Nuclear Discrepancy for Single-Shot Batch Active Learning

Tom J. Viering^{@*} Jesse H. Krijthe^{*} Marco Loog^{*†}

[@]t.j.viering@tudelft.nl, *TU Delft, [†]University of Copenhagen

In active learning, an algorithm tells you which samples to annotate. By selecting 'smart' samples the model learns faster.

Single Shot Batch Active Learning

Input: labeling budget n, unlabeled pool \hat{P} 1. Active learner selects $\hat{Q} \in \hat{P}$ such that $sim(\hat{P}, \hat{Q})$ minimal must be quadratic. For kernel: $K_{MMD}(x_i, x_j) = K_{model}(x_i, x_j)^2$. Now MMD takes loss and model class into account.

New result: now the Discrepancy bound is tighter under a worst-case assumption. Better AL performance? No! Why not?

Probabilistic Analysis

- 2. Request labels for \hat{Q}
- 3. Train Kernel Ridge Regression model on \hat{Q}
- 4. Evaluate model on unseen test set

Domain Adaptation Bounds for AL

 $L_P(h, f) \le L_{\hat{Q}}(h, f) + \sin(\hat{P}, \hat{Q}) + C + \eta$

- h is trained model, f is (unknown) true labeling function
- $L_P(h, f)$ loss on (unknown) distribution P
- $L_{\hat{O}}(h, f)$ loss on training set
- C complexity (e.g. VCdim), η model misspecification
- $sim(\hat{P}, \hat{Q})$ similarity measure

Domain Similarity Measures sim (\hat{P}, \hat{Q})

- Discrepancy [Mansour 2009]
- Maximum Mean Discrepancy (MMD) [Gretton 2012] • Nuclear Discrepancy (proposed)

Let $M = \frac{1}{n_{\hat{P}}} X_{\hat{P}}^T X_{\hat{P}} - \frac{1}{n_{\hat{O}}} X_{\hat{Q}}^T X_{\hat{Q}}$, and e_1, \ldots, e_d be the eigenvectors such that $|\lambda_1| \geq \ldots \geq |\lambda_d|$. Let u = h - f and $v_i = u^T e_i$. The Discrepancy is given by $disc(\hat{P},\hat{Q}) = 4\Lambda^2 |\lambda_1|$. In the worst case $u \propto e_1$: meaning we only consider a very specific u. The following should be small for good performance:

$$|L_{\hat{P}}(h,f) - L_{\hat{Q}}(h,f)| = |u^T M u| \le \sum_{i}^{d} |\lambda_i| v_i^2$$

All eigenvalues λ_i should be minimized if $u \neq e_1!$ Furthermore, for the Discrepancy, the true labeling function f depends on our choice of sample \hat{Q} . This assumption is too pessimistic for AL.

Solution: Nuclear Discrepancy

We build a probabilistic bound where $u \sim p(u)$:

- From symmetry arguments p(u) should be symmetric
- Our choice of samples should not influence f or p(u)

We choose p(u) to be uniform on a ball centered on the origin. In that case, one should optimize the Nuclear Discrepancy:

Which similarity measure is best for AL and why?

Setting

Measure performance in terms of mean squared error (MSE). We use the Kernel Ridge Regression model and indicate the model class by $H = \{h \in \mathcal{H} : ||h||_K \leq \Lambda\}$ (\mathcal{H} is the RKHS).

Theoretical Analysis

To derive bounds, we need to bound the following quantity: $|L_{\hat{P}}(h, f) - L_{\hat{O}}(h, f)|$. Assume $f \in \mathcal{H}$. Discrepancy bound:

 $\max_{h, f \in \mathcal{H}} |L_{\hat{P}}(h, f) - L_{\hat{Q}}(h, f)| = \mathsf{disc}(\hat{Q}, \hat{P})$

Note it depends on model class and loss function! We transform the MMD to also depend on model class and loss function:

$$\mathsf{ND}(\hat{P}, \hat{Q}) = 4\Lambda^2 \sum_{i}^{d} |\lambda_i|$$

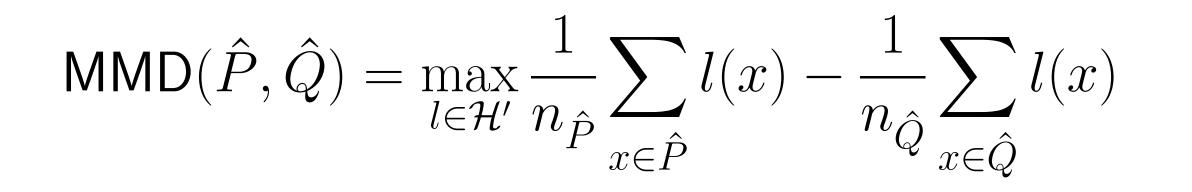
Experiments and Results

- Evaluate on 15 datasets
- Performance: Area Under the MSE Learning Curve (AULC)
- Preprocess data s.t. $\eta = 0$ (realizeable)

Bound \setminus Setting	Worst-case	Our $p(u)$	Performance
Discrepancy	Tightest	Loosest	Worst
MMD	Medium	Medium	Medium
ND (ours)	Loosest	Tightest	Best
• Our assumption on $p(u)$ explains observed performance!			

Discussion and Conclusion

• Further investigation needed to understand agnostic case $(\eta \neq 0)$, but similar trend observed



l can approximate the loss: $(h(x) - f(x))^2$. We choose \mathcal{H}' so it contains the loss function. Example: if h, f are linear in x, \mathcal{H}'

- Bound tightness not most important, accurate assumptions are just as important
- MMD with squared loss: squared kernel is a natural choice





Code available at https://github.com/tomviering/NuclearDiscrepancy

Pattern Recognition Laboratory: http://prb.tudelft.nl