

In The Name of Allah



Digital Media Laboratory
Sharif University of Technology

Statistical Pattern Recognition

Active Learning

Hamid R. Rabiee
Mohsen Fadaee

Spring 2012

<http://ce.sharif.edu/courses/90-91/2/ce725-1/>

Agenda

- ✧ **What is Active Learning?**
- ✧ **Active Learning vs. Passive Learning**
- ✧ **Active Learning Scenarios**
- ✧ **Query Strategy Frameworks**
 - ✧ **How to sample queries?**
- ✧ **Practical Considerations**



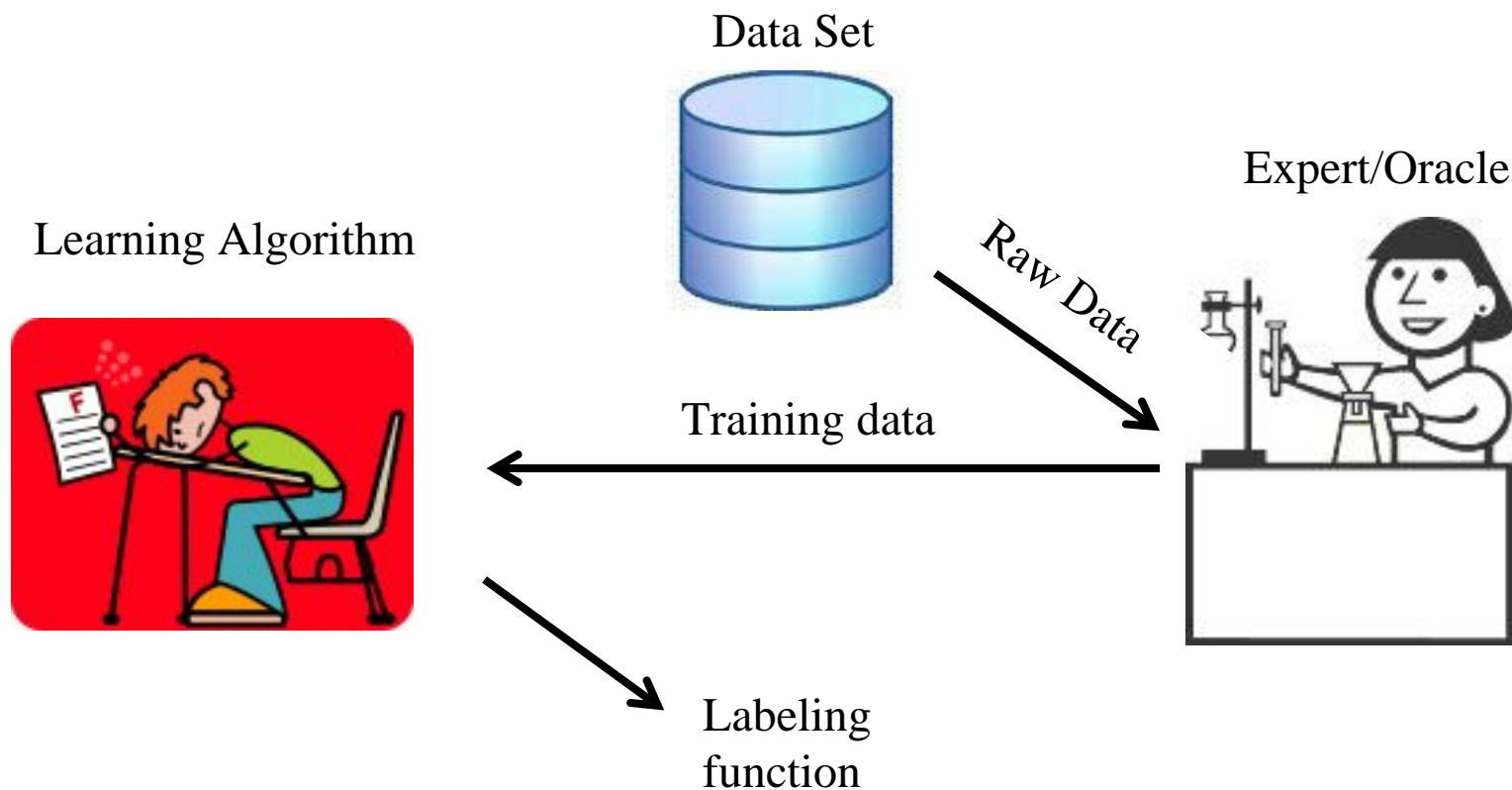
What is a Active Learning?

- ✧ **A learner, in supervised or semi-supervised mode, uses a set of labeled data as training set.**
- ✧ **There are two approaches to access labeled data.**
 - ✧ **Passive Learning: Supervisor gives labeled data to learner.**
 - ✧ **Active Learning: Learner selects data for labeling.**



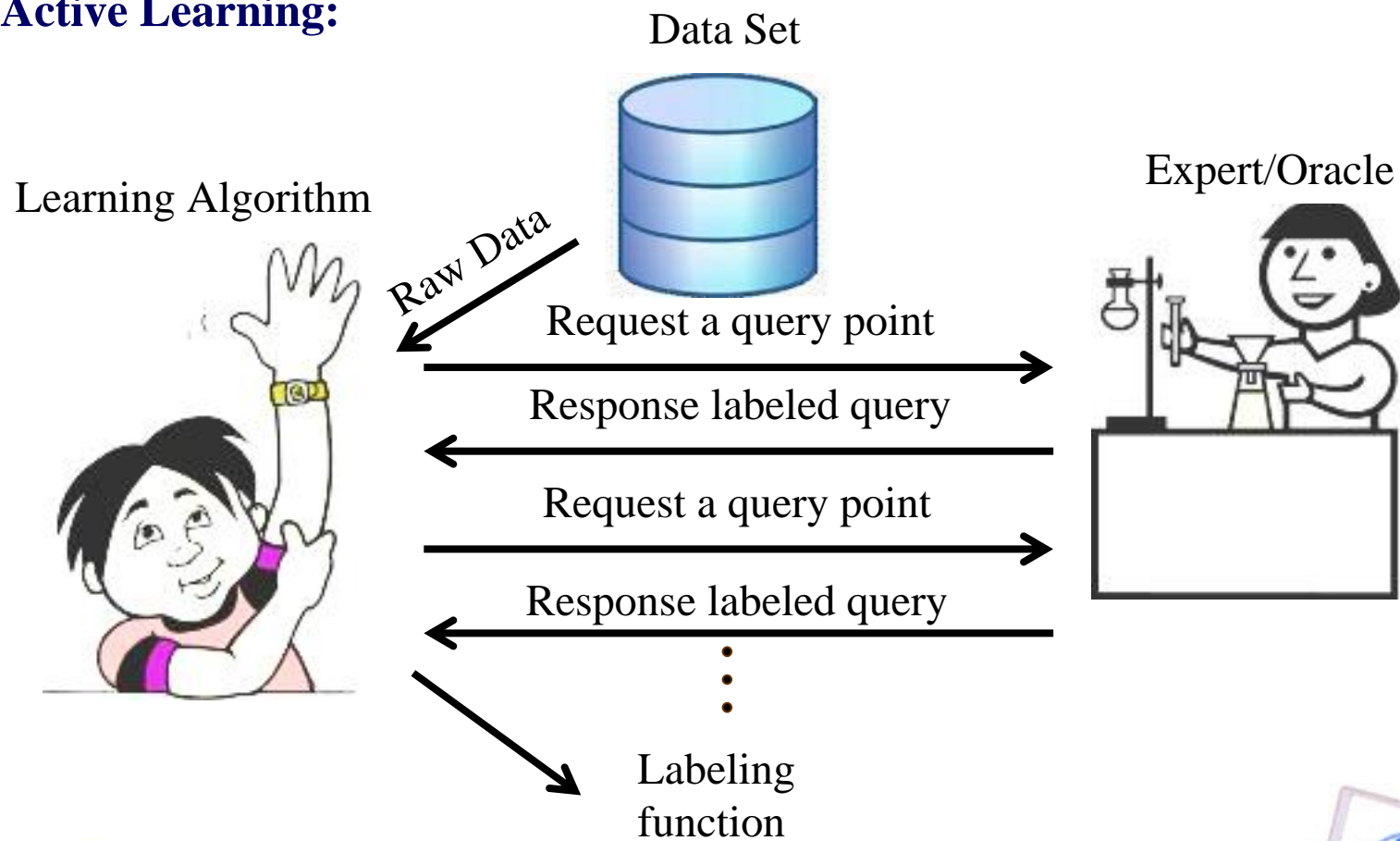
Active Learning vs. Passive Learning

✧ Passive Learning:



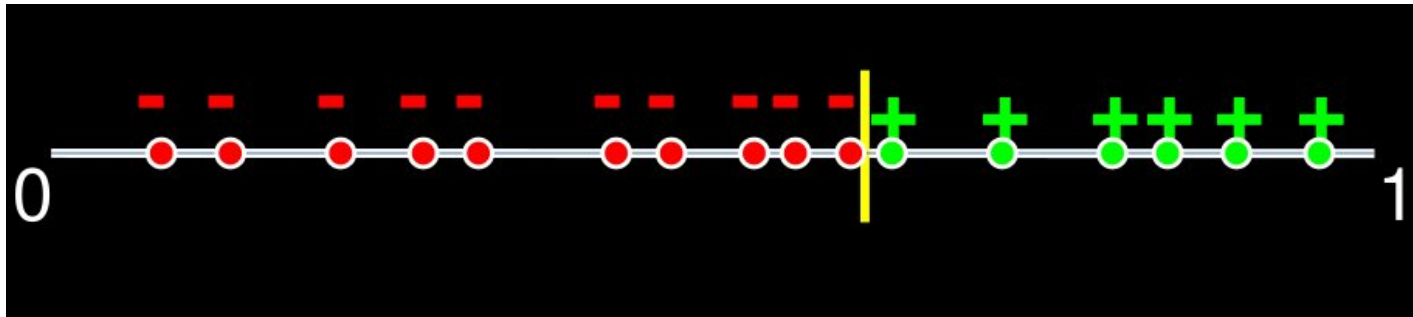
Active Learning vs. Passive Learning

❖ Active Learning:



A Simple Example

- ✧ How many random labeled points are sufficient to find the threshold, where ϵ is the maximum error rate?



- ✧ **Passive:** It's enough to draw $O\left(\frac{1}{\epsilon}\right)$ random labeled instances.
- ✧ **Active:** By using a simple binary search through these unlabeled instances, a classifier with error less than ϵ can be achieved with $O\left(\log\frac{1}{\epsilon}\right)$



Scenarios

✧ Membership Query Synthesis

- ✧ The learner generates queries de novo, rather than those sampled from some underlying natural distribution.
- ✧ Not suitable for a human oracle.
- ✧ But suitable for a "robot scientist" which can execute a series of autonomous experiments.

✧ Stream-Based Selective Sampling

- ✧ Each unlabeled instance is typically drawn one at a time from the data source. The learner must decide whether to query or discard it.

✧ Pool-Based Sampling ← We will focus more on this scenario.

- ✧ Large collections of unlabeled data can be gathered at once.
- ✧ Suitable for many real-world learning problems!



How to sample queries?

✧ Uncertainty Sampling Criteria

✧ Least confidence

$$x^* = \arg \min_x P(\hat{y} | x)$$

✧ Margin sampling

$$x^* = \arg \min_x P(\hat{y}_1 | x) - P(\hat{y}_2 | x)$$

✧ \hat{y}_1 and \hat{y}_2 are the first and second most probable class labels.

✧ Entropy

$$x^* = \arg \min_x \left(- \sum_i P(y_i | x) \log P(y_i | x) \right)$$

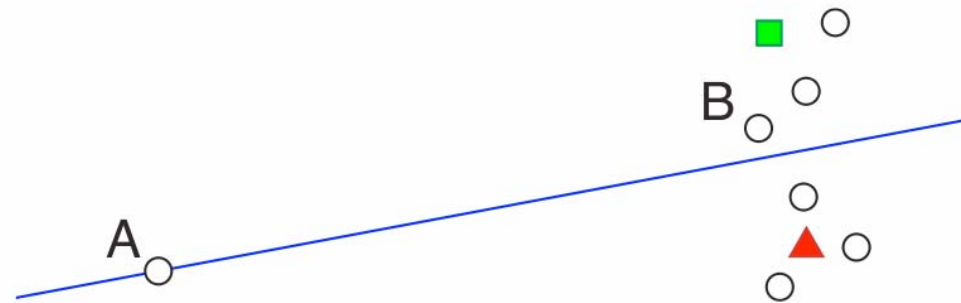
✧ y_i ranges over all possible labelings.



How to sample queries?

✧ Density-Weighted Methods

- ✧ In some cases uncertainty is not sufficient.
 - ✧ The instance should be representative too.
- ✧ Point A or B? On which we are more uncertain and which one is more representative?



$$x^* = \arg \max_x \text{Uncertainty}(x) \left(\frac{1}{W} \sum_{i=1}^W \text{Similarity}(x, x^{(i)}) \right)^\beta$$



How to sample queries?

✧ Query By Committee (QBC)

- ✧ Maintaining a committee of learners, all trained on the current labeled set
- ✧ The goal is finding the best learner for the given dataset.
- ✧ The fundamental premise behind it is minimizing the version space.
 - ✧ A version space is the subset of all learners that are consistent with the training set.
- ✧ QBC needs:
 1. A committee of models that represent different regions of the version space.
 2. Some measure of disagreement among committee members
- ✧ QBC selects the sample with maximum disagreement among the committee.



How to sample queries?

✧ **Expected Model**

- ✧ **Select an instance which knowing its label can make the greatest change to the current model.**
- ✧ **Example:**
 - ✧ **Selecting an instance that leads to maximum Kullback–Leibler distance between old and updated models.**



Practical Considerations

✧ **Batch-Mode Active Learning**

✧ **It is required when:**

- ✧ Serial mode is time consuming
- ✧ Distributed, and parallel labeling environment is available
- ✧ **Q-best is not efficient. Since it fails to consider the overlap in information content among the "best" instances.**

✧ **Noisy Oracle**

- ✧ **If labels come from an empirical experiment.**
- ✧ **Even if come from human experts, may not always be reliable.**
- ✧ **Recently, "crowdsourcing", attempt to "average out" some of this noise by cheaply obtaining labels from multiple non-experts:**
 - ✧ Amazon's Mechanical Turk
 - ✧ Online annotation games



Practical Considerations

✧ Variable Labeling Costs

- ✧ Reducing the number of labeled instances does not necessarily guarantee a reduction in overall labeling cost.
- ✧ There is labeling cost, and mis-classification cost (time, material).
- ✧ Ignoring cost may perform no better than random selection.
- ✧ Different instances: different costs.
- ✧ Cost may vary based on the annotator person.



References

- ✧ **Settles, B., *Active learning literature survey*, Computer Sciences Technical Report 1648, University of Wisconsin-Madison, 2009.**
- ✧ **Hanneke, S., *Advanced Statistical Methods I: Active Learning*, Course 36-781, Carnegie Mellon University, Fall, 2010.**



Any Question

End of Lecture

Thank you!

Spring 2012

<http://ce.sharif.edu/courses/90-91/2/ce725-1/>

