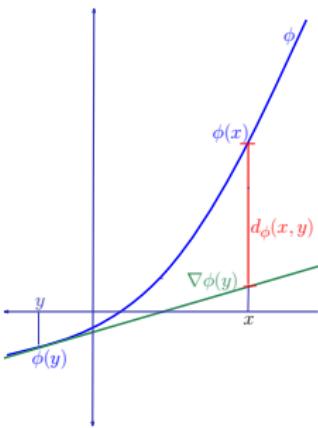
Lovasz Bregman  
Divergences



# Bregman Divergences

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

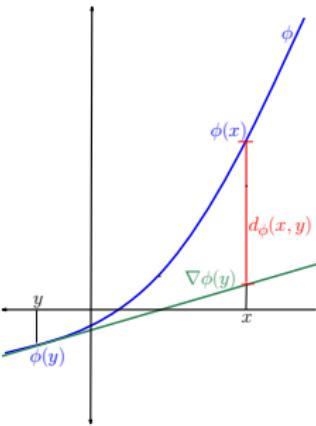
$$d_\phi(x, y) = \phi(x) - \phi(y) - \langle \nabla \phi(y), x - y \rangle.$$



# Bregman Divergences

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

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

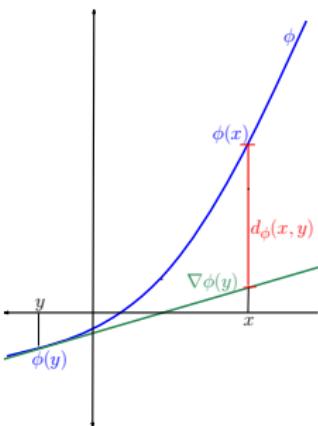


- Occur naturally in many machine learning applications:

# 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:
- Subsumes many useful distance measures (e.g., Squared Euclidean, KL-divergence, Itakura Saito etc.)

# Definition and notation

- Submodular functions: special class of set functions.

$$f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B \quad (3)$$

# Definition and notation

- Submodular functions: special class of set functions.

$$f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B \quad (3)$$

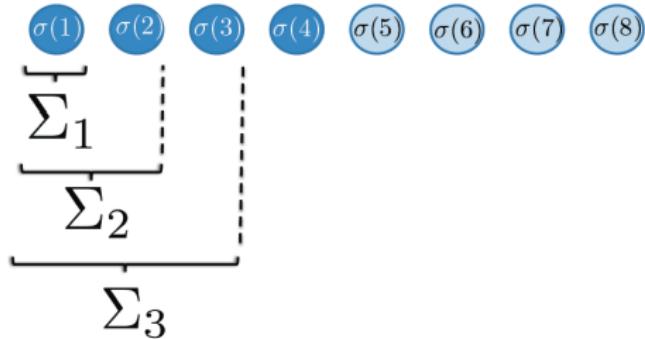
- A bit more notation we will use: Given a vector  $y$ , define permutation  $\sigma_y$  that “sorts”  $y$ , in that:  $y[\sigma_y(1)] \geq \dots \geq y[\sigma_y(n)]$ .

# Definition and notation

- Submodular functions: special class of set functions.

$$f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B \quad (3)$$

- A bit more notation we will use: Given a vector  $y$ , define permutation  $\sigma_y$  that “sorts”  $y$ , in that:  $y[\sigma_y(1)] \geq \dots \geq y[\sigma_y(n)]$ .
- Also, define cumulative unions  $\Sigma_k = \{\sigma(1), \sigma(2), \dots, \sigma(k)\}$ :



# Lovász extension of a submodular function (Lovász, 1983)

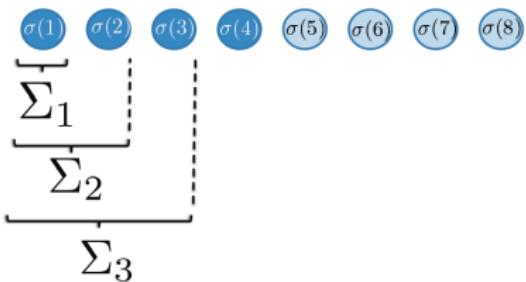
- The Lovász Extension:

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

where:

$$h_{\sigma_y}^f(\sigma_y(k)) = f(\Sigma_k) - f(\Sigma_{k-1}), \forall k \quad (5)$$

and  $\Sigma_k$  is defined previously:



# Lovász extension of a submodular function (Lovász, 1983)

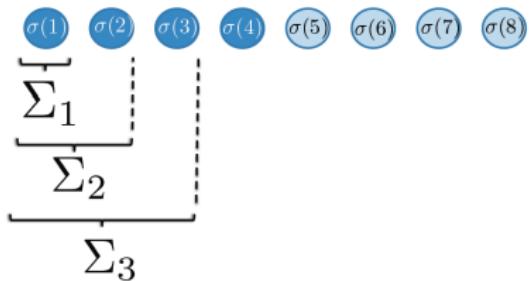
- The Lovász Extension:

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

where:

$$h_{\sigma_y}^f(\sigma_y(k)) = f(\Sigma_k) - f(\Sigma_{k-1}), \forall k \quad (5)$$

and  $\Sigma_k$  is defined previously:



- If the point  $y$  is totally ordered,  $\hat{f}$  has a unique subgradient at  $y$ .

# Lovász extension of a submodular function (Lovász, 1983)

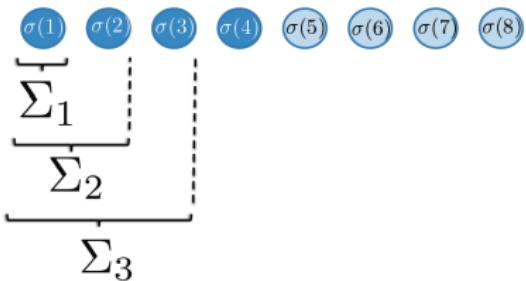
- The Lovász Extension:

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

where:

$$h_{\sigma_y}^f(\sigma_y(k)) = f(\Sigma_k) - f(\Sigma_{k-1}), \forall k \quad (5)$$

and  $\Sigma_k$  is defined previously:



- If the point  $y$  is totally ordered,  $\hat{f}$  has a unique subgradient at  $y$ .
- Moreover, the subgradient  $h_{\sigma_y}^f$  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:

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

# The Lovász Bregman divergence

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

$$\begin{aligned} d_{\hat{f}}(x, y) &= \hat{f}(x) - \langle h_{\sigma_y}^f, x \rangle \\ &= \langle x, h_{\sigma_x}^f - h_{\sigma_y}^f \rangle \end{aligned} \tag{6}$$

# The Lovász Bregman divergence

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

$$\begin{aligned} d_{\hat{f}}(x, y) &= \hat{f}(x) - \langle h_{\sigma_y}^f, x \rangle \\ &= \langle x, h_{\sigma_x}^f - h_{\sigma_y}^f \rangle \end{aligned} \tag{6}$$

- $d_{\hat{f}}(x, y)$  depends on  $y$  only via its permutation  $\sigma_y$ .

# LB divergence as a Score based permutation divergence

- Lovász Bregman is a score based permutation based divergence!

$$d_{\hat{f}}(x||\sigma) = \langle x, h_{\sigma_x}^f - h_\sigma^f \rangle \quad (7)$$

# LB divergence as a Score based permutation divergence

- Lovász Bregman is a score based permutation based divergence!

$$d_{\hat{f}}(x||\sigma) = \langle x, h_{\sigma_x}^f - h_\sigma^f \rangle \quad (7)$$

## Lemma

$d_{\hat{f}}(x||\sigma) = 0$  if and only if  $\sigma_x = \sigma$ .

# LB divergence as a Score based permutation divergence

- Lovász Bregman is a score based permutation based divergence!

$$d_{\hat{f}}(x||\sigma) = \langle x, h_{\sigma_x}^f - h_\sigma^f \rangle \quad (7)$$

## Lemma

$d_{\hat{f}}(x||\sigma) = 0$  if and only if  $\sigma_x = \sigma$ .

- Akin to the permutation metrics, except for additional dependence on valuations.

# Examples of Lovász Bregman

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

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

# Examples of Lovász Bregman

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

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

- Akin to the Kendall  $\tau$ .

# Examples of Lovász Bregman

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

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

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

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

- 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:**  $f(X) = g(|X|)$ ,

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

# Examples of Lovász Bregman

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

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

- 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:**  $f(X) = g(|X|)$ ,

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

- 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).

# 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).

# 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!

# 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).

# 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). A special instance of the LB divergence corresponding to cut functions!

# Properties of the Lovász Bregman

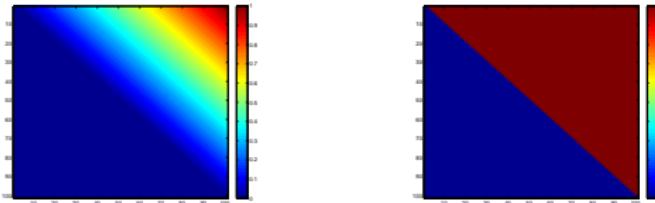
- **Convexity:** The Lovász Bregman  $d_{\hat{f}}(x||\sigma)$  is convex in  $x$  for a given  $\sigma$ .

# 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  $d_{\hat{f}}(x||\sigma)$  when  $\sigma$  given as a top  $k$  list or a partial ordering.
- **Lovász Mallows model:** Forms of Mallows model and Generalized Mallows model:

- **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  $d_{\hat{f}}(x||\sigma)$  when  $\sigma$  given as a top  $k$  list or a partial ordering.
- **Lovász Mallows model:** Forms of Mallows model and Generalized Mallows model:

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

We shall see interesting connections to web ranking!

# Lovász Bregman Rank Aggregation

## Combining Permutations

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

# Lovász Bregman Rank Aggregation

## Combining Permutations

- Combine permutations:  $\sigma_1, \sigma_2, \dots, \sigma_n$ .
- $\sigma = \operatorname{argmin}_{\pi} \sum_{i=1}^n d(\sigma_i, \pi)$

# Lovász Bregman Rank Aggregation

## Combining Permutations

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

# Lovász Bregman Rank Aggregation

## Combining Permutations

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

## Combining Scores

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

# Lovász Bregman Rank Aggregation

## Combining Permutations

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

## Combining Scores

- Often we have a collection of scores  $\{x_1, x_2, \dots, x_n\}$ :
- $\sigma = \operatorname{argmin}_{\pi} \sum_{i=1}^n d_{\hat{f}}(x_i || \pi)$

# Lovász Bregman Rank Aggregation

## Combining Permutations

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

## Combining Scores

- Often we have a collection of scores  $\{x_1, x_2, \dots, x_n\}$ :
- $\sigma = \operatorname{argmin}_{\pi} \sum_{i=1}^n d_{\hat{f}}(x_i || \pi)$
- Can be solved in closed form!

# Lovász Bregman Rank Aggregation

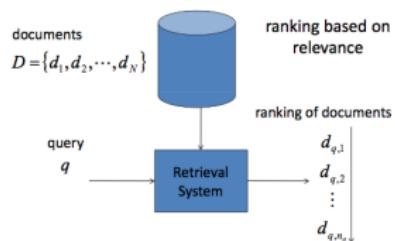
## Combining Permutations

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

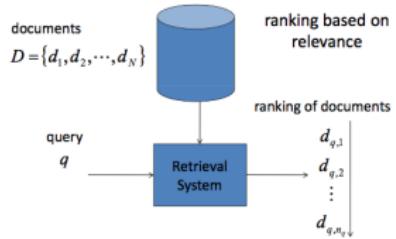
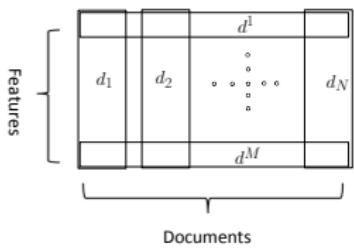
## Combining Scores

- Often we have a collection of scores  $\{x_1, x_2, \dots, x_n\}$ :
- $\sigma = \operatorname{argmin}_{\pi} \sum_{i=1}^n d_{\hat{f}}(x_i || \pi)$
- 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

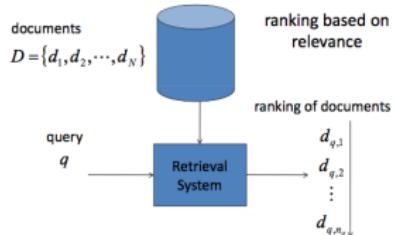
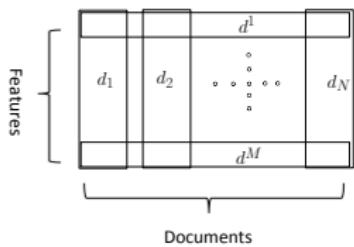


# A new view of web ranking



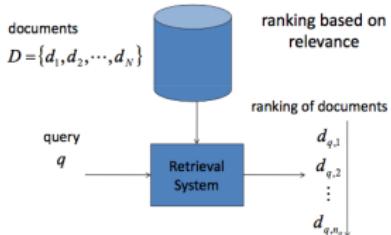
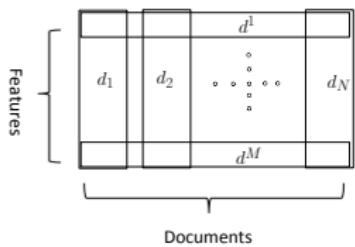
- $\mathcal{D} = \{d_1, d_2, \dots, d_N\}, d_j \in \mathbb{R}^M$ .

# A new view of web ranking



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

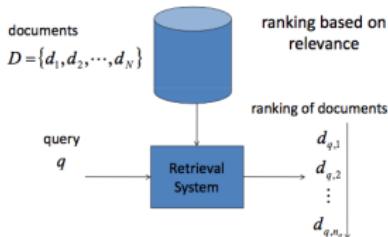
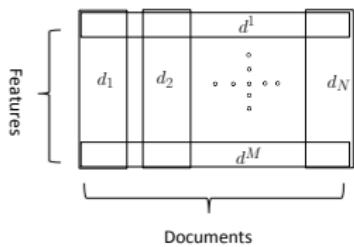
# A new view of web ranking



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

Inference Problem:  $\sigma = \operatorname{argmin}_{\pi} \Psi(\mathcal{D}, \pi) = \operatorname{argmin}_{\pi} \sum_{i=1}^M w^i d_{\hat{f}}(d^i || \sigma)$

# A new view of web ranking

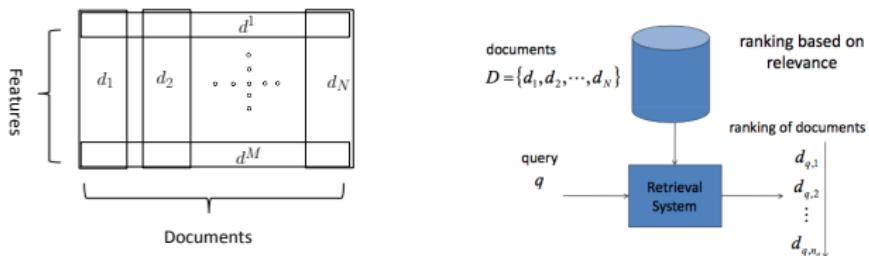


- $\mathcal{D} = \{d_1, d_2, \dots, d_N\}, d_j \in \mathbb{R}^M$ .
- Feature vectors:  $d^1, d^2, \dots, d^M$  and hence  $d^i(j) = d_j(i)$ .

Inference Problem:  $\sigma = \operatorname{argmin}_{\pi} \Psi(\mathcal{D}, \pi) = \operatorname{argmin}_{\pi} \sum_{i=1}^M w^i d_{\hat{f}}(d^i || \sigma)$

- $\sigma = \sigma_\mu, \mu = \sum_{i=1}^M w_i d^i$  and  $\mu(j) = \langle w, d_j \rangle$ .

# A new view of web ranking



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

Inference Problem:  $\sigma = \operatorname{argmin}_{\pi} \Psi(\mathcal{D}, \pi) = \operatorname{argmin}_{\pi} \sum_{i=1}^M w^i d_{\hat{f}}(d^i || \sigma)$

- $\sigma = \sigma_\mu, \mu = \sum_{i=1}^M w_i d^i$  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:

$$p(\sigma|\Theta, \mathcal{D}) \propto \exp(-\Psi(\mathcal{D}, \sigma)) \propto \exp\left(-\sum_{i=1}^M w^i d_{\hat{f}}(d^i || \sigma)\right) \quad (10)$$

# Conditional Models for ranking

- Conditional probability models for ranking:

$$p(\sigma|\Theta, \mathcal{D}) \propto \exp(-\Psi(\mathcal{D}, \sigma)) \propto \exp\left(-\sum_{i=1}^M w^i d_{\hat{f}}(d^i || \sigma)\right) \quad (10)$$

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

# Conditional Models for ranking

- Conditional probability models for ranking:

$$p(\sigma|\Theta, \mathcal{D}) \propto \exp(-\Psi(\mathcal{D}, \sigma)) \propto \exp\left(-\sum_{i=1}^M w^i d_{\hat{f}}(d^i || \sigma)\right) \quad (10)$$

- 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

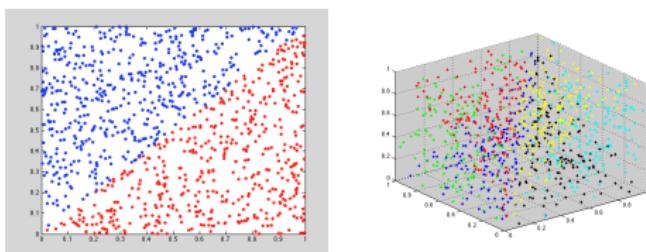
- K-means style clustering algorithm for clustering ordered vectors.

# 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.

# 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

