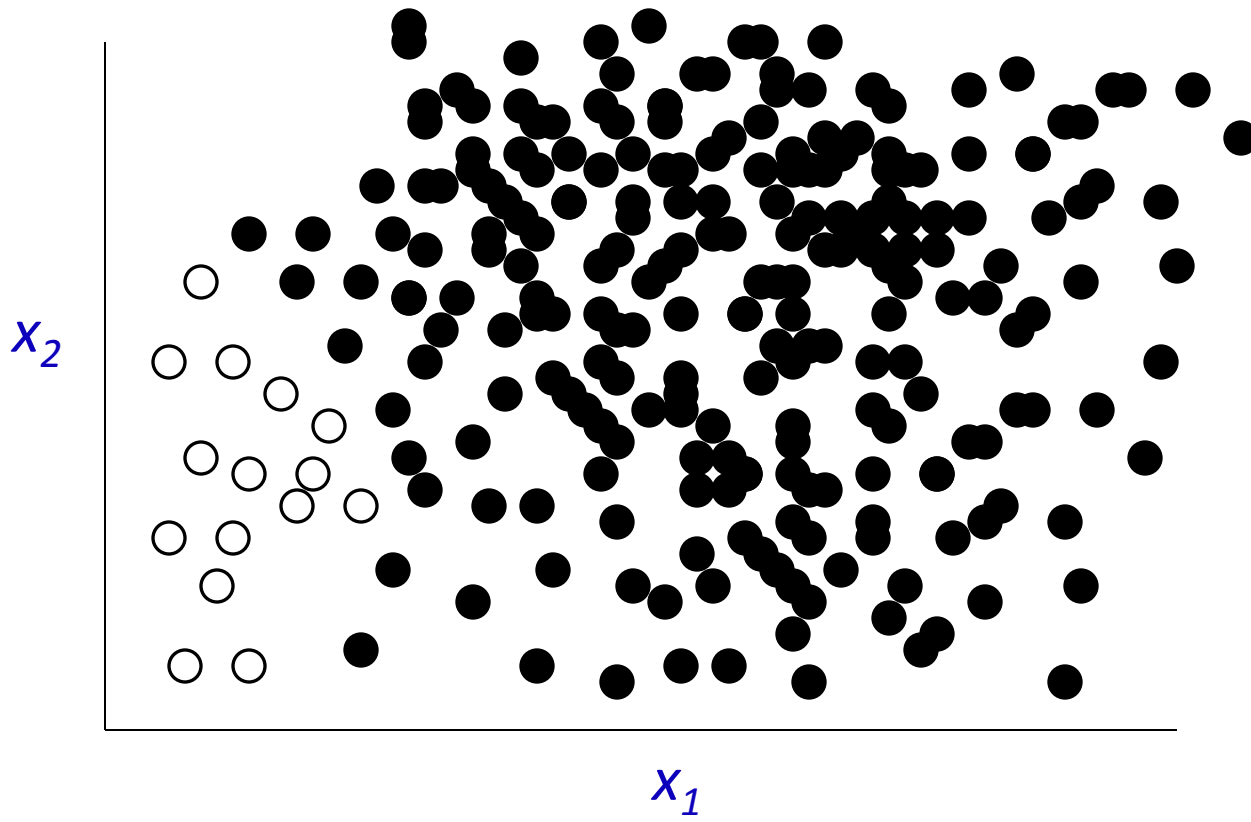
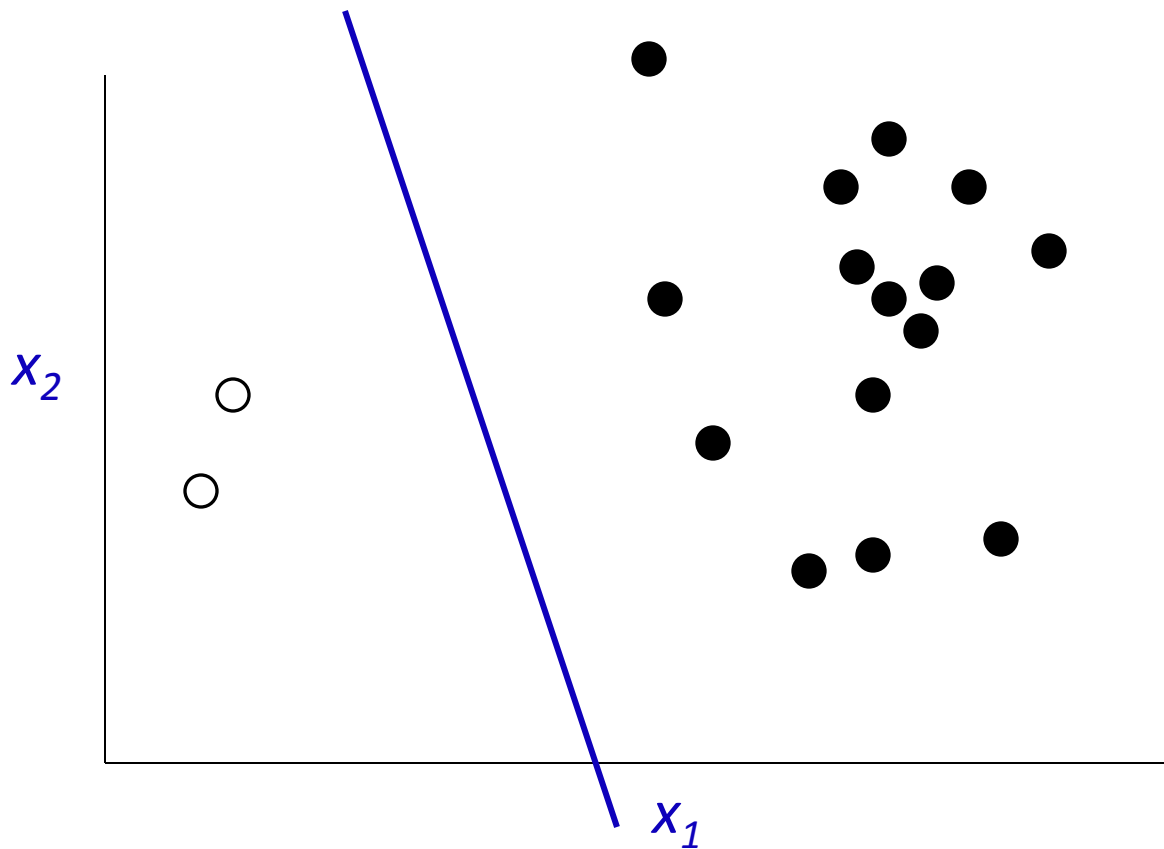


Active Learning

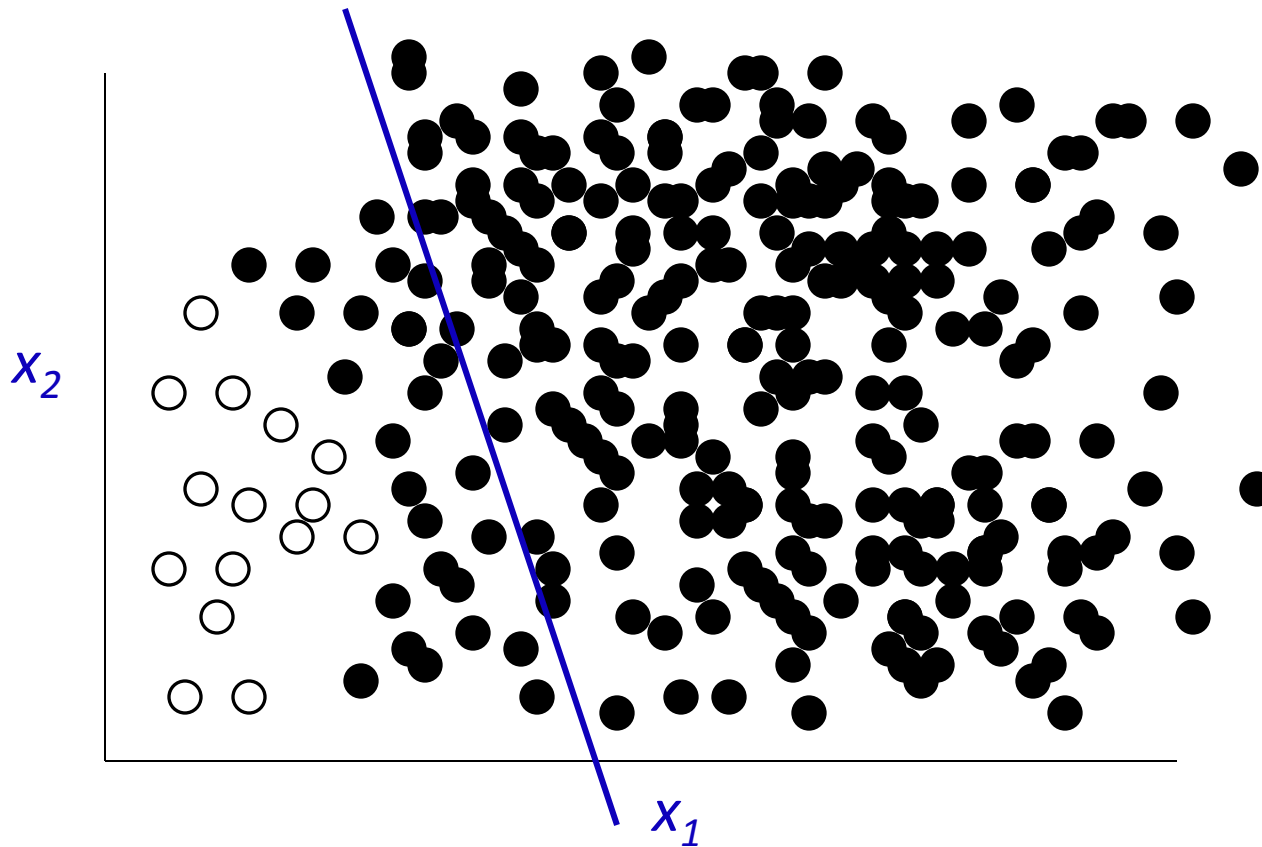
Active Learning: Motivation



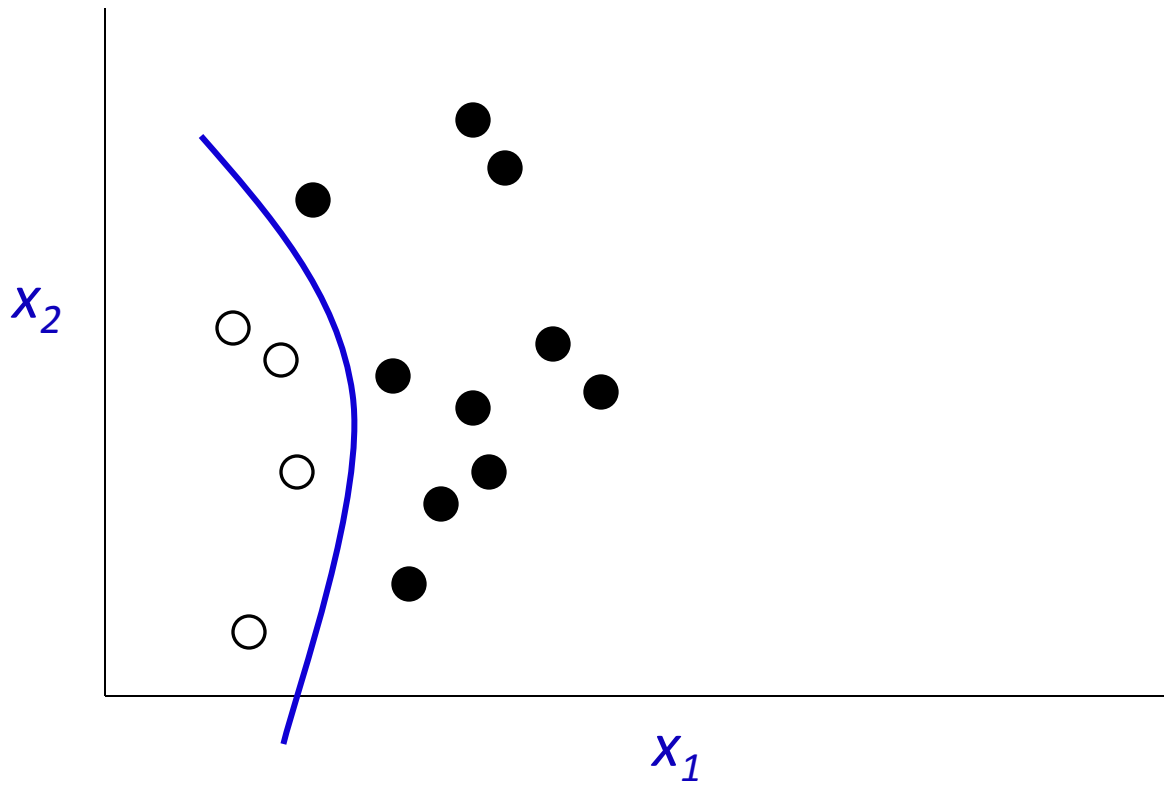
Active Learning: Motivation



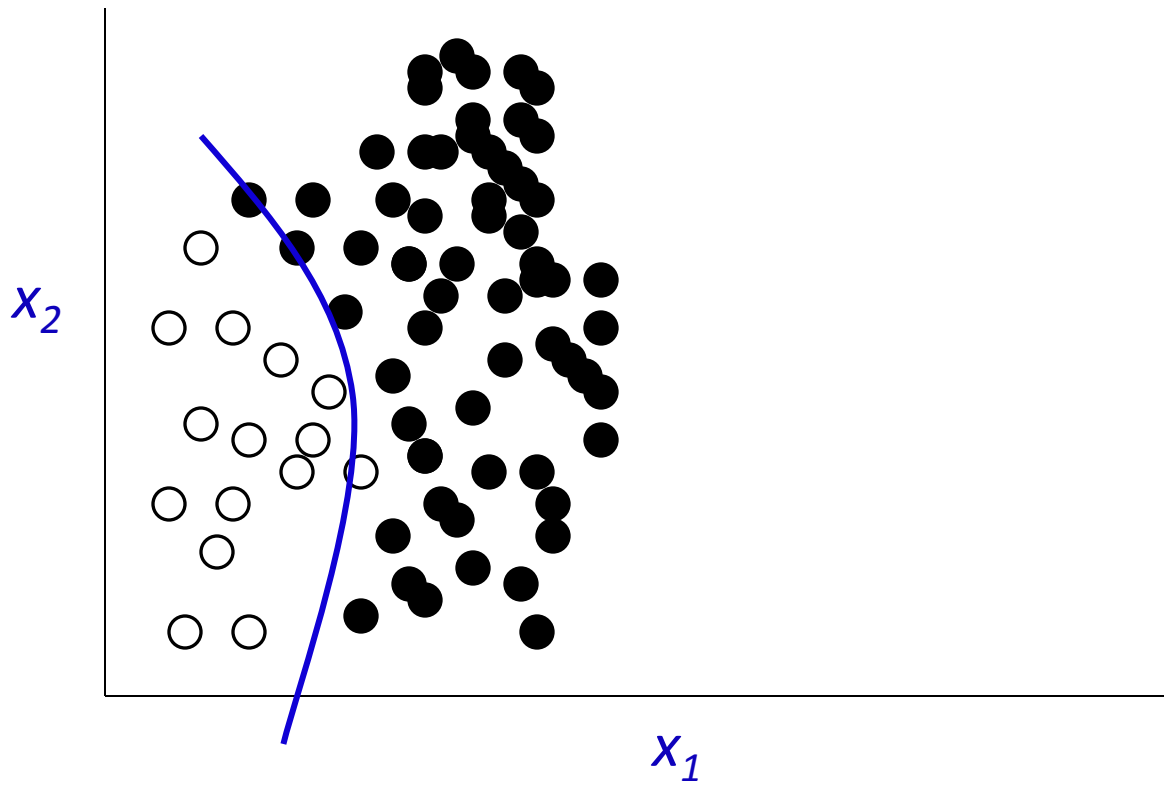
Active Learning: Motivation



Better training examples



Better performance

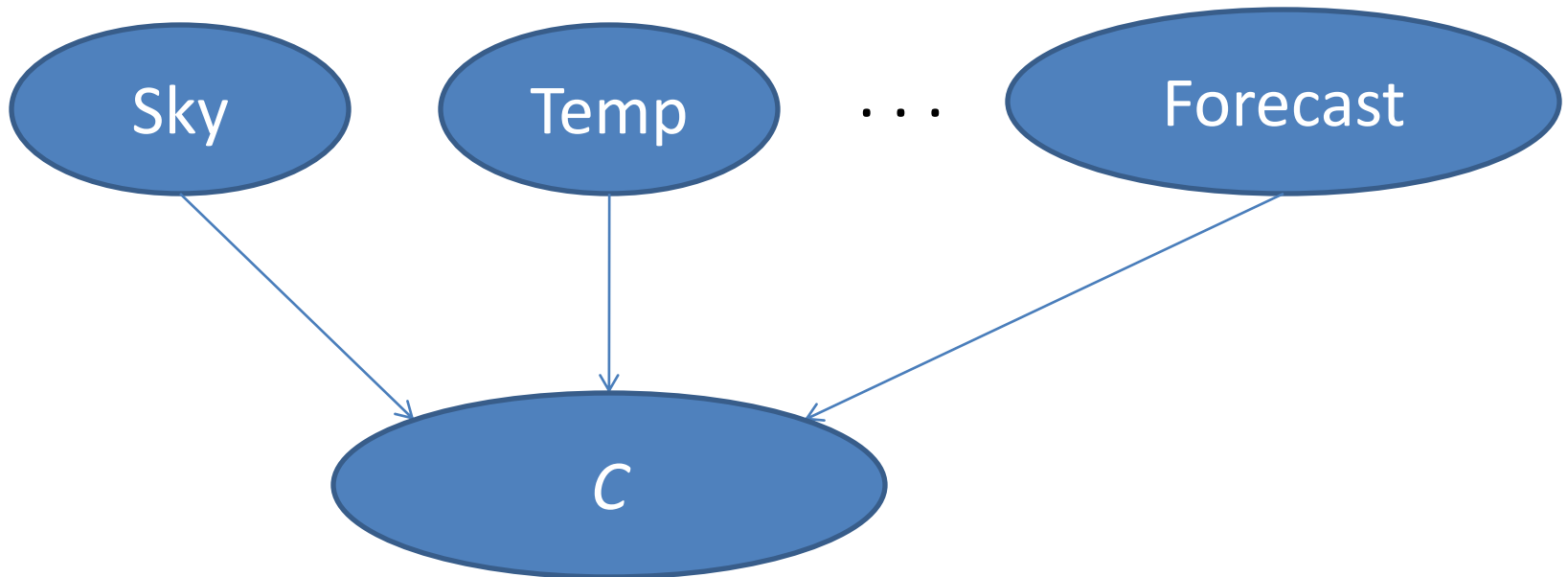


- “Days on which my friend Aldo enjoys his favorite water sport”

INPUT OUTPUT

Sky	Temp	Humid	Wind	Water	Forecast	C(x)
sunny	warm	normal	strong	warm	same	1
sunny	warm	high	strong	warm	same	1
rainy	cold	high	strong	warm	change	0
sunny	warm	high	strong	cool	change	1

INPUT						OUTPUT
Sky	Temp	Humid	Wind	Water	Forecast	$C(x)$
sunny	warm	normal	strong	warm	same	1
sunny	warm	high	strong	warm	same	1
rainy	cold	high	strong	warm	change	0
sunny	warm	high	strong	cool	change	1



Learning a function

- Consider a hypothesis space H of conjunctions of constraints on attributes
- Each constraint can be:
 - A specific value : e.g. *Water=Warm*
 - A don't-care value : e.g. *Water=?*
- Example hypotheses:

Sky	Temp	Humid	Wind	Water	Forecast
sunny	?	?	?	?	?
?	warm	?	?	?	same

Two options

- Passive learning:
 - Watch Aldo and record what he does
- Problem: it takes a long time for every combination to come about => inaccuracy
- Alternative: *active* learning
 - Ask Aldo what he thinks about particular weather configurations

INPUT OUTPUT

Sky	Temp	Humid	Wind	Water	Forecast	C(x)
sunny	warm	normal	strong	warm	same	1
sunny	warm	high	strong	warm	same	1
rainy	cold	high	strong	warm	change	0
sunny	warm	high	strong	cool	change	1

New training examples:

rainy	cold	normal	strong	warm	change	0
-------	------	---------------	--------	------	--------	---

Useless! Already knew Humid=?

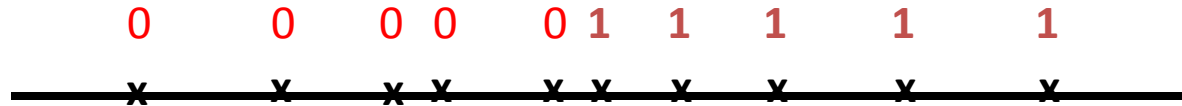
rainy	warm	normal	strong	warm	same	0 or 1
--------------	------	--------	--------	------	------	--------

Learn Sky=sunny

Learn Sky=?



Simpler Example: threshold function



Unlabeled data: labels are all 0 then all 1 (left to right)

Classifier (threshold function): $h_w(x) = 1$ if $x > w$ (0 otherwise)

Goal: find transition between 0 and 1 labels in minimum steps

Random choice

- Requires $O(n)$ training data to find underlying classifier

Binary search for transition between 0 and 1

- Requires $O(\log n)$ training data to find underlying classifier

Questions!

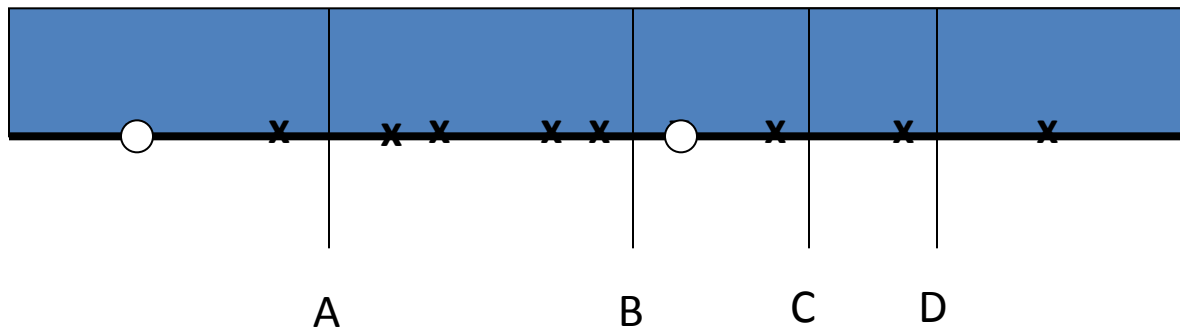
- How do we choose which questions to ask?
- Lots of possibilities
- Today:
 - Entropy
 - Entropy plus "representativeness"
 - Query by committee

Entropy plus representativeness

- Entropy Approach: query least certain examples
 - Don't waste time querying things we already know
 - But what's a problem with this?
- Alternative: *sample* instances weighted by entropy

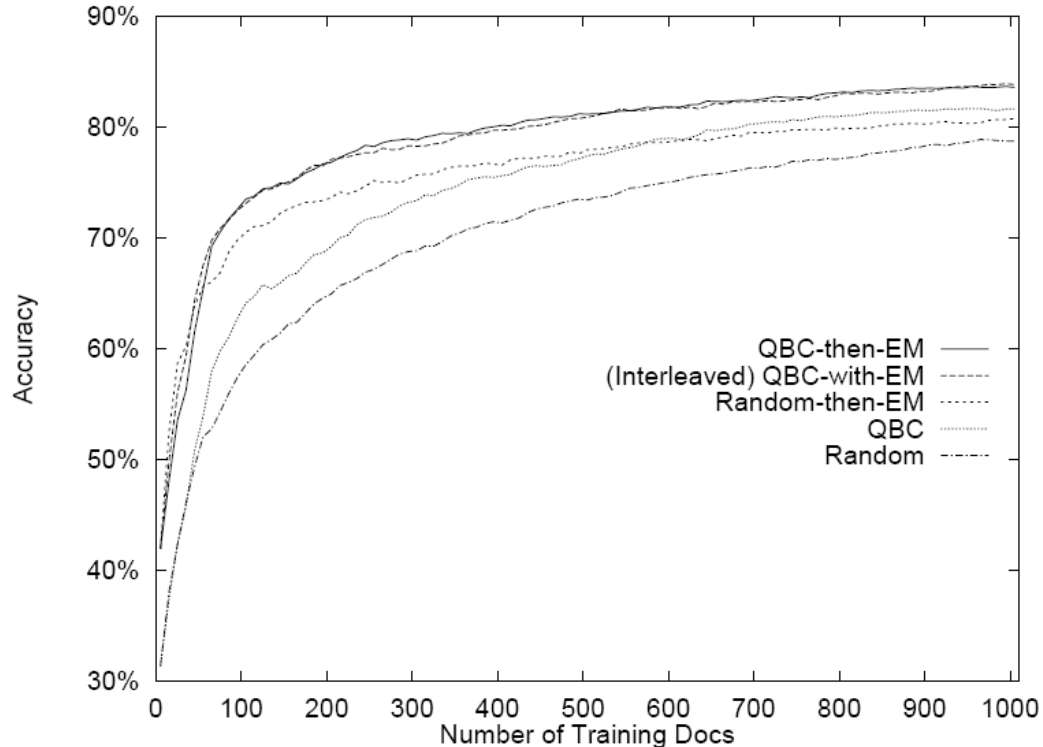
Query by Committee

- Queries an example based on the degree of disagreement between committee of classifiers
 - Ensemble of classifiers/hypotheses (e.g. bagging, boosting)



Query by Committee Application

- Used naïve Bayes model for text classification in a Bayesian learning setting (20 Newsgroups dataset)



Another direction: Active Learning with multiple sources

- Mechanical Turk, ESP game
 - Multiple, heterogeneous annotators
- Input can be adversarial
 - Novel operation: query annotators for most *certain* examples
 - Why? Gauge annotator quality
- Exploration vs. exploitation trade-off