Scoring Rule Markets as Machine Learning Contests*

Rafael Frongillo
CU Boulder
raf@colorado.edu

Bo Waggoner
Penn
bwag@seas.upenn.edu

Abstract

A market-based framework for collaborative machine-learning contests was introduced by Abernethy and Frongillo [2011], Waggoner et al. [2015] in which participants iteratively propose updates or improvements to a central public hypothesis. This note, based on Frongillo and Waggoner [2017], investigates how the underlying choice of loss function and hypothesis necessarily impact the structure and desirable properties of such a contest. A presentation or poster at CiML would include significant background and motivation from Abernethy and Frongillo [2011], Waggoner et al. [2015] as well as new results of Frongillo and Waggoner [2017] and their implications for practice.

Despite the prevalence of winner-takes-all competitions in data science and machine learning, such as those hosted by Kaggle, DrivenData, and even NIPS, WSDM, and several other academic conferences, Abernethy and Frongillo [2011] argue that the incentives in these competitions are not necessarily aligned with the patron of the content. Instead, a *collaborative* contest paradigm is proposed, wherein participants all modify a global, publicly-viewable learning algorithm, and upon evaluation on the test set, are compensated according to how well their edits increased the relative performance of the global algorithm. Roughly speaking, the protocol is as follows:

1. Designer announces a loss function $\ell(h, x, y)$ and obtains private, secret test data set. The final goal is to obtain a hypothesis from contestants that generalizes well, which will be measured using the test set.
2. Designer announces an initial hypothesis $h^{(0)}$ to all.
3. First participant arrives and proposes an updated hypothesis $h^{(1)}$, which is announced to all.
4. $\ldots$
5. Final participant T arrives and proposes $h^{(T)}$, which is announced to all.
6. Designer releases the test set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ and rewards each participant t with improvement in empirical loss of her update:

$$\text{reward}(t) = \left(\sum_{j=1}^{n} \ell(h^{(t-1)}, x_j, y_j) \right) - \left(\sum_{j=1}^{n} \ell(h^{(t)}, x_j, y_j) \right).$$

This design is in fact based on *prediction markets* [Hanson, 2003], which use *proper scoring rules* instead of loss functions, probability distributions rather than hypotheses, and a future event rather than a test set. Indeed, subsequent advances in the prediction market literature have addressed practical considerations in this original framework. One such consideration is the "low-hanging fruit" problem: the largest drop in loss to be had is at the beginning, and each percentage point thereafter is hard-fought, yet the payout in the mechanism is proportional to the drop in loss and not the effort required to achieve it. This issue (in theory) could be solved by raising the stakes of the contest over

*Based on Frongillo and Waggoner [2017], https://arxiv.org/abs/1709.10065

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

time, which has been shown to be broadly feasible in prediction markets [Abernethy et al., 2014]. Additionally, Waggoner et al. [2015] extend the original framework to kernel methods and discuss how participants can use private data sets to propose reasonable updates to the central hypothesis.

In Frongillo and Waggoner [2017], the work that we focus on most here, we ask what "market" qualities one would like in such a mechanism, and having identified several, which loss functions yield these qualities. In particular, several loss functions yield contests closely resembling real financial markets: participants buy and sell "shares" in different parameters of the hypothesis, which results in updates that move those parameters respectively upward or downward. (Examples include the cost-function-based market framework of Abernethy et al. [2013], Abernethy et al. [2014].) These market-like mechanisms have several advantages. First, participants can make arbitrarily small or large changes (which we call the "bounded budget" axiom). Second and most important, participants who have previously made some risk-carrying update (for example, they stand to lose a significant amount if the test data set turns out a certain way) can always make a *neutralizing* update that cancels out this risk, i.e. "selling back their shares" (we call this the "trade neutralization" axiom). We would argue that at least a weak form of neutralization is crucial for a well-run collaborative contest, as it implies that new information or techniques can always be incorporated to improve one's position—in essence, it guarantees that participants can never be "stuck" by their previous submissions.

The central question we ask, therefore, is which loss functions yield frameworks with these properties, and in particular, which allow for trade neutralization. To answer this, we use tools from *information elicitation* [Lambert et al., 2008], which formally investigates the relationship between loss functions and the statistics elicited by them. For example, squared loss $\ell(h,x,y) = (h(x)-y)^2$ "elicits the mean", as do Bregman divergences more generally; absolute loss $\ell(h,x,y) = |h(x)-y|$ elicits the median, and so on. Formally, if one has a conditional distribution $\mathcal{D}$ on y, then the *property* of $\mathcal{D}$ elicited by the loss function $\ell(r,y)$ is denoted $\Gamma(\mathcal{D})$ and is defined as

$$\Gamma(\mathcal{D}) = \arg\min_{r} \mathop{\mathbb{E}}_{y\sim\mathcal{D}} \ell(r,y).$$

Our main result characterizes markets that satisfy trade neutralization, showing that they use Bregman divergences and elicit an expectation, or are some discretized version of these. We also show that any scoring rule market for the mode or indeed any "finite property" (categorical hypothesis) cannot satisfy either bounded budget or trade neutralization; while any market for the median or indeed any quantile can satisfy bounded budget, but not trade neutralization. Finally, while the main characterization is somewhat restrictive, implying that only loss functions which elicit the mean have these nice properties, we also give several examples of losses which satisfy a weaker version of the axiom call "weak neutralization", which practically speaking is just as useful.

One such example of losses satisfying weak neutralization are those which elicit a *ratio of expectations* $\Gamma(\mathcal{D}) = \mathbb{E}_{y\sim\mathcal{D}} f(y)/\mathbb{E}_{y\sim\mathcal{D}} g(y)$. A classic example is the squared relative error, $L(r,y) = (r-y)^2/r^2$, which elicits the ratio of second and first moments of $\mathcal{D}$, i.e., $f(y) = y^2, g(y) = y$. After showing that the contest framework resulting from such losses satisfy weak neutralization, we examine the contest more closely through the lens of convex conjugate duality. We find the following natural interpretation of such a contest. In traditional markets, one buys and sells a security $f(\cdot)$, worth $\$f(y)$ upon outcome y, in exchange for money. A purchase of 1 unit of f for $\$c$ therefore expresses the belief $\mathbb{E}[f] \geq c$, and similarly a sale at the same price expresses the belief $\mathbb{E}[f] \leq c$, whence we conclude the equilibrium market price should be the market consensus belief about the value of $\mathbb{E}[f]$. In the ratio of expectations market, one buys or sells the security f in exchange for some units of *another security* g. Analogously then, a purchase of 1 unit of f for $c \cdot g$ expresses the belief $\mathbb{E}[f] \geq c\,\mathbb{E}[g]$, and similarly a sale at the same "price" expresses the belief $\mathbb{E}[f] \leq c\,\mathbb{E}[g]$, so the market price should reflect the consensus belief about $\mathbb{E}[f]/\mathbb{E}[g]$.

In conclusion, we study in Frongillo and Waggoner [2017] the collaborative machine learning contest framework proposed by Abernethy and Frongillo [2011], which conveys several advantages over the traditional winner-takes-all paradigm. In particular, we ask for which loss functions this framework gives a "fluid" mechanism, where traders can make arbitrarily small adjustments to the global hypothesis or algorithm, and are always able to recover from mistakes in previous adjustments. While motivated by deepening our understanding of loss functions and prediction markets, we believe our work has relevance to the practical implementation of collaborative machine learning contests.

References

Jacob Abernethy, Yiling Chen, and Jennifer Wortman Vaughan. Efficient market making via convex optimization, and a connection to online learning. *ACM Transactions on Economics and Computation*, 1(2):12, 2013.

Jacob D. Abernethy and Rafael M. Frongillo. A collaborative mechanism for crowdsourcing prediction problems. In *Advances in Neural Information Processing Systems 25*, NIPS '11, pages 2600–2608, 2011.

Jacob D. Abernethy, Rafael M. Frongillo, Xiaolong Li, and Jennifer Wortman Vaughan. A general volume-paramterized market making framework. In *Proceedings of the 15th Conference on Economics and Computation*, EC '14, 2014.

Rafael Frongillo and Bo Waggoner. An axiomatic study of scoring rule markets. 2017. URL https://arxiv.org/abs/1709.10065.

Robin Hanson. Combinatorial information market design. *Information Systems Frontiers*, 5(1): 107–119, 2003.

Nicolas S Lambert, David M Pennock, and Yoav Shoham. Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, EC '09, pages 129–138. ACM, 2008.

Bo Waggoner, Rafael Frongillo, and Jacob D. Abernethy. A market framework for eliciting private data. In *Advances in Neural Information Processing Systems 28*, NIPS '15, pages 3492–3500, 2015.