Introduction to Active Learning

Thursday, May 20, 2021

Presented by: Yongchao Zhou

VECTOR INSTITUTE

Outline

- What is active learning?
- What kind of examples are most informative?
- How does AL fit into the ML workflow?

What is Active Learning?

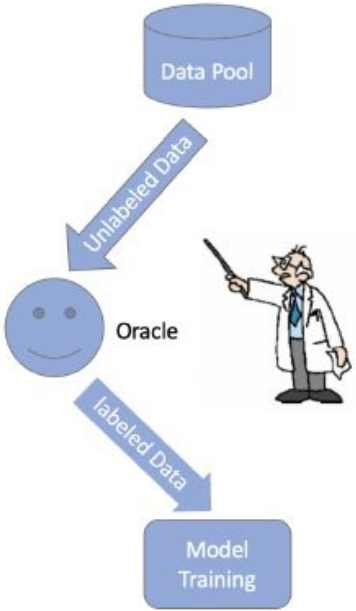

Figure 1: Passive Machine Learning

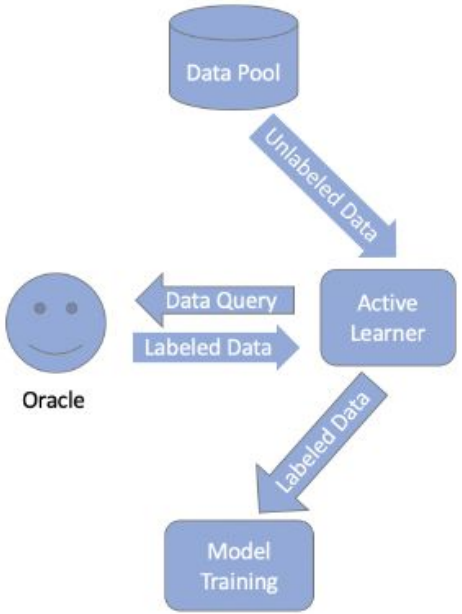

Figure 2: Active Machine Learning

Active Learner & Label Efficiency

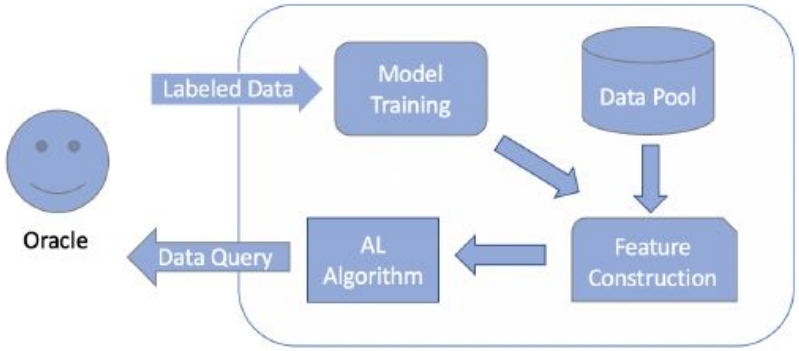

Figure 3: Active Learner

- AL's goal: Build a **high-quality dataset** such that the model trained on it can achieve as high test accuracy as possible.

Typical Results

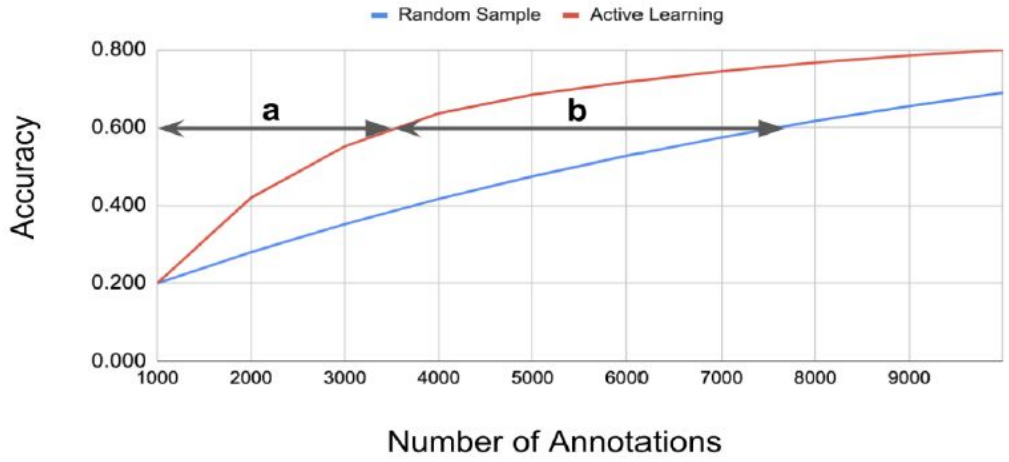

Figure 4: Typical Active Learning Result

- Given a target accuracy, the AL algorithm reduces the labeling effort by x%.
- Given a fixed labeling budget, the AL algorithm improves the performance by x%.

What kind of examples are most informative?

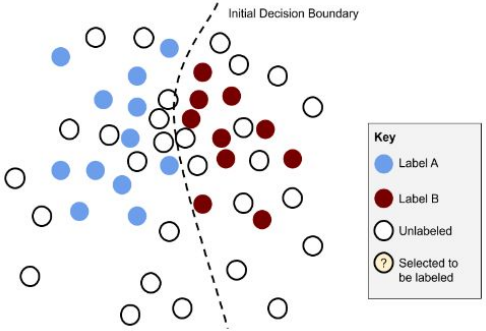

Figure 5: Current Decision Boundary

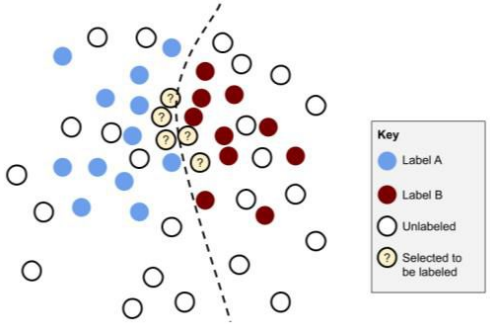

Figure 6: Uncertainty Sampling

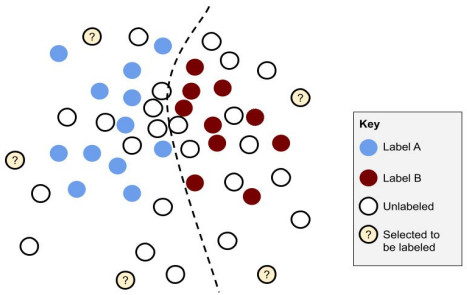

Figure 7: Diversity Sampling

Uncertainty Sampling

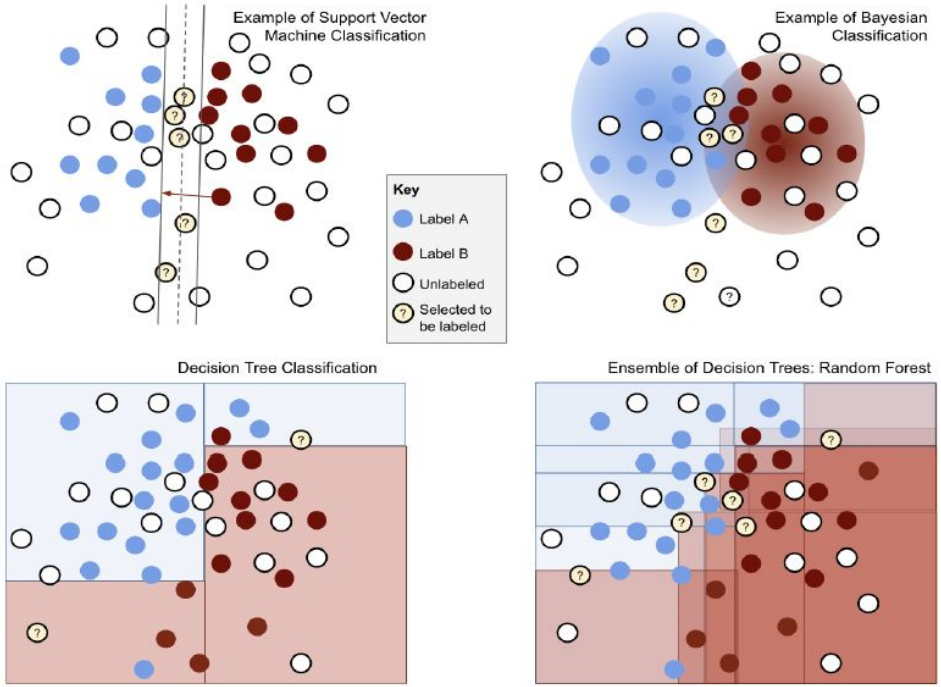

Figure 8: Uncertainty Sampling for different ML Models

Diversity Sampling

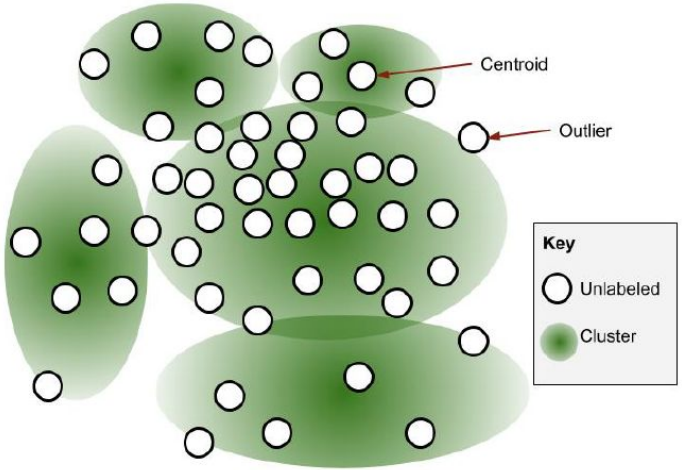

Figure 10: Cluster-based Sampling

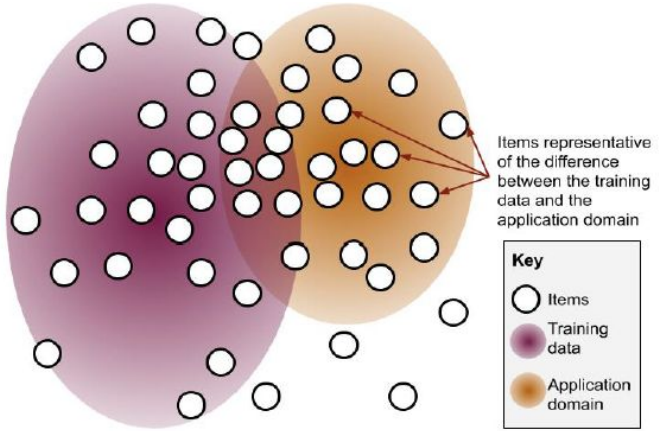

Figure 11: Representative Sampling

Hybrid Method

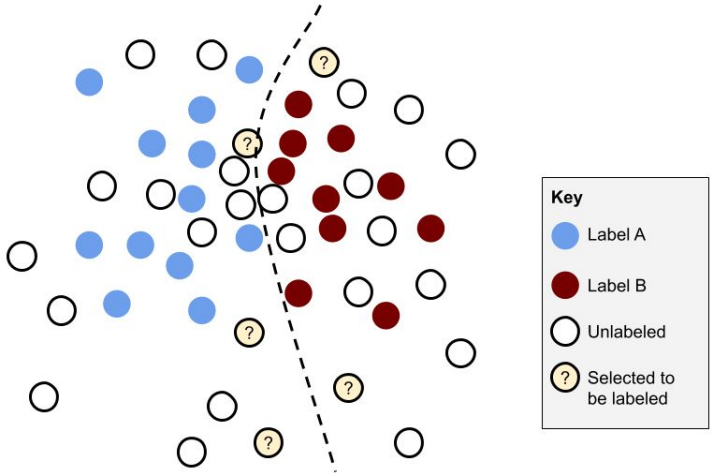

Figure 12: Hybrid Method

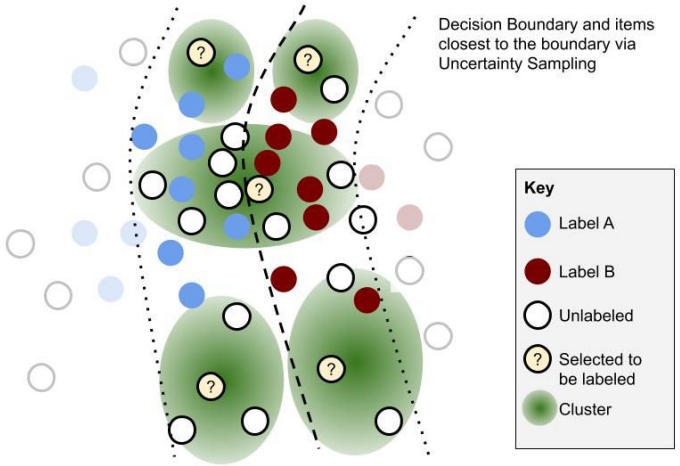

Figure 13: Uncertainty + Clustering

Active Learning for Neural Networks

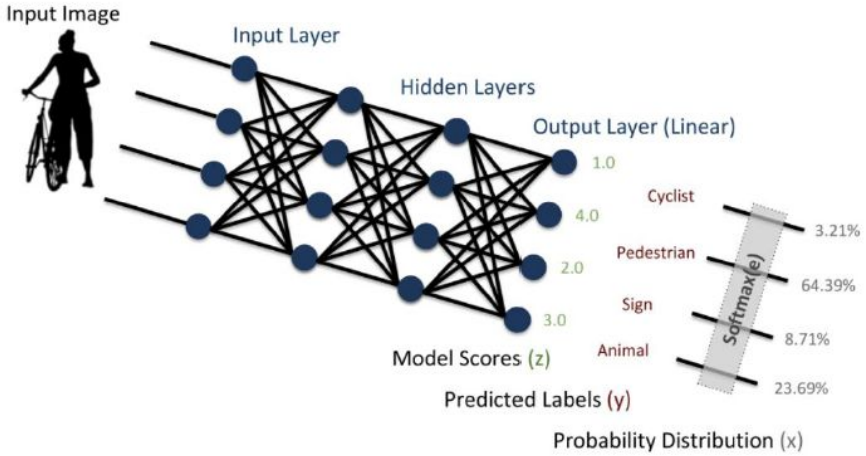

Figure 9: Neural Network Visualization

- Prediction-based
  - Least Confidence
  - Max Entropy
  - Margin Sampling
  - Bayes-Coreset
- Model-based
  - Expected Parameter Change
  - Maximum Variance Reduction
  - Adversarial
  - Coreset
  - BADGE
- Ensemble-based
  - Query by Committee
  - Bayesian Disagreement

Pros and Cons of different AL algorithm

Table 1: Pros and Cons of different AL algorithms

AL Type	Pros	Cons
Uncertainty Sampling	Simple to implement	Select redundant points
Diversity Sampling	Good sample diversity	Select easy points
Hybrid Method	Consider both uncertainty and diversity	Poor scalability

Active Learning - An iterative process

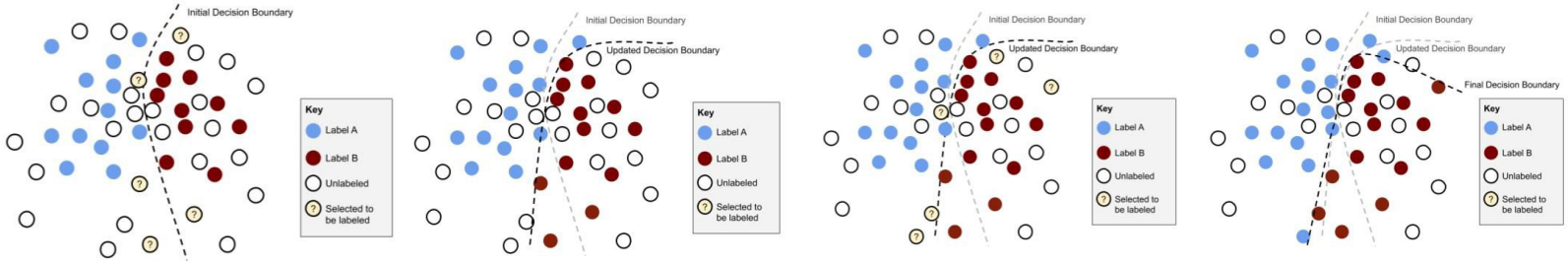

Figure 14: Active Learning Process

- Step 1: Apply Active Learning to sample items that require a human label to create additional training items.
- Step 2: Retrain the model with the new training items, resulting in a new decision boundary.
- Step 3: Apply Active Learning again to select a new set of items that require a human label.
- Step 4: (and beyond): Retrain the model again, and repeat the process to keep getting a more accurate model.

How does AL fit into the ML workflow?

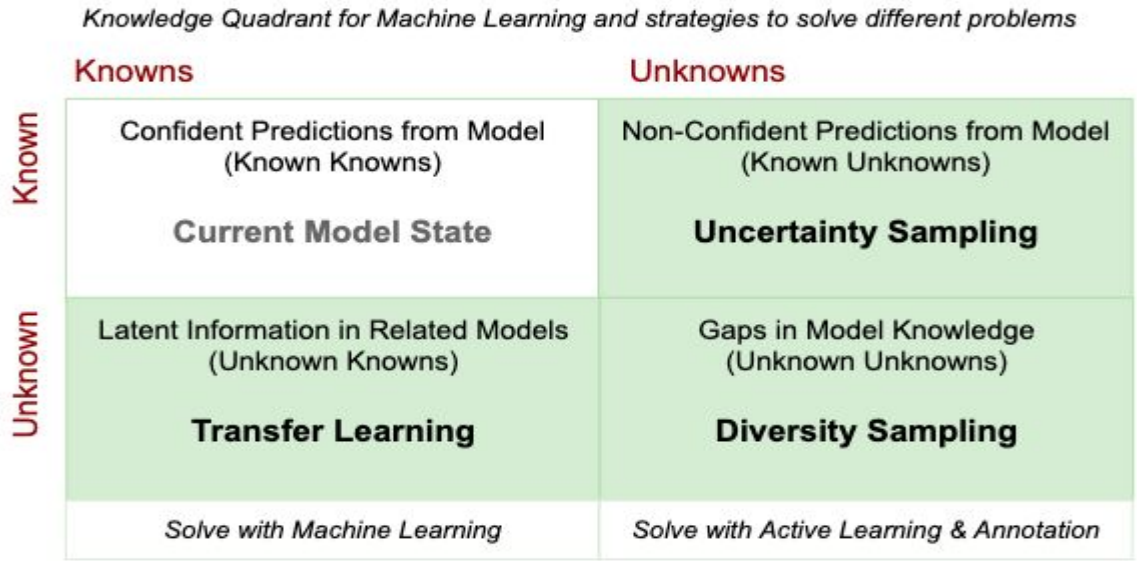

Figure 15: Knowledge Quadrant for ML

Human-in-the-Loop Machine Learning

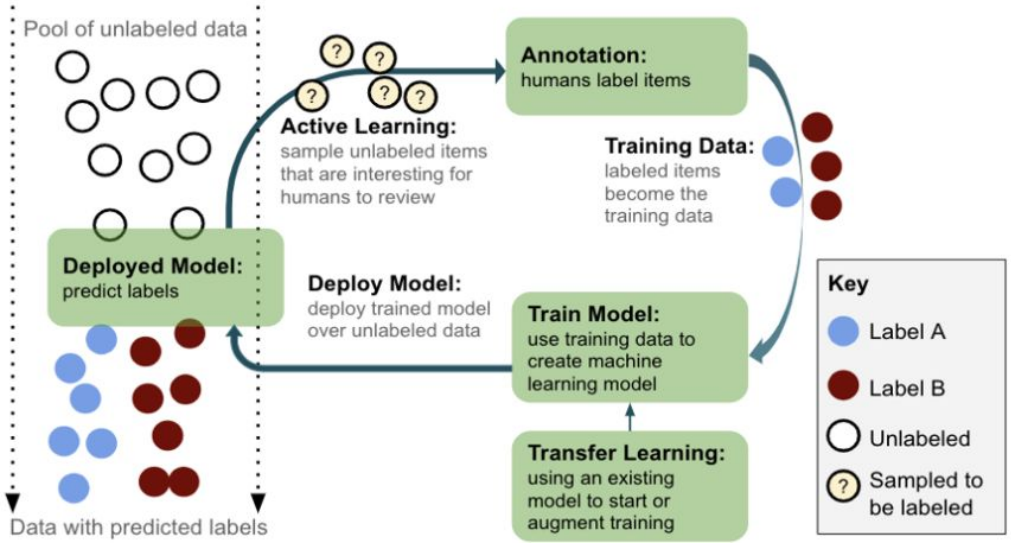

Figure 16: Human-in-the-loop Machine Learning Workflow

Questions?

Thank you!

VECTOR INSTITUTE