Gini regularized Optimal Transport with an application to Spatio-Temporal Forecasting

Lucas Roberts, Leo Razoumov, Lin Su*, Yuyang Wang

Amazon.com Inc and NCSU*

NIPS 2017 Optimal Transport Workshop

December 9th, 2017

Context

Lucas Roberts, Leo Razoumov, Lin Su*, Yuyang Wang

Problem Background
Experiments
Take Aways

1. Global vs Local sales forecasts
2. $MAWE = \sum_{i \in T} |\hat{F}_i - s_i| w(s_i)$
3. Where $\hat{F}_i$ is the forecast and s_i is the sales.
4. All errors are the same cost.
5. Goal is: design a metric with the business cost of errors.

Context

1. Global vs Local sales forecasts
2. $MAWE = \sum_{i \in T} |\hat{F}_i - s_i| w(s_i)$
3. Where $\hat{F}_i$ is the forecast and s_i is the sales.
4. All errors are the same cost.
5. Goal is: design a metric with the business cost of errors.

Optimal Transport Metrics

Given an $n \times k$ cost matrix $\boldsymbol{M}$, the optimal transport problems are:

1. $\min_{\boldsymbol{P} \in U(\boldsymbol{\mu}, \boldsymbol{\nu})} \langle \boldsymbol{P}, \boldsymbol{M} \rangle = tr(P^T M)$,
   $U(\boldsymbol{\mu}, \boldsymbol{\nu}) = \{\boldsymbol{P} \in \mathbb{R}_+^{n \times k} | \boldsymbol{P}\mathbf{1}_n = \boldsymbol{\mu}, \boldsymbol{P}^T \mathbf{1}_k = \boldsymbol{\nu}\}$,
2. $\min_{\boldsymbol{P} \in U(\boldsymbol{\mu}, \boldsymbol{\nu})} \langle \boldsymbol{P}, \boldsymbol{M} \rangle - \frac{1}{\lambda} R(\boldsymbol{P})$,

where $R(\boldsymbol{P})$ is either $\mathcal{G}(\boldsymbol{P}) = \sum_{i=1}^n \sum_{j=1}^k P_{ij}(1 - P_{ij})$ (Gini) or $h(\boldsymbol{P}) = -\sum_{i=1}^n \sum_{j=1}^k P_{ij} \log P_{ij}$ (entropy) Here the convention is $0 \log(0) = 0$.

These metrics are: Classical OT (EMD), Entropy regularized OT, and Gini regularized OT

Gini OT Experiments

1. Any QP solver will work with Gini regularizer to OT problem.
2. We examined
    1. CVX a conic solver
    2. Conditional Gradient (Frank-Wolfe) solver
    3. Mirror Descent (MD)*
3. Rewrite Gini OT: $\min_{\boldsymbol{P} \in U(\boldsymbol{\mu}, \boldsymbol{\nu})} \langle \boldsymbol{P}, \boldsymbol{M}_\lambda \rangle + \frac{1}{\lambda} \|\boldsymbol{P}\|_F^2 \doteq f(\boldsymbol{P})$, where $\boldsymbol{M}_\lambda = \boldsymbol{M} - \mathbf{1}_n \mathbf{1}_k^T / \lambda$. and $\| \cdot \|_F$ denotes the Frobenius norm.

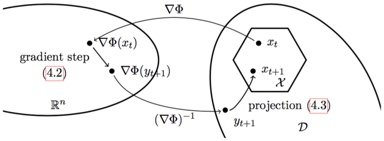

Figure 4.1: Illustration of mirror descent.

Figure: A graphical depiction of Mirror Descent. (taken from Bubeck Convex Optimization book)

Comparison of Approximations

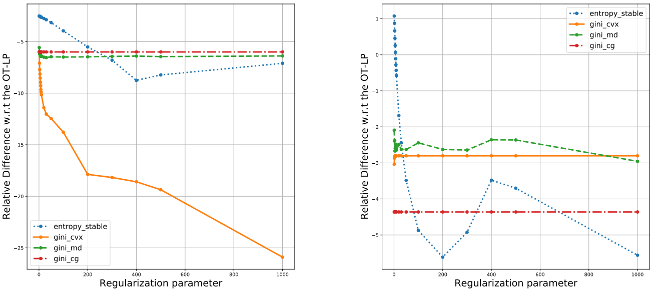

Figure: Accuracy comparison:
relative difference with respect to the OT-LP (in log scale) vs. λ for (left) normalized random uniform; (right) logit-normal.

Scales Well (In The Number of Locales)

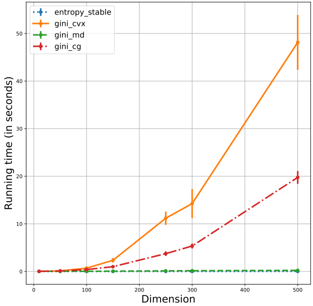

Figure: Scaling comparison:
for each dimension, we repeat the experiments 10 times and report the error bars.

Take Home Points

1. Classical Forecast metrics do not consider the locale of sales.
2. Optimal Transport is a framework for local Accuracy criteria.
3. Use Mirror Descent with Gini OT.
4. Gini OT is insensitive to λ.

For more please come to our Poster