Supervised Learning

- All algorithms so far (kNN, DT, Nbayes, Perceptron, ..., SVM) are passive
- Input: is set of examples and labels
- Output: is classifier.

- What if the learner can be active?
  - What does that mean?

Focusing on Labels

- Often labels are expensive because they require work/time of a domain expert
- or scarce/unavailable because we do not have access to "ground truth"
- Recent methods try to use crowdsourcing to obtain labels - this is related to our topic but out of scope for the course

Active Learning Variants

- Where/how can the learner ask for labels?
  - We will discuss 3 settings.
  - Query Learning
  - Stream based action learning
  - Pool based active learning
  - Various types of questions possible:
    - In Classification
      - Equivalence queries
      - Subset queries
      - Membership queries
    - In preference elicitation:
      - Ranking queries
      - ...
Membership query Learning

• Consider learning a discrete threshold function \((x \geq n)\) where \(n\) is an integer in the range \([-1000,1000]\)
• Learning algorithm using membership queries?
  
  • Naive query strategy: ...
  
  • Smart query strategy: ...

• Consider learning a discrete 2D rectangle where \(x, y\) boundaries are integers in the range \([-1000,1000]\)
• Learning algorithm using membership queries?
  
  • Naive query strategy: ...
  
  • Smart query strategy: ...

  This is much harder

Membership query Learning

• General principle? Ask about an example which will shrink the range of options to the largest extent.
• Since the answers are not known in advance: ask about an example which will shrink the range of options for any possible answer.

  • In noise free case: the version space is the set of hypotheses that are consistent with current data
  • Seek to shrink the version space

Notable Query Learning Systems

• Program synthesis/debugging [S82]: developed methods for interactive construction of Prolog programs via equivalence queries (and counter examples) and membership (I/O) queries.
  
  • Robot scientist [K04,K09]: automate process of investigation re which genes (in yeast) encode certain enzymes.

Program debugging

\[
\text{isort}([X|Y],Z) \leftarrow \text{isort}(X,Z), \text{insert}(X,Z,Y).
\text{isort}([X],[]).
\text{insert}(X,[Y],Z) \leftarrow Y>X, \text{insert}(X,Y,Z).
\text{insert}(X,[Y],Z) \leftarrow X\leq Y.
\text{insert}(X,[]).
\]

We first test \text{isort} on \([2,1,3]\),

\[
\text{isort}([2,1,3],X).
\]

\[
X = [2,3,1]
\]
• Automatically design yeast growth experiments
• Implement in hardware
• Observe results, and continue

• 1000 combinations a day
• Goal: identify which genes encode specific enzymes/functions
• Novel discoveries

Stream based and Pool based active learning avoid this difficulty by allowing questions only on examples that already exist.

Query Learning: summary

• Main idea: try to reduce uncertainty regarding possible true concepts
• Impressive results in some cases
• Hard to apply in general
• Mainly because of the need to construct meaningful examples or questions

Pool Based Active Learning

• Assume that a set of unlabeled examples is given
• Learner can ask for labels of examples in this set
• A simple baseline will draw a random subset of examples and ask for their labels.
  • Can we do better?

Pool Based Active Learning

1) Obtain initial classifier
2) While expert is willing to label
   a) Make predictions with current classifier
   b) Identify “useful instance(s)"
   c) Request labels for “useful instances"
   d) Retrain and Goto 2

• How to select instances in step 2?
  - Classifier specific schemes?
  - Classifier independent schemes?

• Uncertainty sampling:
  • Instance selected for which current hypothesis is least confident in its prediction
    - Uses a single hypothesis to determine uncertainty
    - How do we define uncertainty for k-NN, decision trees, N Bayes, SVM?
Pool Based Active Learning

- If we have or can approximate a probabilistic prediction and label probabilities are (from max to min)
  
  \[ p_1, p_2, \cdots, p_k \]

- Instance with smallest \( p_1 \)
- Instance with smallest \( (p_1 - p_2) \)
- Instance with largest entropy \( (p_1, p_2, \cdots, p_k) \)

(same for 2-class but different in multi-class)

Experiments on ad dataset

- Distinguish Advertisement Images from Non-Ads (UCI)
- 3279 instances, 458 ads, \([0,1]\) input, \(|S_l(0)|=25, |S_{select}|=5\)
- Used Averaged Perceptron with Margin

Pathological Example for Uncertainty Sampling

- Uncertainty sampling
  - Easy to implement in many algorithms
  - Successful in many cases
  - Can be trapped into missing important distinctions
  - Decisions based on one/current hypothesis and may not reflect diversity in version space
Pool/Stream based Active Learning

- Ideally take a vote among all the hypotheses in the version space
- Picking a point with high disagreement will reduce the version space regardless of what the true label is!
- and will reduce uncertainty

- But this is almost never feasible
- Alternative: form a “committee”

Query by Committee

- Ideally take a vote among all the hypotheses in the version space
- Instead take a vote among a few selected hypotheses – the committee
- Pick an example that maximizes disagreement

- How to pick committee? And its size?
- Works for Pool based or stream based

Stream Based Active Learning

- Query by Committee

1) Calculate initial hypothesis space $H_1$ and prior $\pi$
2) For $t = 1, \ldots, T$
   a) Retrieve unlabeled instance $x_t$
   b) Draw $h_1, h_2 \sim (\pi, H_t)$
   c) If $h_1(x_t) \neq h_2(x_t)$, request label from expert
   d) Generate $H_{t+1}$ and recalculate $\pi$

- Here committee size is 2
- And we are assuming ability to sample from “posterior” over hypotheses

Query by Committee

150 random samples
150 QBC-variant queries

Size of Query Batch

- Repeat ad data experiment

- Small size works well

- Small size often works well
- But not practical in many cases where we want to send a batch of questions
- How should the batch be chosen?
Active Learning using Pre-clustering

- If data clusters, we only require a few representative instances from each cluster to label data

[Nguyen & Smeulders 2004]

Systematic Reviews [WSBT2011]

- Systematic review: an exhaustive assessment of all the published medical evidence regarding a precise clinical question
- e.g., “Is aspirin better than leeches in inducing more than 50% relief in patients with tension headaches?”
- Must find all relevant studies

Systematic Reviews [WSBT2011]

- The problem

<table>
<thead>
<tr>
<th>19 million</th>
<th>PubMed search</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,000</td>
<td>Potentially eligible screen</td>
</tr>
<tr>
<td>500</td>
<td>Relevant</td>
</tr>
</tbody>
</table>

Doctors read these. They’d rather be doing something else.

Citation Screening [WSBT2011]

- An additional challenge of performing active learning when the class distribution is very skewed (very few positive instance)
- Must get 100% recall
- ...
- Reliably reduce labeling effort by X%
- System deployed and used in practice

Active Learning

- Query learning allows to synthesize query examples
- Active Learning focus on existing examples
- Pool based/stream based selection
- Uncertainty sampling is simple and effective - based on the current hypothesis
- Query by Committee uses disagreement amongst an ensemble of classifiers
- Several refinements and improvements
- Margin based selection with linear separators a popular approach