

Active Online Domain Adaptation

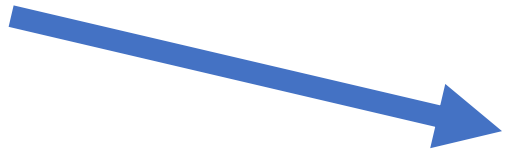
Yining Chen (Stanford), Haipeng Luo (USC), Tengyu Ma (Stanford), Chicheng Zhang
(University of Arizona)

Domain shift: challenge for ML system



Online learning: a classical framework for domain shift

- At timestep $t = 1, \dots, T$:
 - Input x_t is revealed.
 - Learner predicts \hat{y}_t , suffers some loss.
 - Label y_t is revealed.

 Expensive!

-> This work: **active** online learning

Learner can decide whether to query the label.

Active online regression: Setup

- Hypothesis class \mathcal{F}
- Realizable setting: $y_t = f^*(x_t) + \xi_t$
- At timestep $t = 1, \dots, T$:
 - x_t is revealed.
 - Learner predicts \hat{y}_t , suffers loss $(y_t - \hat{y}_t)^2$.
 - Learner decides whether to query y_t .
- Metrics: (1) # queries Q (2) Regret $R = \sum_{t=1}^T (\hat{y}_t - f^*(x_t))^2$
- Goal: Minimize R subject to $Q \leq B$.

Linear regression

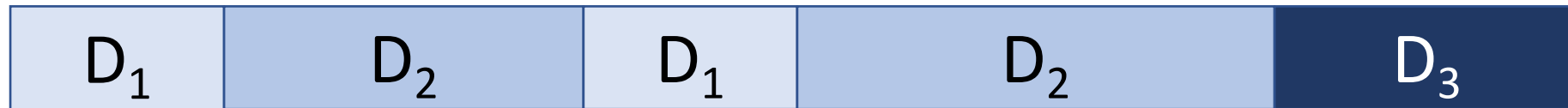
- Hypothesis class: $\mathcal{F} = \{\theta^\top x : \theta \in \mathbb{R}^d\}$
- Prediction strategy:
 - Follow the regularized leader on all **queried** examples till t-1:

$$\hat{\theta}_t = \operatorname{argmin}_{\theta} \sum_{i < t, q_i = 1} (\theta^\top x_i - y_i)^2 + \lambda \|\theta\|^2$$

- What's the query strategy?
 - Uniformly random queries are optimal without domain structure.

Can we do better if the data is structured?

- Domain structure:



- m domains
- Domain u : support dimension d_u , duration T_u

Main contribution: optimal tradeoff with **unknown** domain structure

- QuFUR (Query in the Face of Uncertainty for Regression):
 - Query **uncertain** examples
- Linear Regression: For **any** domain partition, with high probability:

$$R = \tilde{O}\left(\left(\sum_u \sqrt{d_u T_u}\right)^2 / B\right).$$

- Matching lower bound.
- Generalize to general hypothesis classes w/ bounded Eluder Dimension.