

A decorative header at the top of the slide features four overlapping spheres. From left to right, they are light green, light blue, light red, and light yellow. The spheres are partially cut off by the top edge of the slide.

Combined Regression and Ranking

D. Sculley
Google Pittsburgh



The Claim

- Many applications require models that give **both**:
 - good **regression** performance and
 - good **ranking** performance

Example: Predicting Star Ratings



digital camera

Search

About 145,662 results (0.49 seconds)

[Advanced Product Search](#)

Everything

Shopping

Show only:

- Google Checkout
- Free shipping
- New items

Any category

- [Digital Cameras](#)
- [Camera Bags](#)
- [Camera Batteries](#)
- [Digital Camera Accessories](#)
- [Photography Books](#)
- [More »](#)

[Top 10 Digital Cameras](#)

[Cameras.NetShoppers.com](#)

Digital Camera Sale Now On! Low Prices + Free Shipping

Tax and shipping for Pittsburgh, PA 15206 - [Change](#)



[Kodak EASYSHARE CD40 4 MP Digital Camera](#)

MultiMediaCard, SD Memory Card, 1y warranty, F/4.5

Through the years, Kodak has led the way with an abundance of new products and processes that have made photography simpler, more useful and more enjoyable. Today, the ...

★★★★☆ 5 reviews - [Add to Shopping List](#)



[Sony Cyber-shot DSC-W70 7.2 MP Digital Camera \(Silver\)](#)

Memory Stick Duo, Memory Stick PRO Duo, F/2.8-5.2

The Sony Cyber-shot DSC-W70 is an ideal camera for starting out right in digital photography, combining the quality of 7.2 megapixels and Carl Zeiss 3X optical zoom, the ...

★★★★☆ 117 reviews - [Add to Shopping List](#)



Example: Click Prediction



football tickets

Search

About 55,900,000 results (0.20 seconds)

[Advanced search](#)

Everything

News

More

Any time

Latest

Past 3 weeks

More search tools

Something different

[basketball tickets](#)

[nfl tickets](#)

[soccer tickets](#)

[hockey tickets](#)

[sports tickets](#)

[Football Tickets](#)

Sponsored links

www.TicketLiquidator.com/Steelers

Steelers Football Tickets - Cheap. Check Our Prices. Save 10% Or More.

TicketLiquidator.com is rated ★★★★★ on Google Products ([31 reviews](#))

[Football Tickets](#)

www.StubHub.com/Football

NFL & College Football Tickets. StubHub - Official Site.

[50% Off Football Tickets](#)

www.TicketZoom.com

Save up to \$50 - Use Code JUL36. Last Minute Deals. Save up to 50%.

[Football Tickets at StubHub! NFL Football Tickets, College ...](#) ☆

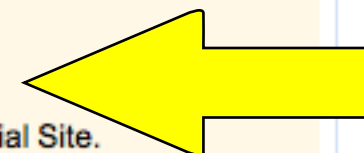
Football Tickets - Buy and sell **Football Tickets** and other Sports Tickets at StubHub! Where Fans Buy & Sell Tickets.

www.stubhub.com > [Sports tickets](#) - [Cached](#) - [Similar](#)

[NFL Tickets - Football Tickets](#) ☆

Largest inventory of NFL **tickets** online at Vividseats.com. Get **ticket** prices for all NFL teams, the NFL Playoffs and the Super Bowl.

www.vividseats.com/nfl/ - [Cached](#) - [Similar](#)



- 
- A decorative header at the top of the slide features four overlapping spheres: a green one on the far left, a blue one in the middle, a red one partially behind the blue one, and a yellow one on the right. A thin black horizontal line runs across the page just below these spheres.
- Why not just use existing methods?

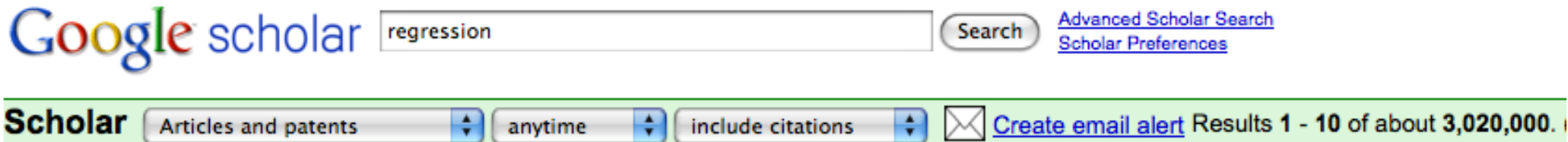
Standard Methods Can Fail Badly

- Rank-based models may do arbitrarily badly at regression
- Perfect regression implies perfect ranking, but...
- Even "good" regression can have bad ranking performance

Our Approach

- Novelty: optimize ranking and regression simultaneously
 - primary goal: try and get "best of both" performance
 - do as well at ranking as a ranking-only method
 - do as well at regression as a regression-only method
 - secondary goal: improved regression through ranking?
- We'll build this up in pieces

Supervised Regression (birds eye view)



- Goal: learn a model w that predicts a *real valued* target y
- Examples:
 - Least mean squares
 - Ridge Regression
 - LASSO
- Often solved using empirical risk minimization



Supervised Regression (review)

$$\min_{\mathbf{w} \in \mathbb{R}^m} L(\mathbf{w}, D) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2$$

Supervised Regression (review)

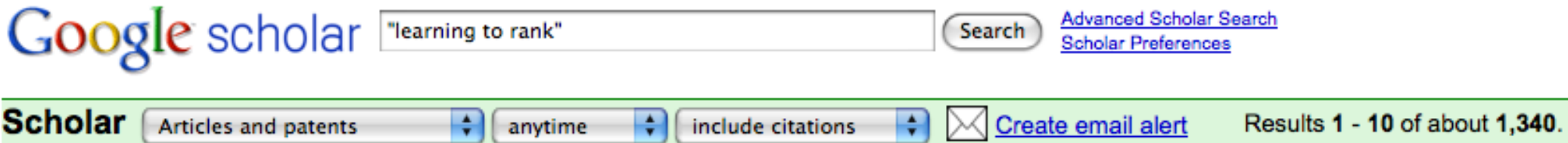
Loss
Function

$$\min_{\mathbf{w} \in \mathbb{R}^m} L(\mathbf{w}, D) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2$$

Examples: Squared Loss,
Logistic Loss, etc.

Regularization
Term

Supervised Ranking (review)



- Goal: learn a model w that puts unseen data in the correct preference order
- Several known methods:
 - RankSVM (Joachims, 2002)
 - Voted Perceptron variant (Elsas et al., 2008)
 - Boosting variants: AdaRank-MAP, AdaRank-NDCG (Xu and Li, 2007)
 - Listwise approach (Cao et al., 2007)



Supervised Ranking (review)

$$\min_{\mathbf{w} \in \mathbb{R}^m} L(\mathbf{w}, P) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2$$

Supervised Ranking (review)

$$\min_{\mathbf{w} \in \mathbb{R}^m} L(\mathbf{w}, P) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2$$

Candidate Pairs:
pairs (a,b) of
comparable
examples with
different ranks

Su



supervised ranking

Search

About 636,000 results (0.08 seconds)

[Advanced search](#)

Everything

Videos

More

All results

[Related searches](#)

More search tools

[\[PDF\] Supervised Rank Aggregation](#) ☆

File Format: PDF/Adobe Acrobat - [Quick View](#)

by YT Liu - [Cited by 27](#) - [Related articles](#)

Supervised Rank Aggregation, in which learning is formalized an ... meta-searches show that **Supervised Rank Aggregation** can ...

www2007.org/papers/paper286.pdf - [Similar](#)

[Supervised rank aggregation](#) ☆

by YT Liu - 2007 - [Cited by 27](#) - [Related articles](#)

We refer to the approach as **Supervised Rank Aggregation**. We set up a general framework for conducting **Supervised Rank Aggregation**, in which learning is ...

portal.acm.org/citation.cfm?id=1242638 - [Similar](#)

[Supervised ranking in open-domain text summarization](#) ☆

by T Nomoto - 2002 - [Cited by 3](#) - [Related articles](#)

Supervised ranking in open-domain text summarization. Full text, Publisher Site , Pdf (142 KB). Source, Annual Meeting of the ACL archive ...

portal.acm.org/citation.cfm?id=1073161

[Show more results from portal.acm.org](#)

[\[PDF\] Supervised Ranking in Open-Domain Text Summarization](#) ☆

File Format: PDF/Adobe Acrobat - [Quick View](#)

by T Nomoto - [Cited by 3](#) - [Related articles](#)

2 **Supervised Ranking** with Probabilistic. Decision Tree. One technical problem associated with the use of a decision tree as a summarizer is that it is not ...

www ldc.upenn.edu/acl/P/P02/P02-1059.pdf

**Candidate Pairs:
pairs (a,b) of
comparable
examples with
different ranks**



Supervised Ranking (review)

$$\min_{\mathbf{w} \in \mathbb{R}^m} L(\mathbf{w}, P) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2$$

Warning: P is quadratic in $|D|$

- 
- A decorative header at the top of the slide features four overlapping spheres. From left to right, they are light green, light blue, light red, and light yellow. The spheres are partially cut off by the top edge of the slide.
- Joint optimization...

Combined Ranking and Regression

$$\min_{\mathbf{w} \in \mathbb{R}^m} \alpha L(\mathbf{w}, D) + (1 - \alpha) L(\mathbf{w}, P) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2$$

Combined Ranking and Regression

Regression
Term

Ranking Term

$$\min_{\mathbf{w} \in \mathbb{R}^m} \alpha L(\mathbf{w}, D) + (1 - \alpha) L(\mathbf{w}, P) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2$$

Tradeoff
Parameter

Regularization
Term

Combined Ranking and Regression

$$\min_{\mathbf{w} \in \mathbb{R}^m} \alpha L(\mathbf{w}, D) + (1 - \alpha) L(\mathbf{w}, P) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2$$

Convexity Maintained

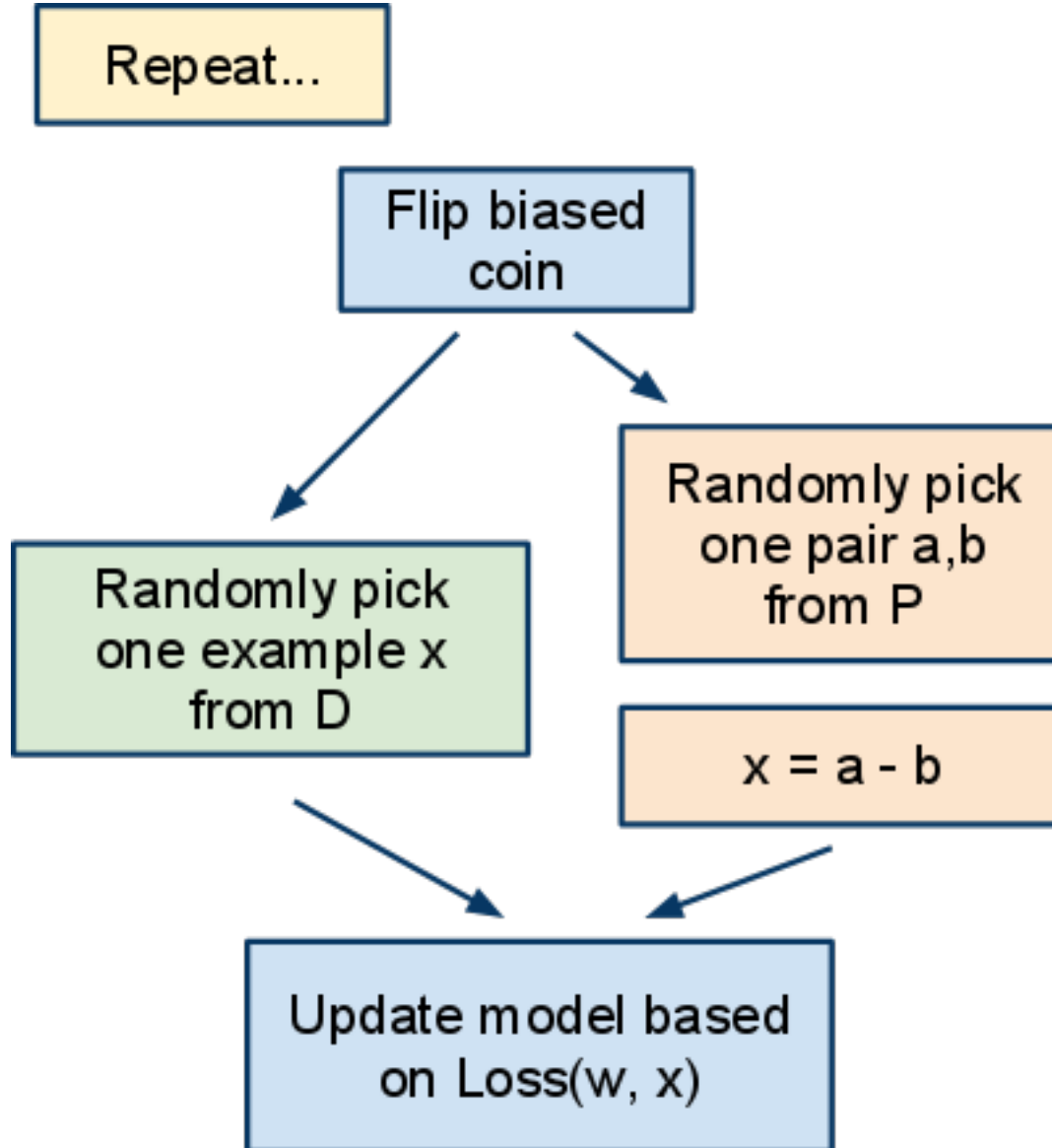


- 
- A decorative header at the top of the slide features four overlapping spheres. From left to right, they are light green, light blue, light red, and light yellow. The spheres are partially cut off by the top edge of the slide.
- What about dealing with size of P ? This is quadratic in $|D|$.

Efficient Sampling from P

- We don't want to look at $O(n^2)$ training pairs
- How to sample pairs from P ?
- Fastest solution is to index the training data:
 - $O(\log|Q| + \log|Y|)$ in general
 - $O(1)$ for common scenarios
- When data is too large to index, can use rejection sampling

Solving CRR Efficiently



Scalability



- Like other stochastic gradient descent algorithms, CRR is fast for large data
- RCV1 experiments
 - 780,000 training examples
 - Less than 3 CPU sec's on normal laptop

Non-linear Models

- CRR optimization problem is defined using a linear model w
- If we want non-linearity, use a trick from Balcan and Blum:
 - Pick a set of k reference examples r_1, \dots, r_k
 - Map each example x into a new feature space of dimension k
 - Value for feature i in new space is $\text{kernel}(x, r_i)$
- Still efficient

Experimental Overview

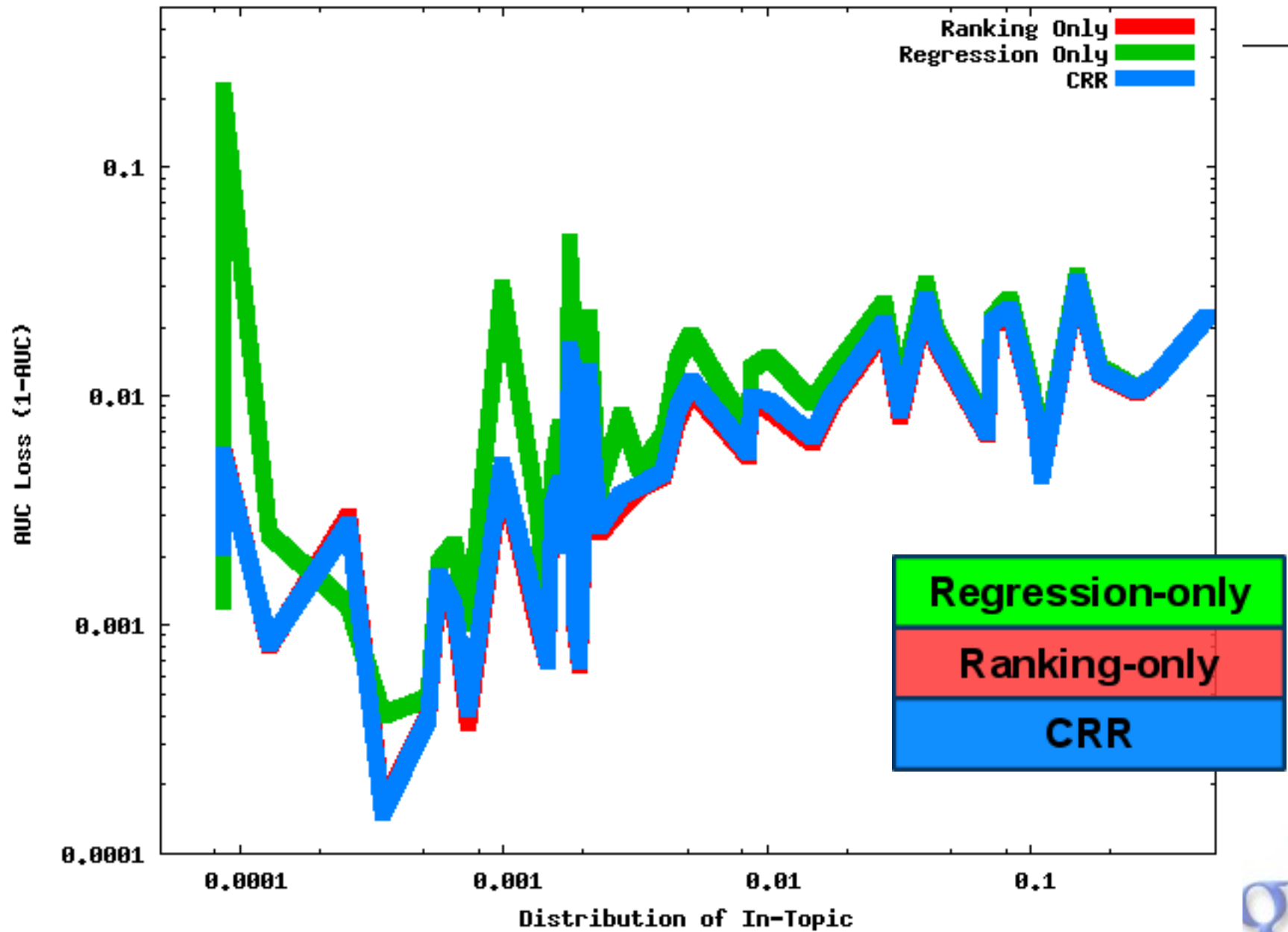
- Data sets:
 - RCV1 text classification
 - LETOR learning to rank benchmark data
 - Click prediction data for sponsored search (private)
- Comparison methods:
 - Regression-only, Ranking-only
 - Parameters tuned with cross validation on training data or on separate validation data
- Evaluation metrics:
 - Mean Squared Error (MSE)
 - AUC Loss (1 - Area Under ROC Curve)
 - Normalized Discounted Cumulative Gain (NDCG)
 - Mean Average Precision (MAP)



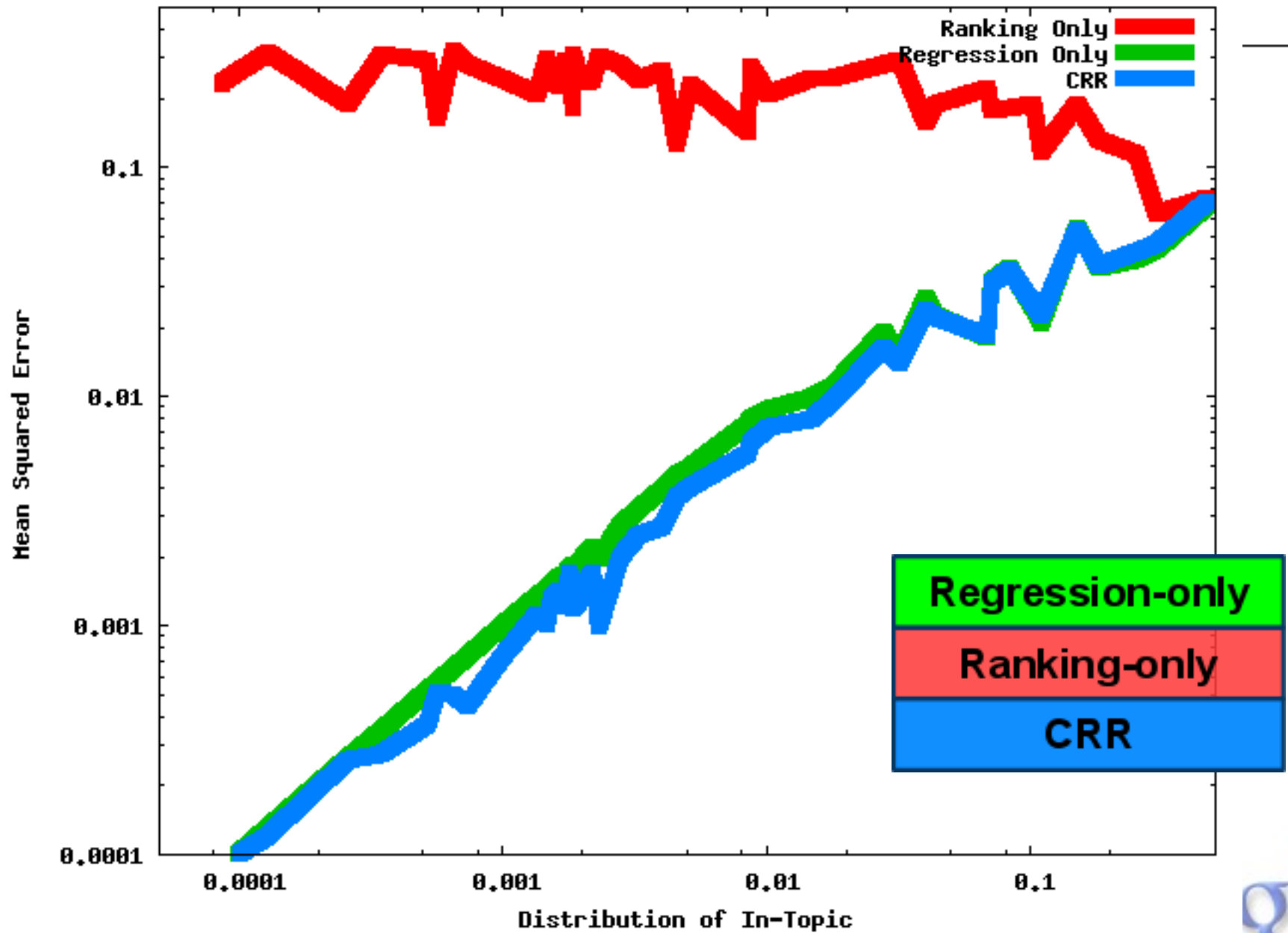
RCV1 Setup

- Benchmark text mining data set
- Tested 40 per-topic tasks
- ~780k training examples
- ~23k test examples
- ~50k sparse features
- Some topics contain extreme minority class distributions, with only 0.02% "positive"
- Used logistic loss on $\{0, 1\}$ targets

RCV1 Ranking Results



RCV1 Regression Results



RCV1 Results

- CRR achieves "best of both" metrics on 16 out of 40 tasks
 - Within 0.001 of best on 19 additional tasks
 - Always gives best performance on at least one of the two metrics
- Adding rank-based constraints can help regression:
 - CRR out-performs regression-only on MSE on 20 of 33 extreme minority class topics
 - gives equal performance on remainder

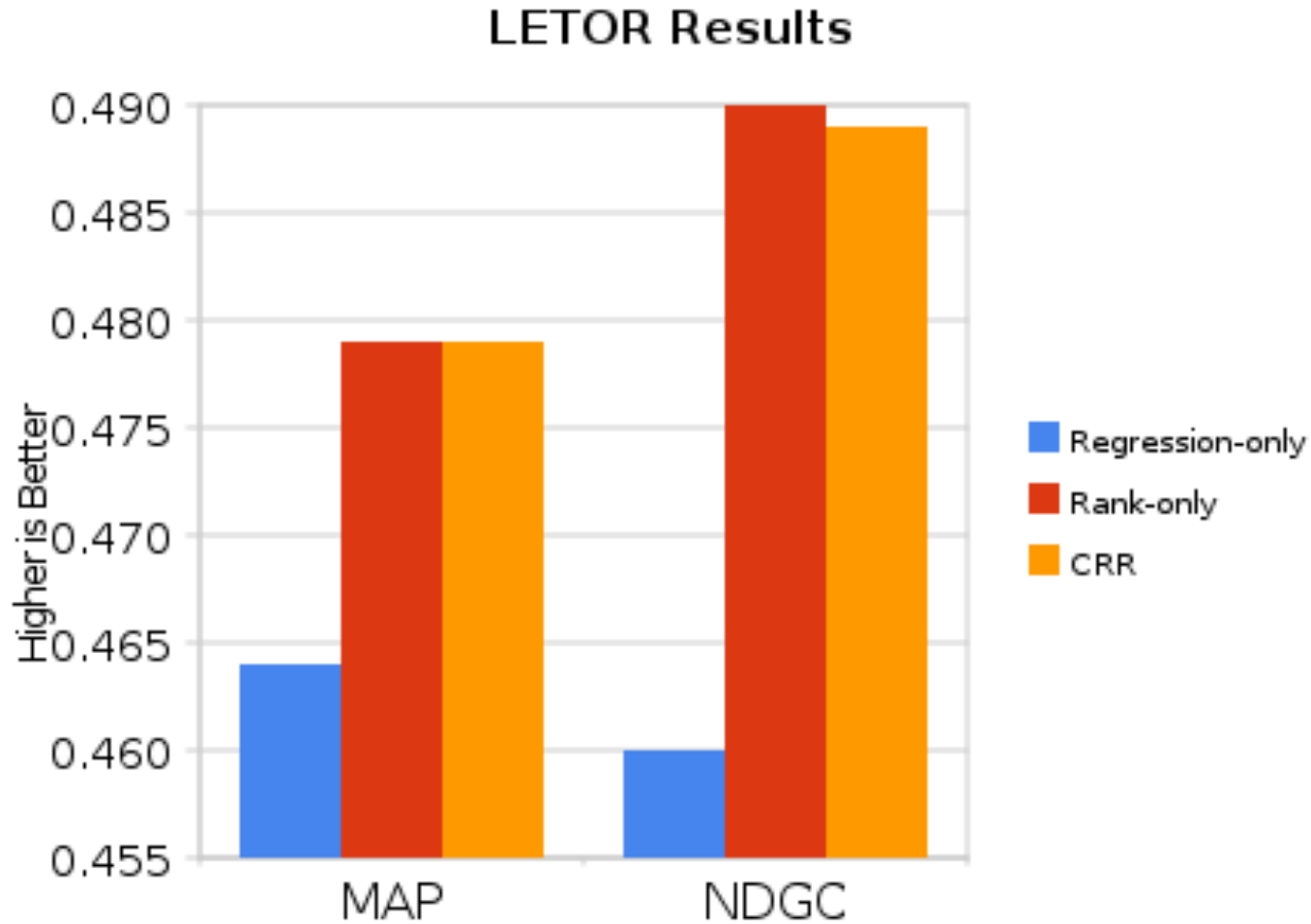
Why Would Ranking Help Regression?

- Rank-based constraints are informative, especially when observations are rare
- Imagine you had two biased coins
 - A comes up heads with probability 0.02
 - B comes up heads with probability 0.03
- Knowing that coin C is between A and B is extremely helpful if we don't have much other data

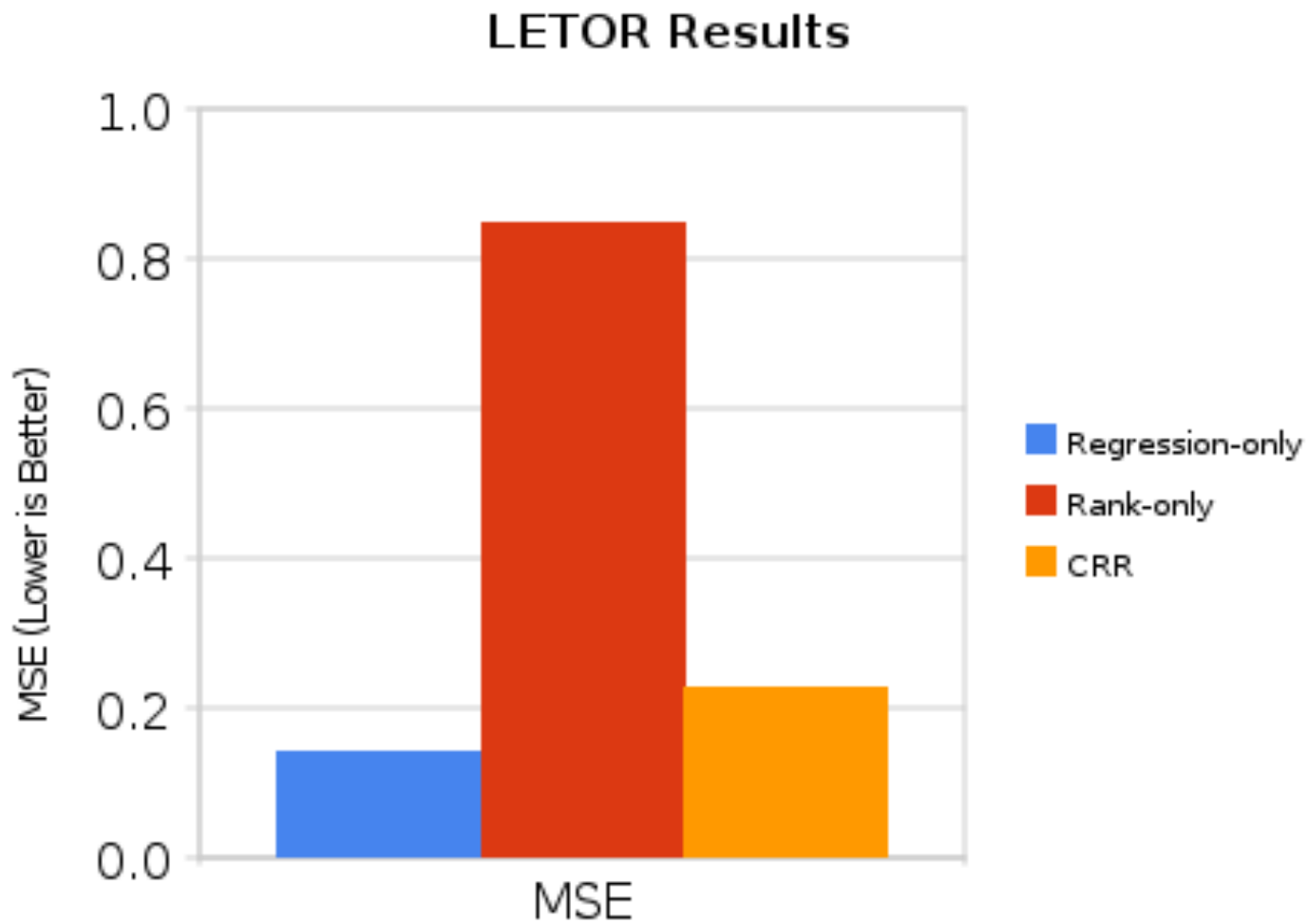
LETOR Experiments

- LETOR: benchmark learning to rank data
- Tasks with multiple relevance levels: 1, 2, or 3 stars
- Used squared loss; regression predicts ordinal values

LETOR Ranking Results



LETOR Regression Results



Click Prediction Experiments

- Test data set of several million ads
- Labels of "clicked" and "not clicked"
- Very high dimensional feature space
- Logistic loss used

Click Prediction Results

Method	Mean Sq. Error	AUC Loss
Ranking-only	0.0935	0.1325
Regression-only	0.0840	0.1334
CRR	0.0840	0.1325

Click Prediction Results

Method	Mean Sq. Error	AUC Loss
Ranking-only	0.0935	0.1325
Regression-only	0.0840	0.1334
CRR	0.0840	0.1325

Improvements
are statistically
significant

11% better
than
ranking-only

0.8% better
than
regression-
only

- 
- A decorative header at the top of the slide features four overlapping spheres. From left to right, they are light green, light blue, light red, and light yellow. The spheres are partially cut off by the top edge of the slide.
- How sensitive is the tradeoff parameter α ?

Combined Ranking and Regression

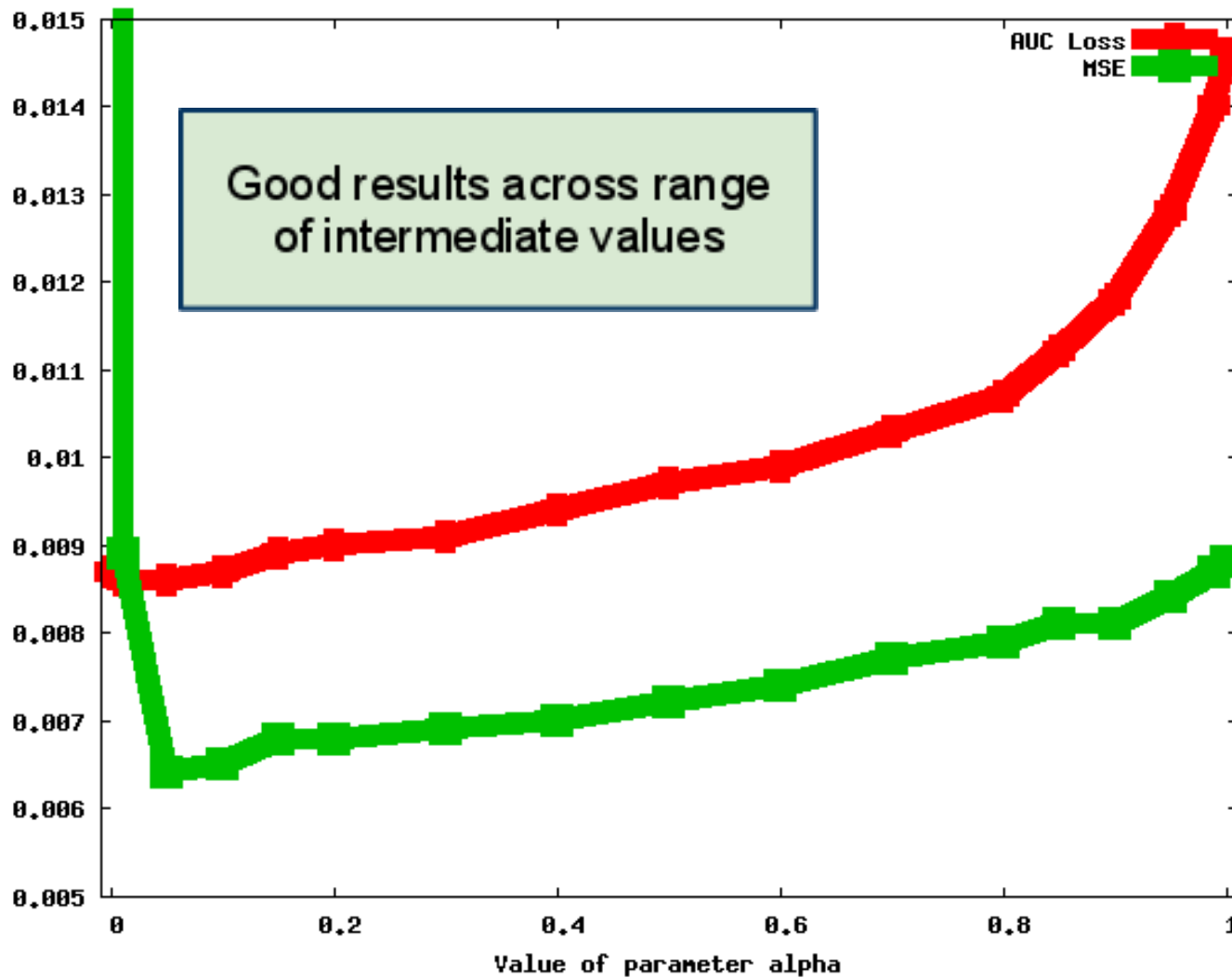
Regression
Term

Ranking Term

$$\min_{\mathbf{w} \in \mathbb{R}^m} \alpha L(\mathbf{w}, D) + (1 - \alpha) L(\mathbf{w}, P) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2$$

Tradeoff
Parameter

Looking at Tradeoff Parameter, alpha



Wrapping Up...

- Combined Ranking and Regression often gives "best of both" performance
- This algorithm uses pairwise method for rank-based component
- Simple, scalable, and robust
- Promising area for additional work
 - consider joint optimizations including MAP or NDCG optimization for ranking component

Thank you!

A decorative header featuring a horizontal line. Above the line, there are several overlapping semi-circles in light green, light blue, light red, and light yellow.

Questions?

Open Source Code: <http://code.google.com/p/sofia-ml>

Email: dsculley@google.com



RCV1

TASK	% POSITIVE	REGRESSION		RANKING		CRR	
		AUC Loss	MSE	AUC Loss	MSE	AUC Loss	MSE
E141	0.05%	0.000	0.001	0.000	0.293	0.000	0.000
GOBIT	0.06%	0.002	0.001	0.001	0.162	0.002	0.001
E61	0.06%	0.002	0.001	0.001	0.320	0.001	0.001
GTOUR	0.10%	0.030	0.001	0.005	0.245	0.005	0.001
C331	0.13%	0.003	0.001	0.001	0.205	0.001	0.001
E143	0.15%	0.001	0.001	0.001	0.296	0.001	0.001
G152	0.15%	0.005	0.001	0.003	0.239	0.003	0.001
G155	0.16%	0.007	0.002	0.004	0.223	0.004	0.001
E411	0.17%	0.002	0.002	0.002	0.289	0.002	0.001
C313	0.18%	0.047	0.002	0.014	0.281	0.016	0.002
E311	0.19%	0.001	0.002	0.001	0.311	0.001	0.001
C32	0.19%	0.019	0.002	0.012	0.180	0.013	0.002
G157	0.19%	0.001	0.002	0.001	0.254	0.001	0.001
C16	0.21%	0.022	0.002	0.012	0.234	0.013	0.002
GWELF	0.22%	0.010	0.002	0.005	0.236	0.006	0.002
E513	0.23%	0.004	0.002	0.003	0.300	0.003	0.001
E14	0.28%	0.008	0.003	0.003	0.281	0.004	0.002
C173	0.33%	0.005	0.003	0.004	0.237	0.004	0.002
E121	0.41%	0.007	0.004	0.004	0.261	0.005	0.003
GENT	0.46%	0.014	0.004	0.008	0.126	0.008	0.004
C34	0.52%	0.018	0.005	0.011	0.231	0.012	0.004
GHEA	0.85%	0.007	0.008	0.005	0.140	0.006	0.006
C183	0.87%	0.013	0.008	0.009	0.275	0.010	0.006
GDEF	1.01%	0.015	0.009	0.009	0.208	0.009	0.007
C42	1.48%	0.009	0.010	0.006	0.242	0.007	0.008
E211	1.76%	0.013	0.011	0.010	0.245	0.010	0.009
E51	2.77%	0.025	0.019	0.019	0.280	0.021	0.016
M12	3.16%	0.010	0.015	0.008	0.288	0.009	0.014
C24	3.98%	0.031	0.027	0.025	0.157	0.026	0.024
GDIP	4.34%	0.019	0.023	0.017	0.188	0.018	0.022
M13	6.89%	0.007	0.018	0.007	0.221	0.007	0.018
GPOL	7.11%	0.021	0.031	0.020	0.175	0.021	0.031
C152	8.34%	0.026	0.036	0.023	0.178	0.024	0.035
C151	10.22%	0.010	0.024	0.009	0.188	0.009	0.025
M14	10.98%	0.005	0.021	0.004	0.115	0.004	0.022
ECAT	14.90%	0.033	0.054	0.030	0.188	0.031	0.053
C15	18.05%	0.013	0.036	0.013	0.132	0.013	0.037
MCAT	25.41%	0.011	0.039	0.010	0.113	0.010	0.043
GCAT	30.11%	0.012	0.043	0.012	0.062	0.012	0.046
CCAT	46.59%	0.022	0.067	0.022	0.073	0.022	0.070



Click Prediction Results

METHOD	AdSet1	
	AUC LOSS	MSE
REGRESSION	0.133	0.084
RANKING	0.132	0.094
CRR	0.132	0.084

0.8% improvement in AUC loss with same MSE
Difference is statistically significant

