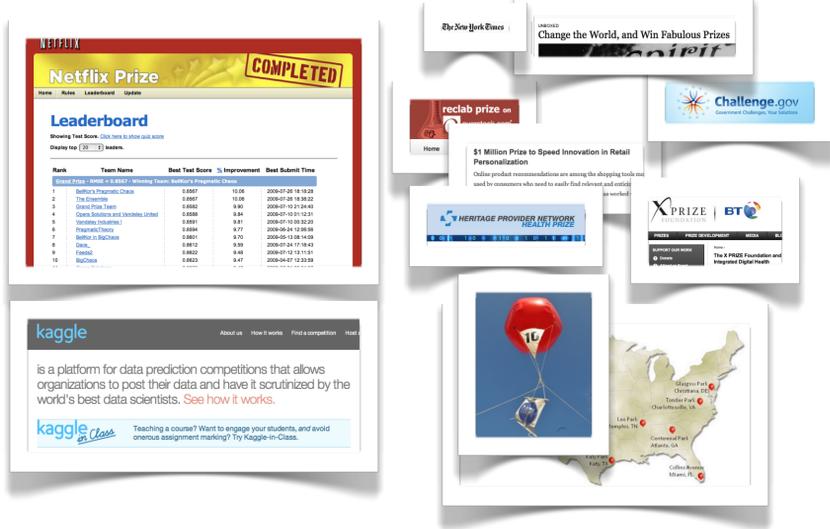


A COLLABORATIVE MECHANISM FOR CROWDSOURCING PREDICTION PROBLEMS

Jacob Abernethy (UPenn) + Rafael Frongillo (UC Berkeley)

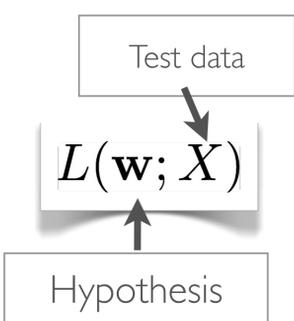
ML Competitions and their Design Problems



- **Improper payout** structure: prize money for a single benchmark, but no bonus for further improvement?
- Competition is **anti-collaborative**: competitors have no incentive to reveal their strategies or predictions
- Winner-take-all **discourages team diversity**, leading instead to team merging

Crowdsourced Learning Mechanism (CLM)!

Loss function



PHASE I:

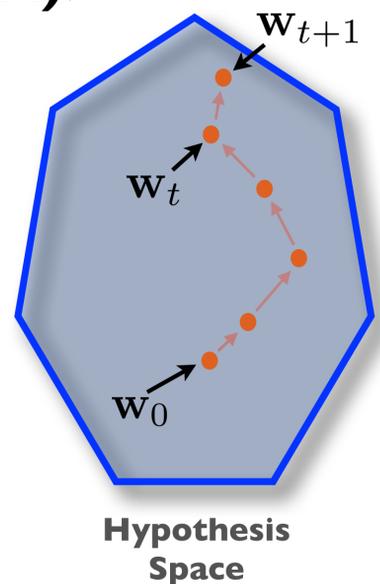
For each t while market open
 Market posts w_t
 Trader "bets" on $(w_t \rightarrow w_{t+1})$
 Trader pays $\text{Cost}(w_t \rightarrow w_{t+1})$
 Market updates to w_{t+1}
 EndWhile

PHASE 2:

Test data X is revealed

PHASE 3:

For every t , trader t earns $\text{Payout}(w_t \rightarrow w_{t+1}; X)$



Participant gets paid by the amount they improved the hypothesis on the test data:

$$\text{Profit}(w_t, w_{t+1}, X) = \text{Payout}(w_t \rightarrow w_{t+1}; X) - \text{Cost}(w_t \rightarrow w_{t+1}) = L(w_t; X) - L(w_{t+1}; X)$$

Intro to Prediction Markets

Contract:	Pays \$1 if outcome is Perry	Pays \$1 if outcome is Cain	Pays \$1 if outcome is Cookiemonster
# Shares Sold	19	17	60
Price:	\$0.04	\$0.03	\$0.93

$$\text{LMSR: } C(\mathbf{q}) := \frac{1}{\eta} \log \left(\sum_i \exp(\eta \mathbf{q}[i]) \right)$$

Share bundle: $\mathbf{r} \in \mathbb{R}^n$

Price of bundle: $C(\mathbf{q} + \mathbf{r}) - C(\mathbf{q})$

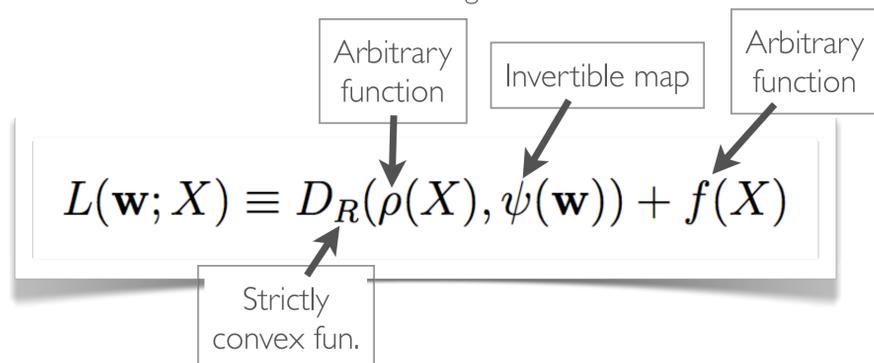
$$\text{Price vector } \nabla C(\mathbf{q}) := \left\langle \frac{\exp(\eta \mathbf{q}[i])}{\sum_j \exp(\eta \mathbf{q}[j])} \right\rangle_i$$

$$\text{Market maker risk} \leq \frac{\log N}{\eta}$$

Prediction Markets as CLMs

When is CLM \Leftrightarrow Prediction Market?

You can use an "share-based" market to implement a CLM iff the loss function can be written as a *divergence*:



The cost function is just the dual $C(\mathbf{q}) = \max_{\mathbf{p}} \mathbf{p} \cdot \mathbf{q} - R(\mathbf{p})$

CLM = Better incentives!

Benefits of share-based CLMs

- Tractable budgeted betting
- Ability to neutralize risk
- Simple betting language

Examples

- Netflix: $L(w; X) = \|w - y\|^2$
- Huffman codes: $L(p; q) = \text{KL}(q \parallel p) + H(q)$

- Offers **aligned incentives**, with payout proportional to contribution
- Mechanism has **built-in aggregation**
- Incentives **encourage diversity**, rewarding partial knowledge

Future considerations

- Too much risk?
 - Use vouchers
- Insufficient feedback?
 - Several rounds of test data
- Overly restricted model?
 - Allow additions to feature space