Introduction to Active Learning

Felipe P A Fraga

Departamento de Informática, PUC-RIO

December 2020

Outline

Presentation built upon paper “The two faces of active learning” by Sanjoy Dasgupta [Das09]

- Motivation
- What is Active Learning
- Main challenge: Sampling Bias
- Exploiting cluster structure in data
- Hierarchical Sampling for Active Learning [DH08]
- Closing Remarks

Motivation for Active Learning

Classification is usually associated with **labeled data**.

However . . .

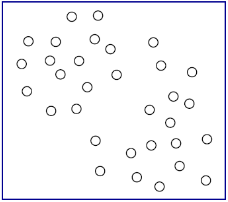
There are many data sources that produce **unlabeled data**:

- Documents on the web
- Speech recognition
- Medical Imaging

Labelling each data point has a **cost** associated with it.

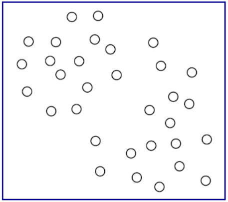
The cost for labelling data can be expensive.

Motivation for Active Learning

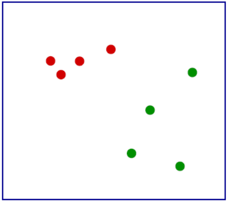


Unlabeled Data

Motivation for Active Learning

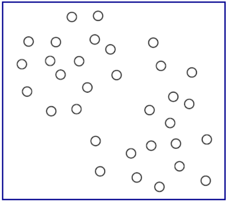


Unlabeled Data

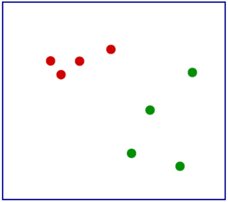


Supervised Learning

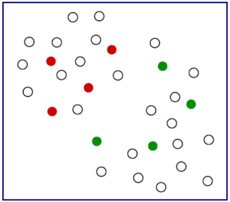
Motivation for Active Learning



Unlabeled Data



Supervised Learning

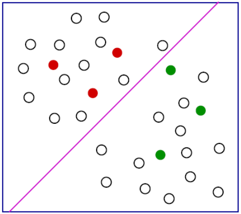


Semisupervised and Active Learning

What is Active Learning

- Start with a pool of unlabeled data
- Pick a few points at random and get their labels
- Repeat:

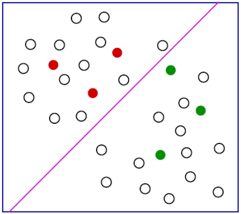
  Fit a classifier to the labels seen so far
  Query the unlabeled point **closest to the boundary**



What is Active Learning

- Start with a pool of unlabeled data
- Pick a few points at random and get their labels
- Repeat:

  Fit a classifier to the labels seen so far
  Query the unlabeled point **closest to the boundary**



Biased sampling: the labeled points are not representative of the underlying distribution

Not Random!

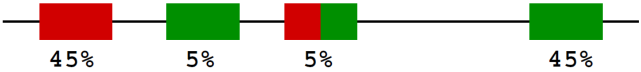
Sampling Bias

Most fundamental **challenge** of Active Learning.

In the Active Learning phase, algorithm starts to sample points closer to the boundary, getting further from the real distribution $\mathbb{P}$ of the data at each new query.

The chosen samples are **biased**, they aren't good representations of the actual distribution $\mathbb{P}$.

Example:



Two Faces of Active Learning

- *Efficient search through the hypothesis space*
  - Looks to shrink the search space as efficiently as possible
  - Solid research already exists, yielding lower label complexity (number of labels queried in order to achieve a given rate of accuracy) than supervised learning
- *Exploiting cluster structure in data*
  - Analyses the underlying structure of data, exploiting existence of clusters with similar labels
  - Still an unexplored area of Active Learning

Two Faces of Active Learning

- *Efficient search through the hypothesis space*
  - Looks to shrink the search space as efficiently as possible
  - Solid research already exists, yielding lower label complexity (number of labels queried in order to achieve a given rate of accuracy) than supervised learning
- *Exploiting cluster structure in data*
  - Analyses the underlying structure of data, exploiting existence of clusters with similar labels
  - Still an unexplored type of Active Learning

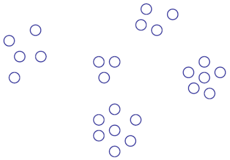
Exploiting cluster structure in data

When dealing with data that may be clustered:

- What may be done to avoid Sampling Bias
- What is done to manipulate clusters
- How to know when to assign labels

Exploiting cluster structure in data

Example:



Initial intuition might be:

"We just need five labels!""

Challenges: In general, the cluster structure (i) is not so clearly defined and (ii) exists at many levels of granularity. And (iii) the clusters themselves might not be pure in their labels.

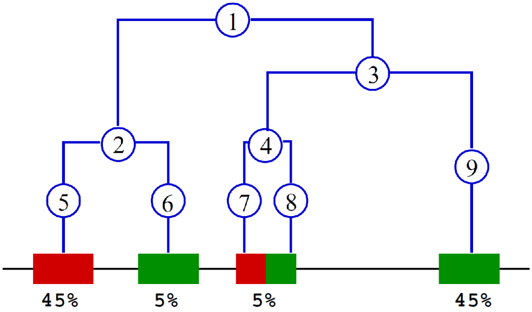
How to exploit whatever structure happens to exist?

Hierarchical Sampling for Active Learning

- Get correct labels (find best classifier) using fewest amount of samples as possible
- Use clustering structure as indication to propagate queried labels to similar points
- Avoid sampling bias

Hierarchical Sampling for Active Learning

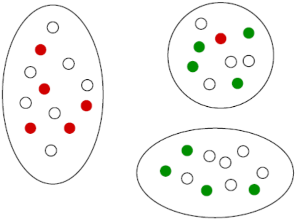
- Get correct labels (find best classifier) using fewest amount of samples as possible
- Use clustering structure as indication to propagate queried labels to similar points
- Avoid sampling bias
- Starting point:



Hierarchical Sampling for Active Learning

Basic functioning:

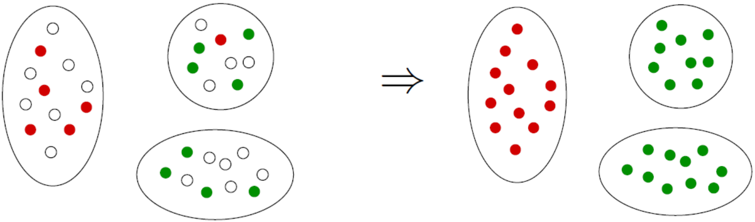
- Start with a clustering of the data
- Obtain the labels for some *randomly-chosen* samples in each cluster
- Assign the majority label to the whole cluster
- Build a classifier from the "fully labelled" data



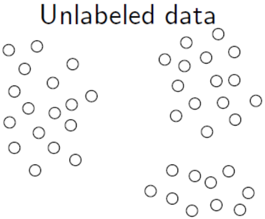
Hierarchical Sampling for Active Learning

Basic functioning:

- Start with a clustering of the data
- Obtain the labels for some *randomly-chosen* samples in each cluster
- Assign the majority label to the whole cluster
- Build a classifier from the "fully labelled" data



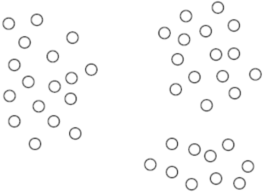
Hierarchical Sampling for Active Learning



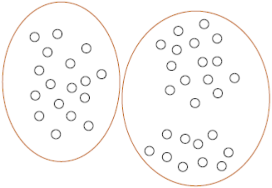
Hierarchical Sampling for Active Learning

Hierarchical Sampling for Active Learning

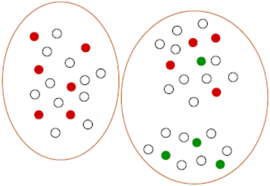
Unlabeled data



Find a clustering

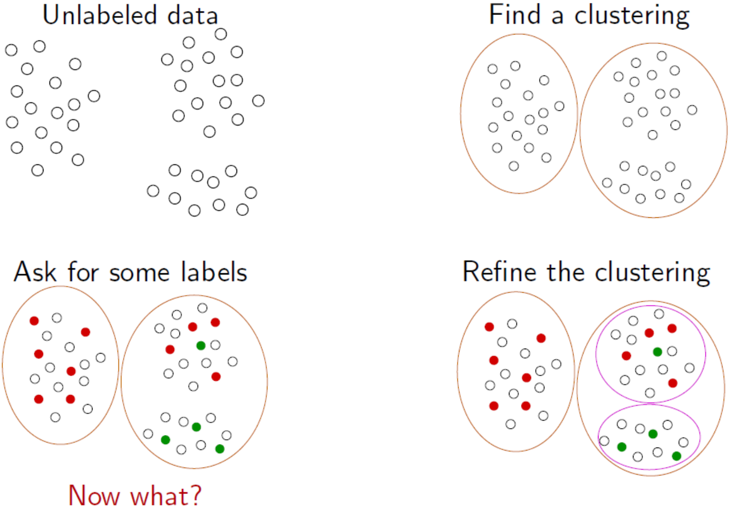


Ask for some labels



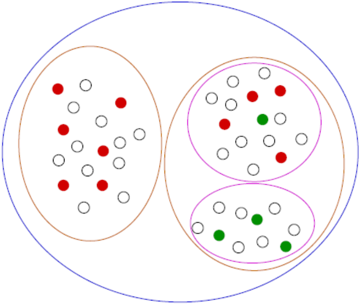
Now what?

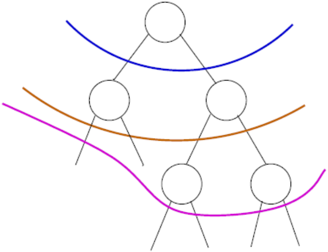
Hierarchical Sampling for Active Learning



Hierarchical Sampling for Active Learning

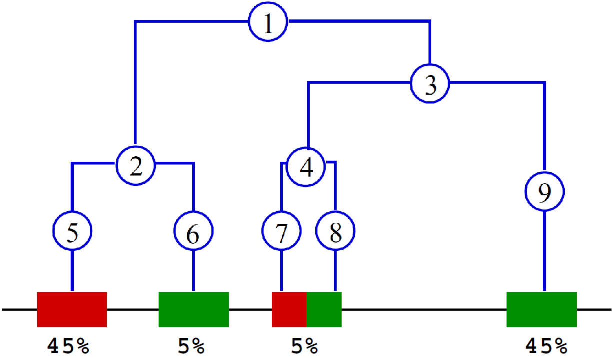
The clustering can be represented by a tree:





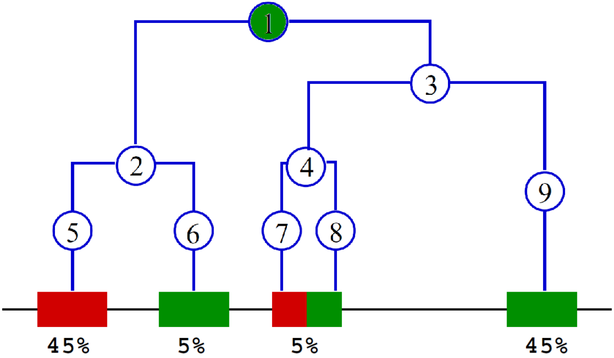
Hierarchical Sampling for Active Learning

Working example:



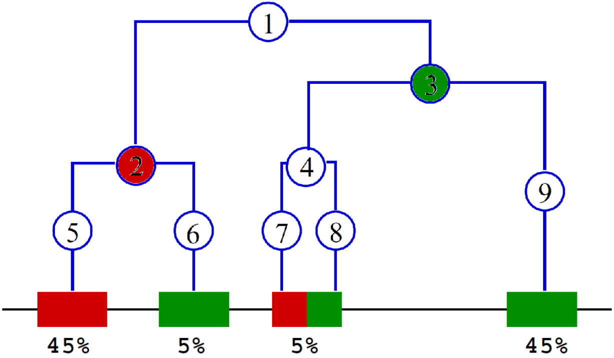
Hierarchical Sampling for Active Learning

Working example:



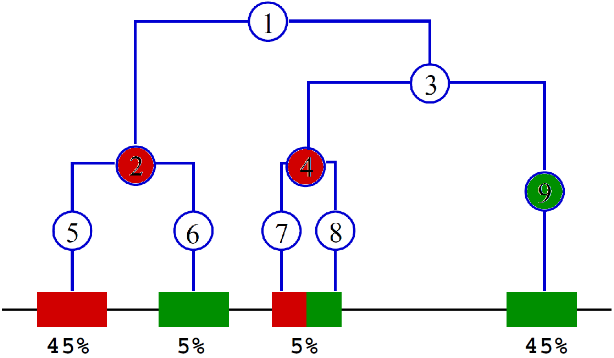
Hierarchical Sampling for Active Learning

Working example:



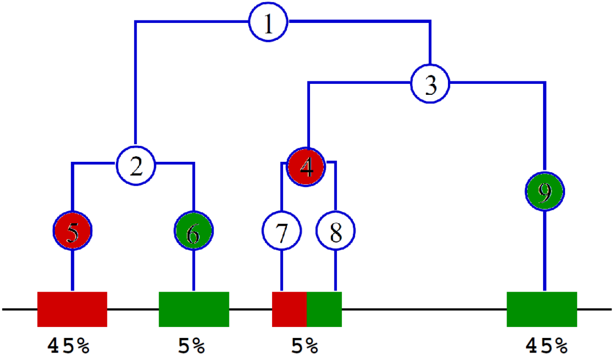
Hierarchical Sampling for Active Learning

Working example:



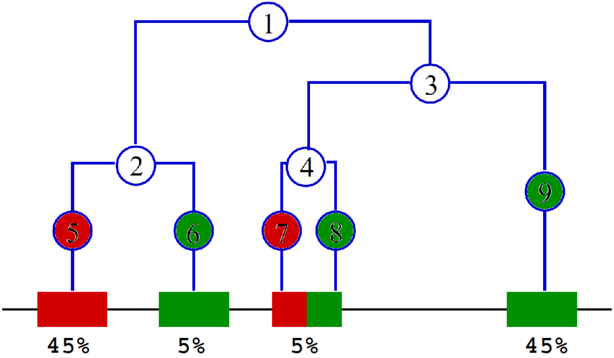
Hierarchical Sampling for Active Learning

Working example:



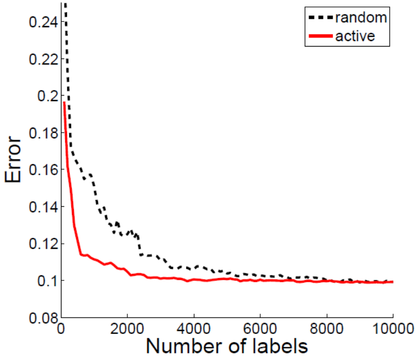
Hierarchical Sampling for Active Learning

Working example:



Active Sampling vs. Random Sampling

Active sampling	vs.	**Random sampling**
Less queries needed	vs.	More queries needed
Reduced sampling bias	vs.	No sampling bias



Algorithm Analysis

Pros:

- Tackles sampling bias efficiently
- Needs considerably less queries to achieve a classifier with a small error than supervised learning

Cons:

- Depends harshly on obtaining the initial hierarchical structure
- Is not suitable for some cluster structures
- (Works best when there exists a pruning of the tree that is small and a significant fraction of its clusters are pure)

Closing Remarks

Active Learning appears to be an interesting area

Can be an interesting optimization tool

Similarities with PAC Learning and Bandits models

References

Sanjoy Dasgupta, *The two faces of active learning*, Algorithmic Learning Theory (Berlin, Heidelberg) (Ricard Gavaldà, Gábor Lugosi, Thomas Zeugmann, and Sandra Zilles, eds.), Springer Berlin Heidelberg, 2009, pp. 1–1.

Sanjoy Dasgupta and Daniel Hsu, *Hierarchical sampling for active learning*, Proceedings of the 25th International Conference on Machine Learning (New York, NY, USA), ICML '08, Association for Computing Machinery, 2008, p. 208–215.