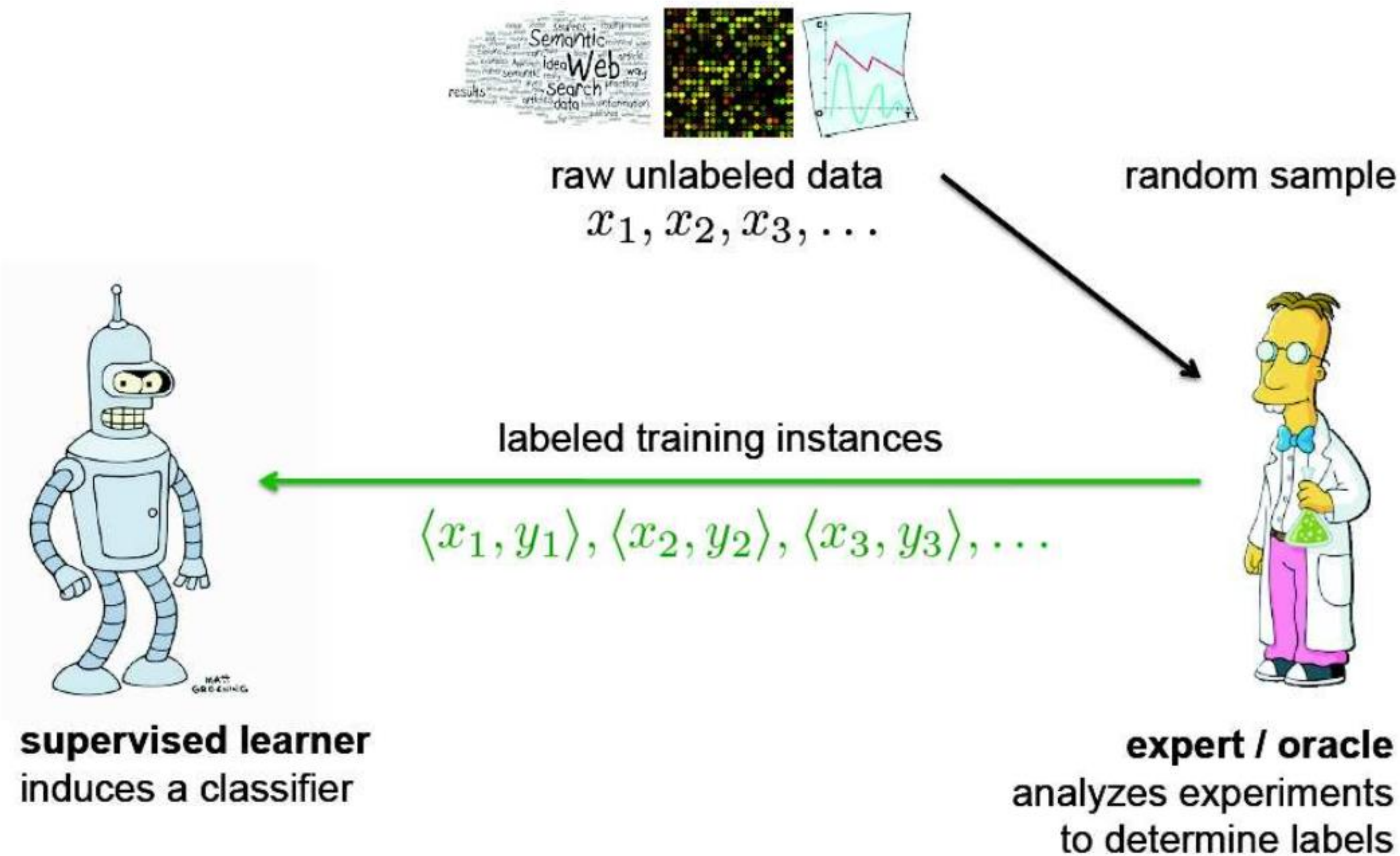
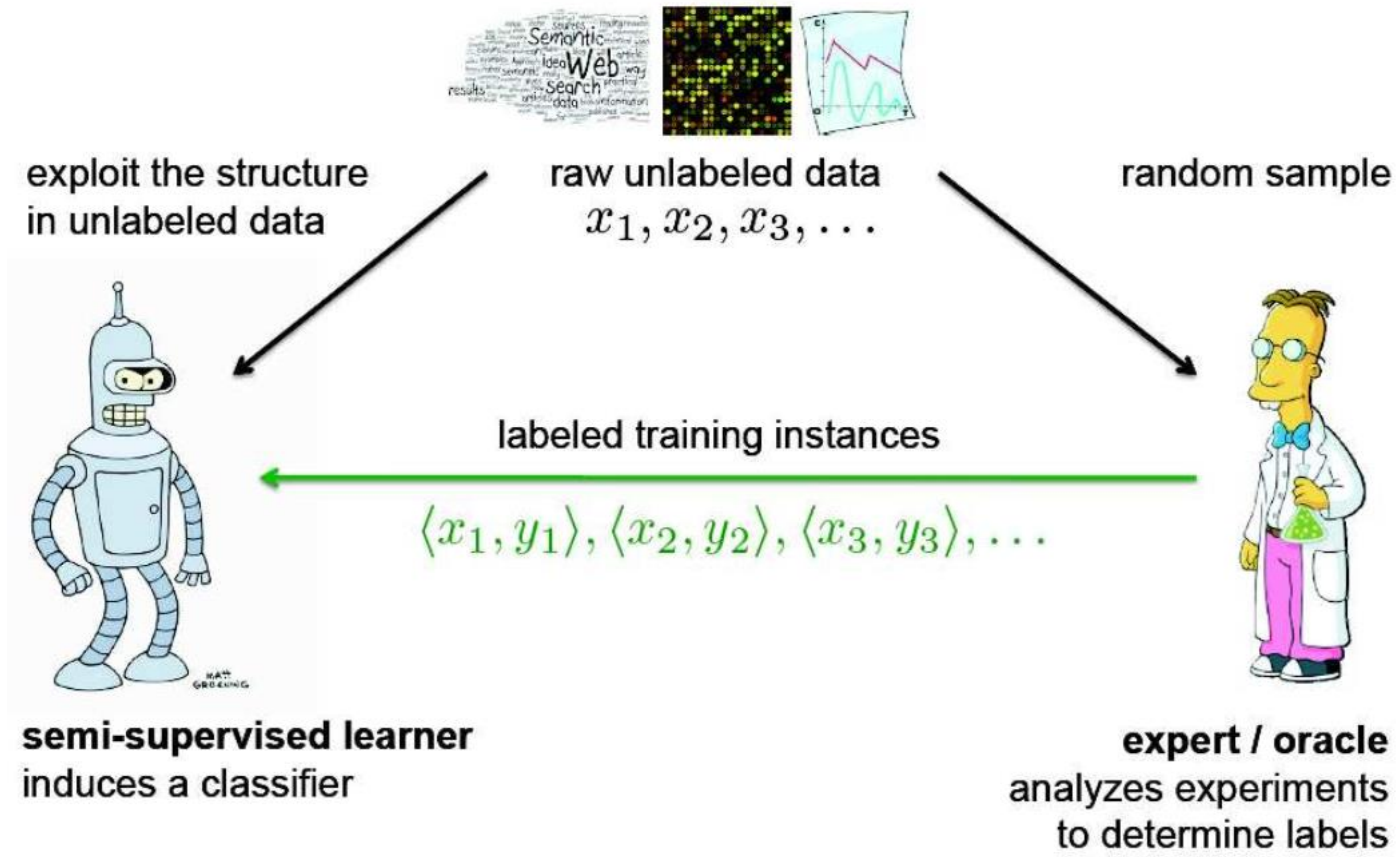
Active Learning

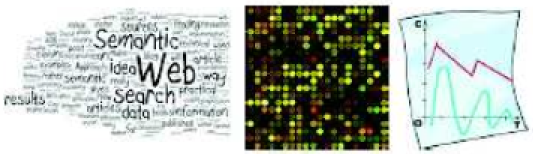
(Passive) Supervised Learning



Semi-Supervised Learning

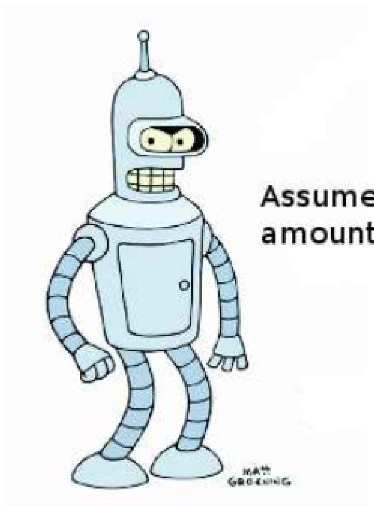


Active Learning



raw unlabeled data

$$x_1, x_2, x_3, \ldots$$

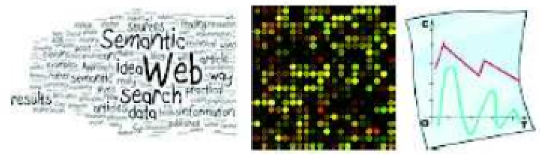


Assumes some small
amount of initial labeled training data

active learner
induces a classifier

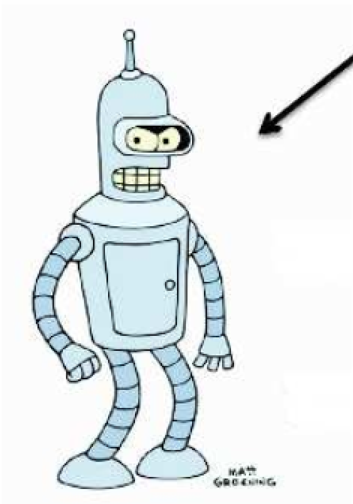
expert / oracle
analyzes experiments
to determine labels

Active Learning



raw unlabeled data
$x_1, x_2, x_3, \ldots$

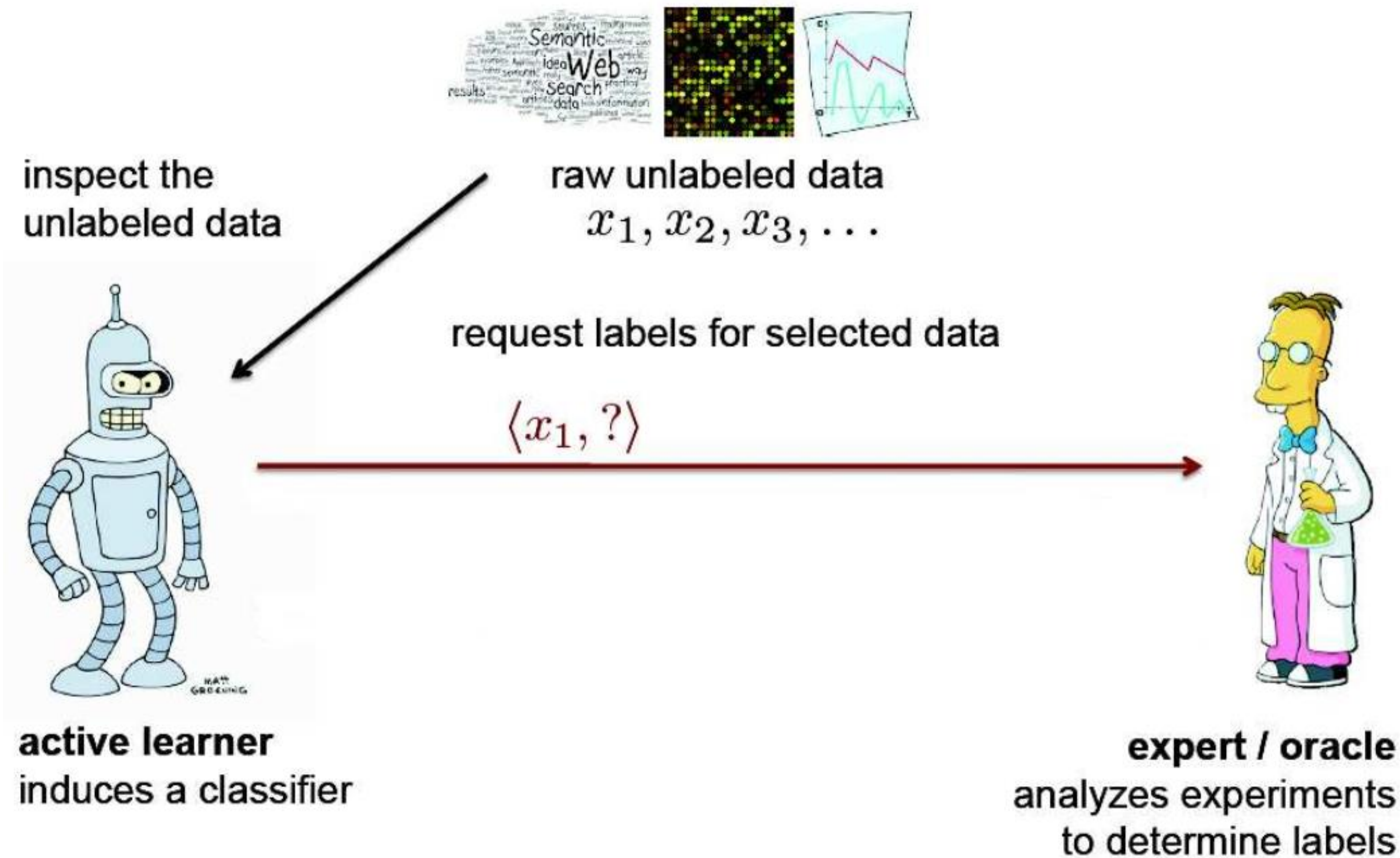
inspect the
unlabeled data



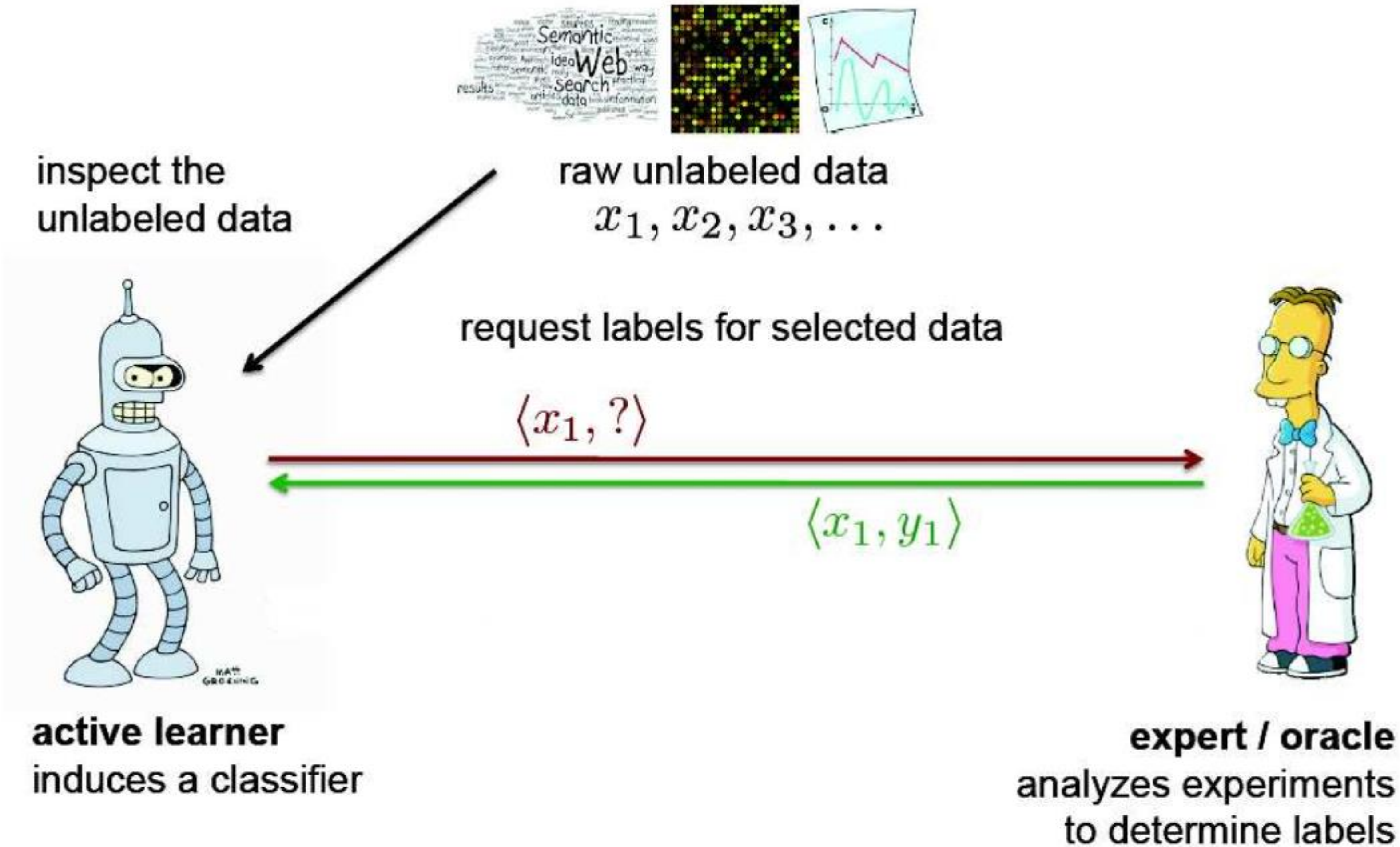
active learner
induces a classifier

expert / oracle
analyzes experiments
to determine labels

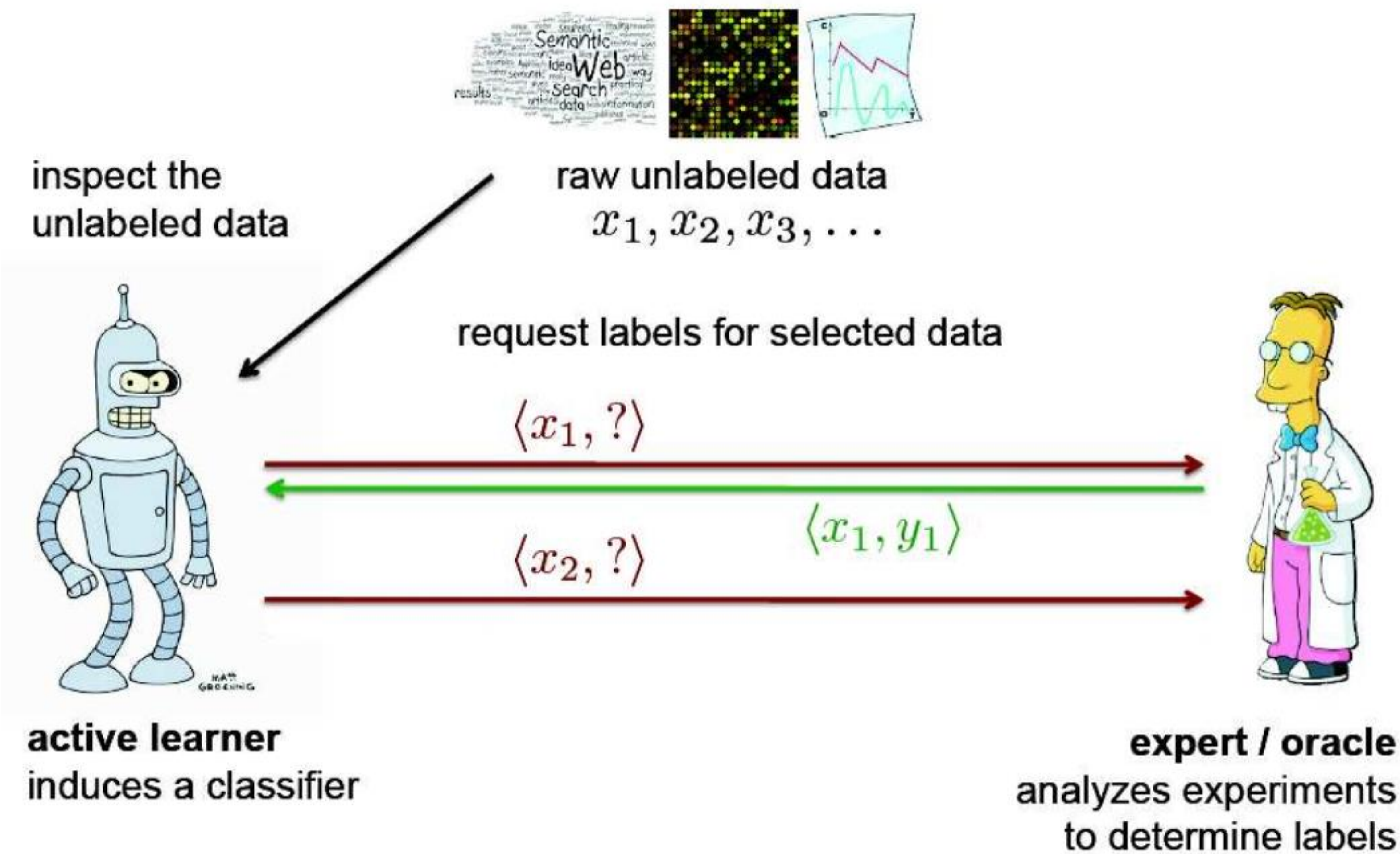
Active Learning



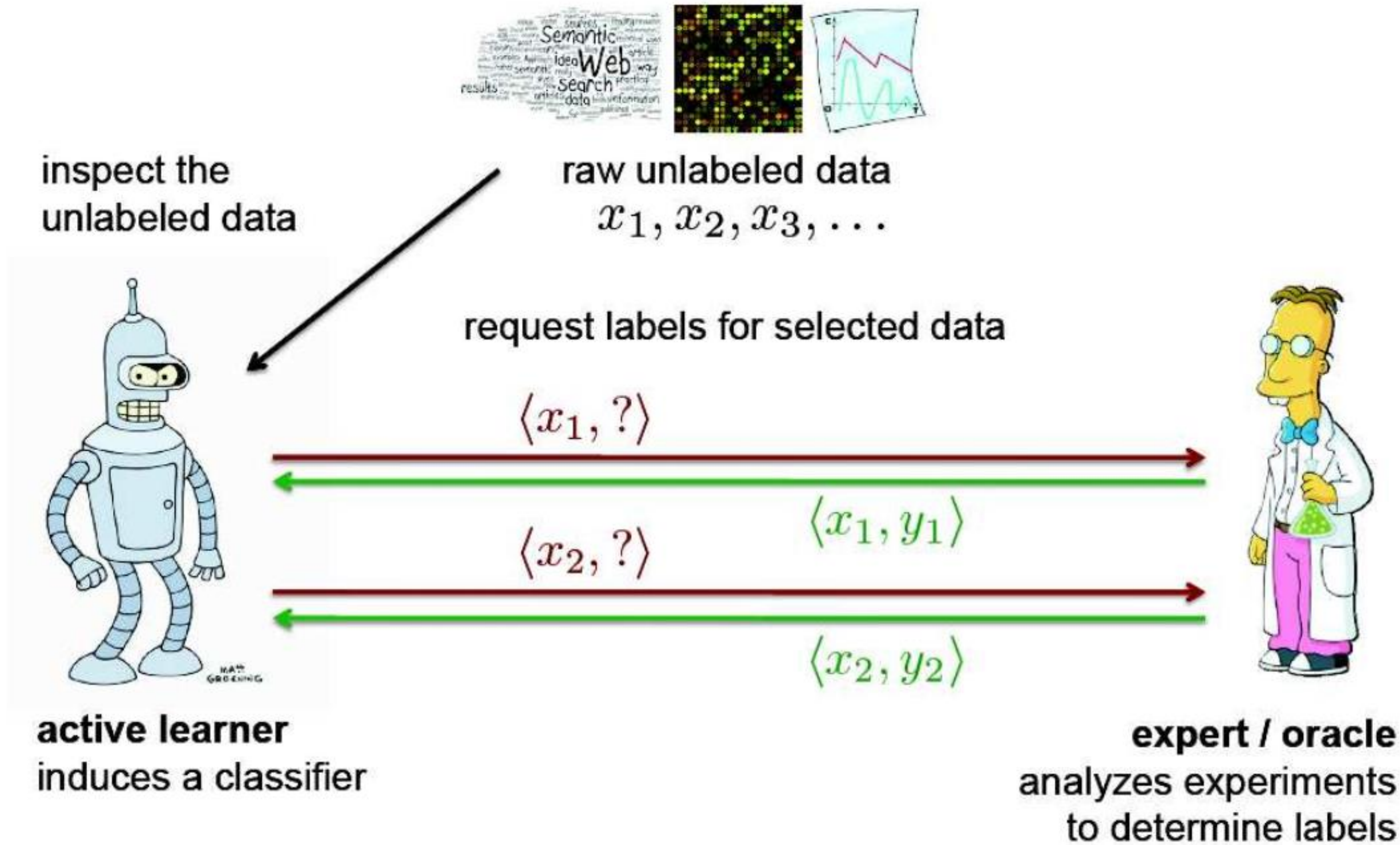
Active Learning



Active Learning

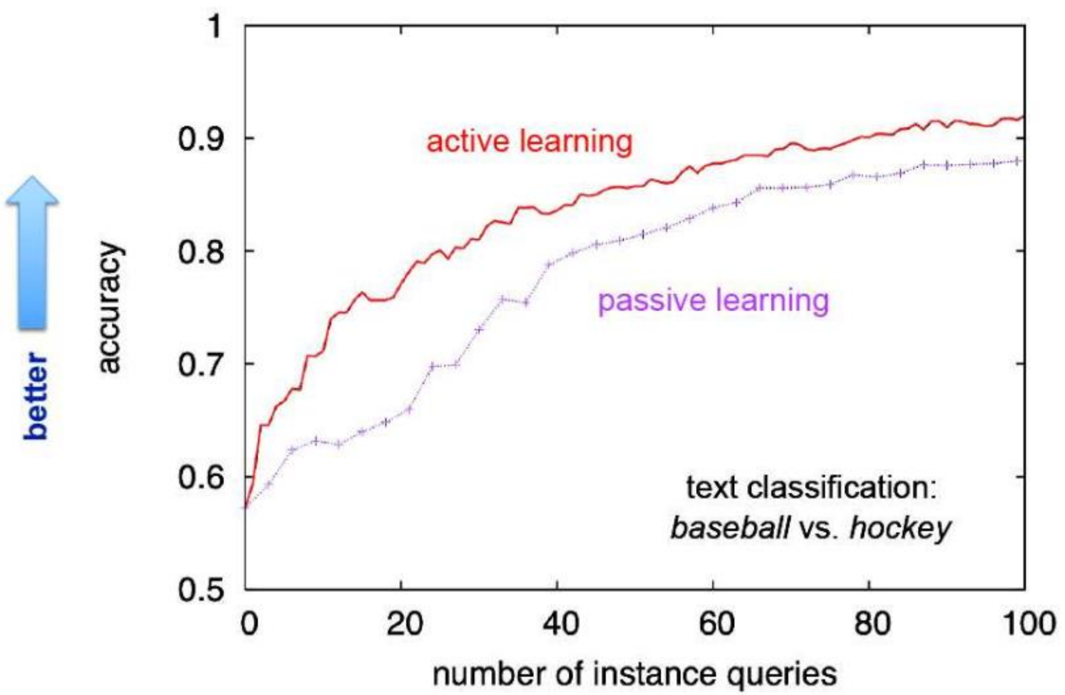


Active Learning



Active Learning vs Random Sampling

- **Passive learning curve**
  - Randomly selects examples to get labels for
- **Active learning curve**
  - Active learning selects examples to get labels for

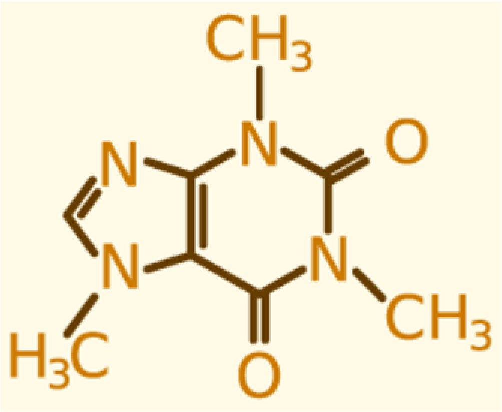


Motivation

- **Why do we need active learning?**
  - Supervised learning can solve all our problems, right?
  - Yes, if we have enough labeled data (input-output pairs)
  - But Labeling is expensive
  - We want to learn a highly-accurate function with few labeled examples
  - Intelligently select the examples for which we want to get labels for (unlabeled data is plentiful and cheap)

Active Learning Example: Drug Design

Goal: find compounds which bind to a particular target



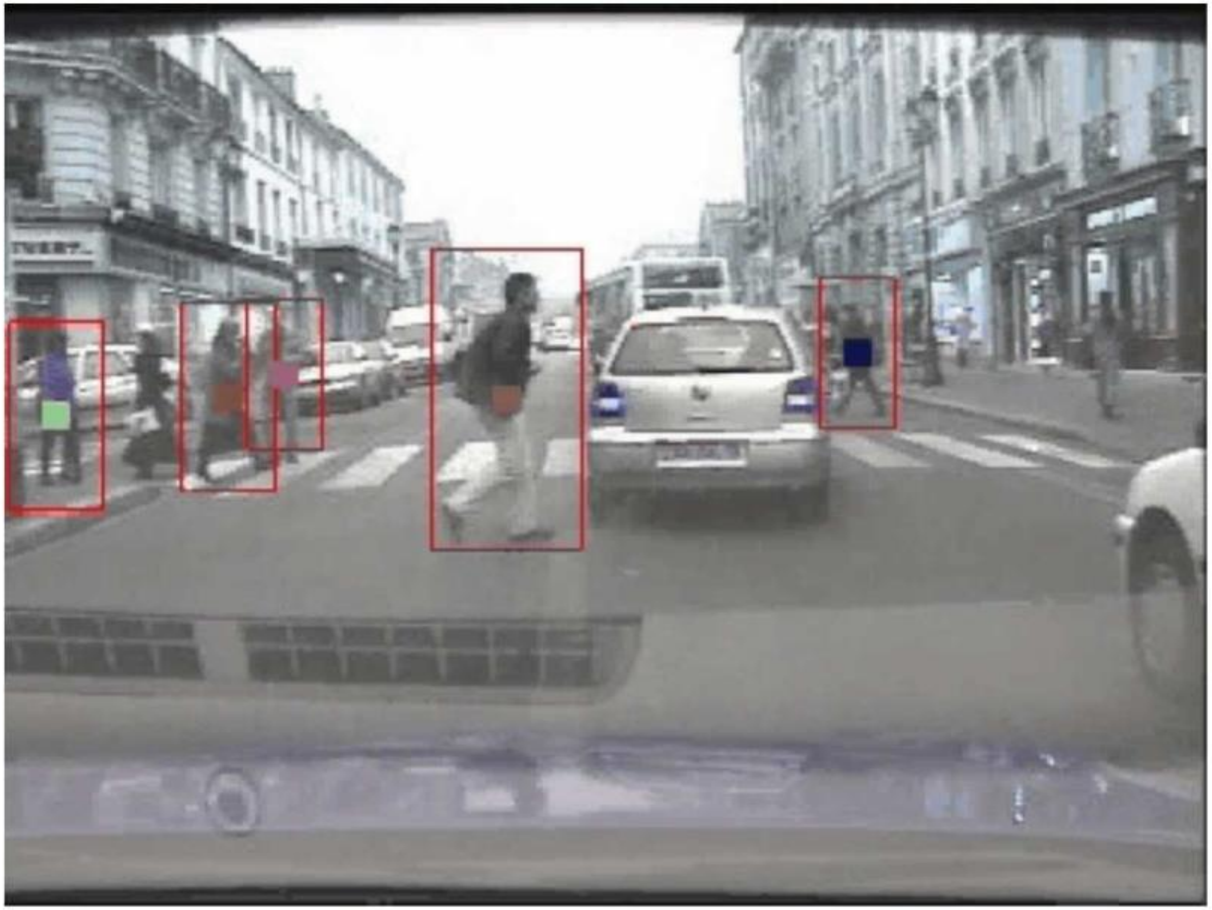
Large collection of compounds, from:

- vendor catalogs
- corporate collections
- combinatorial chemistry

unlabeled point	$\equiv$	description of chemical compound
label	$\equiv$	*active* (binds to target) vs. *inactive*
getting a label	$\equiv$	chemistry experiment

Credit: Sanjoy Dasgupta and John Langford

Active Learning Example: Pedestrian detection



Who uses Active Learning?

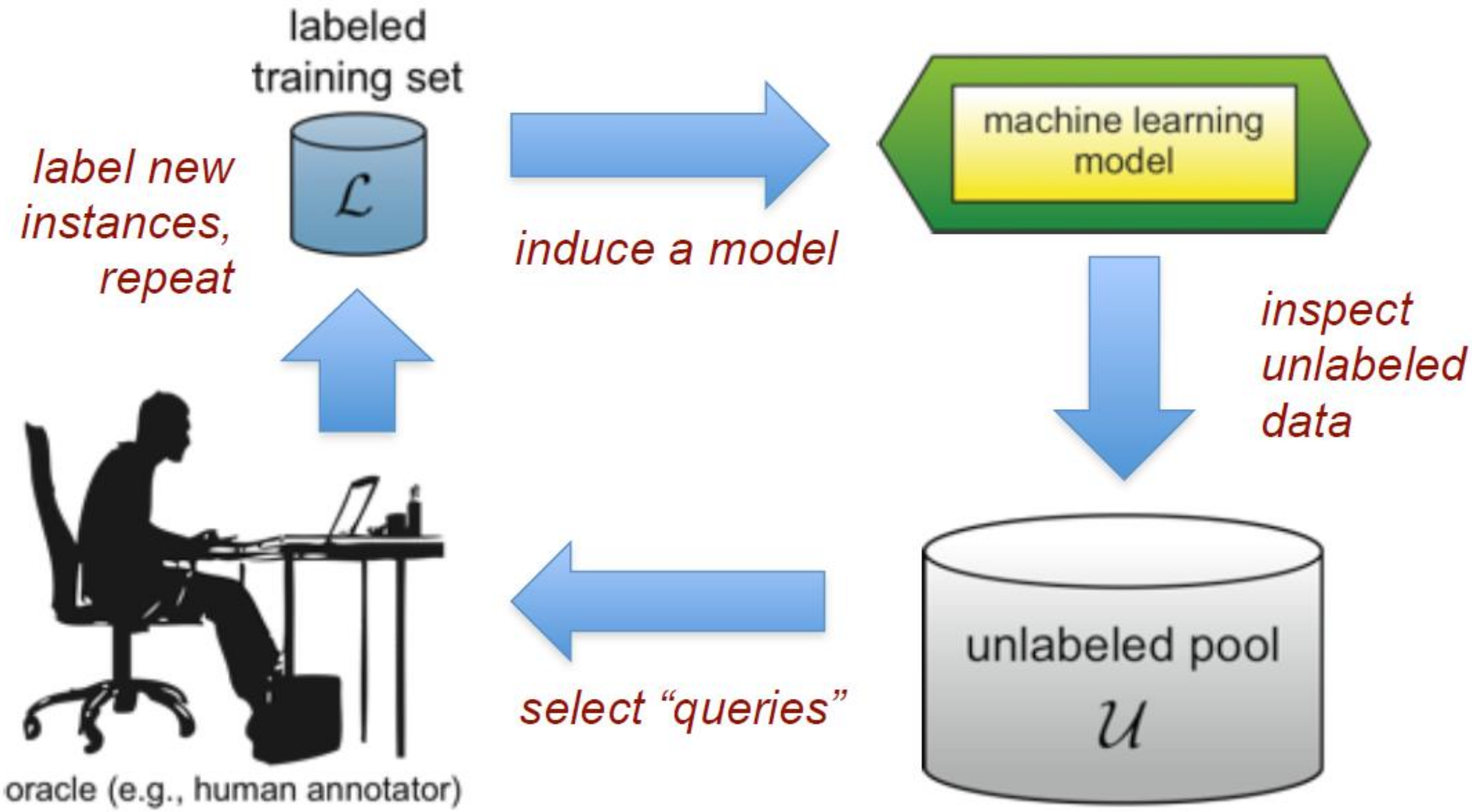
Sentiment analysis for blogs; Noisy relabeling
– *Prem Melville*

Biomedical NLP & IR; Computer-aided diagnosis
– *Balaji Krishnapuram*

MS Outlook voicemail plug-in [Kapoor et al., IJCAI'07]; "A variety of prototypes that are in use throughout the company." – *Eric Horvitz*

"While I can confirm that we're using active learning in earnest on many problem areas… I really can't provide any more details than that. Sorry to be so opaque!"
– *David Cohn*

Pool based Active Learning



Credit: Burr Settles

Query Selection Strategies

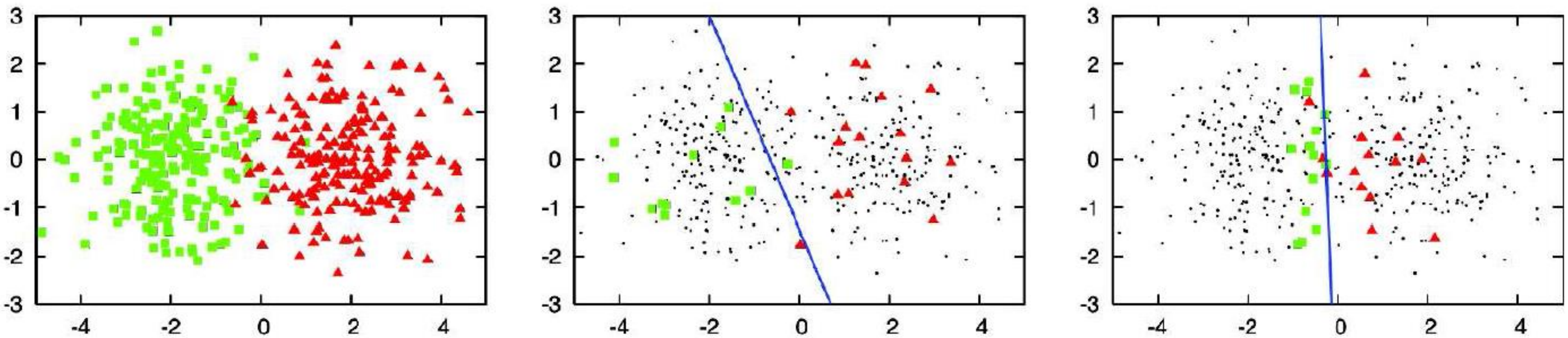
- Any active learning algorithm requires a query selection strategy.
- Some examples
  - Uncertainty Sampling
  - Query By Committee (QBC)
  - Expected Model Change
  - Expected Error Reduction
  - Variance Reduction
  - Density Weighted Methods

Uncertainty Sampling

- Select examples which the current model is most uncertain about
- Many ways to measure uncertainty
  - Based on distance from the hyperplane
  - Using the probability distribution over labels P(y|x) for probabilistic models
- Some examples based on label probabilities
  - **Least Confident**, where confidence is defined as 1 – probability of highest scoring label
  - **Smallest Margin**, where margin is defined as the difference between the probabilities of the first and second best labels

Uncertainty Sampling: Illustration

- Example: uncertainty sampling based on the distance from hyperplane (i.e., margin based)



400 instances sampled from 2 class Gaussians

random sampling
30 labeled instances (accuracy=0.7)

uncertainty sampling
30 labeled instances (accuracy=0.9)

Query By Committee (QBC)

- QBC uses a committee of models (say $h_1, h_2, \ldots, h_k$)
- All models are trained using the currently available labeled data L
- How is the committee constructed?
  - Ensemble methods (e.g., bagging, boosting)
- All models vote their predictions on the unlabeled pool
- The example(s) with maximum disagreement is(are) chosen f labeling
  - simple disagreement rate
  - Entropy over the vote distribution over all labels
- Each model in the committee is re-trained with new example

Summary and Outlook

- Active learning is a label-efficient learning strategy
- Intelligently selects the examples based on their informativeness
- Other variants
  - Different examples having different labeling costs
  - Access to multiple labeling oracles varying in labeling cost and accuracy
  - Active learning on features instead of labels
- Further Reading
  - Active learning survey from Burr Settles
  - http://burrsettles.com/pub/settles.activelearning.pdf