Hierarchical Label Queries With Data-Dependent Partitions

Samory Kpotufe, Ruth Urner, Shai Ben-David
Princeton University, Max Planck Institute, University of Waterloo
Overview

Setting
Classification with expensive labels.

Keywords
Active Learning, Cluster-Assumption, Partitioning trees.

Motivation
Refined analysis of a practical procedure in practical settings.
Overview

Setting
Classification with **expensive labels**.

Keywords
**Active Learning**, **Cluster-Assumption**, Partitioning trees.

Motivation

Refined analysis of a practical procedure in practical settings.
Overview

Setting
Classification with expensive labels.

Keywords
Active Learning, Cluster-Assumption, Partitioning trees.

Motivation
Refined analysis of a practical procedure in practical settings.
Overview

Setting
Classification with **expensive labels**.

Keywords
**Active Learning**, **Cluster-Assumption**, Partitioning trees.

Motivation

Refined analysis of a practical procedure in practical settings.
Partition unlabeled $X_1^n$, query a few labels in each cell. LABEL pure cells, PARTITION impure cells; REPEAT

Label data with error $< \epsilon \implies$ now use supervised learner.

Overall Appeal
A-L: Implementable, C-A: needs only hold approximately. SAFE.
Hierarchical Labeling: Dasgupta and Hsu 2008

Partition unlabeled $X^n_1$, query a few labels in each cell. LABEL pure cells, PARTITION impure cells; REPEAT

Label data with error $< \epsilon \implies$ now use supervised learner.

Overall Appeal

A-L: Implementable, C-A: needs only hold approximately. SAFE.
Hierarchical Labeling: Dasgupta and Hsu 2008

Partition unlabeled $X^n_1$, query a few labels in each cell. LABEL pure cells, PARTITION impure cells; REPEAT

Label data with error $< \epsilon \implies$ now use supervised learner.

Overall Appeal

A-L: Implementable, C-A: needs only hold approximately. SAFE.
**Labeling Goal:** $\leq 1/\epsilon^2$ label-complexity of agnostic-learning.

**Guarantees on Label queries:** from $|T|/\epsilon$ to $1/\epsilon^2$

Depends on niceness of $P_{X,Y}$, and $|T| \equiv$ Data-quantization rate.

**Earlier results (similar label guarantees)**

- [Das., Hsu, 08]: Niceness of sample $X_1^n, Y_1^n$.
- [Urn., Wulff, B-Dav, 13]: Niceness of $P_{X,Y}$, no noise in $Y$, partition $T \perp X_1^n$.

**This result:** (more practical assumptions)

Niceness of $P_{X,Y}$, low noise in $Y$, $T = T(X_1^n) \implies$ smaller $|T|$.
Labeling Goal: $\leq 1/\epsilon^2$ label-complexity of agnostic-learning.

Guarantees on Label queries: from $|T|/\epsilon$ to $1/\epsilon^2$

Depends on niceness of $P_{X,Y}$, and $|T| \equiv$ Data-quantization rate.

Earlier results (similar label guarantees)

- [Das., Hsu, 08]: Niceness of sample $X^n_1, Y^n_1$.
- [Urn., Wulff, B-Dav, 13]: Niceness of $P_{X,Y}$, no noise in $Y$, partition $T \perp X^n_1$.

This result: (more practical assumptions)

Niceness of $P_{X,Y}$, low noise in $Y$, $T = T(X^n_1) \implies$ smaller $|T|$. 
Labeling Goal: $\leq 1/\epsilon^2$ label-complexity of agnostic-learning.

Guarantees on Label queries: from $|T|/\epsilon$ to $1/\epsilon^2$

Depends on niceness of $P_{X,Y}$, and $|T|$ ≡ Data-quantization rate.

Earlier results (similar label guarantees)

- [Das., Hsu, 08]: Niceness of sample $X_1^n, Y_1^n$.
- [Urn., Wulff, B-Dav, 13]: Niceness of $P_{X,Y}$, no noise in $Y$, partition $T \perp X_1^n$.

This result: (more practical assumptions)

Niceness of $P_{X,Y}$, low noise in $Y$, $T = T(X_1^n) \implies$ smaller $|T|$. 
Labeling Goal: $\leq 1/\epsilon^2$ label-complexity of agnostic-learning.

Guarantees on Label queries: from $|T|/\epsilon$ to $1/\epsilon^2$

Depends on niceness of $P_{X,Y}$, and $|T| \equiv$ Data-quantization rate.

Earlier results (similar label guarantees)

- [Das., Hsu, 08]: Niceness of sample $X_1^n, Y_1^n$.
- [Urn., Wulff, B-Dav, 13]: Niceness of $P_{X,Y}$, no noise in $Y$, partition $T \perp X_1^n$.

This result: (more practical assumptions)

Niceness of $P_{X,Y}$, low noise in $Y$, $T = T(X_1^n) \implies$ smaller $|T|$.
Merci! ... See you at the poster!