Annotation on the cheap

Sanjoy Dasgupta

University of California, San Diego

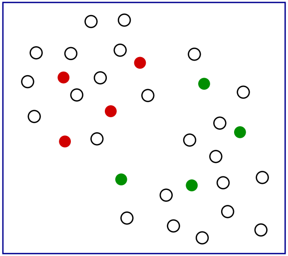
Active learning of classifiers

Unlabeled data is often plentiful and cheap: documents off the web, speech samples, images, video. *But labeling can be expensive.*

Active learning of classifiers

Unlabeled data is often plentiful and cheap: documents off the web, speech samples, images, video. *But labeling can be expensive.*

Active learning: Machine learns a classifier by querying just a few labels, choosing wisely and adaptively.

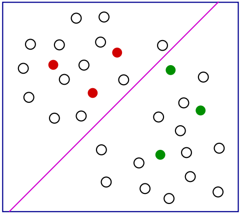


- Good querying schemes?
- Tradeoff between # labels and error rate of final classifier?

Algorithms for active learning

1. **Use the current best classifier to choose the next query.**

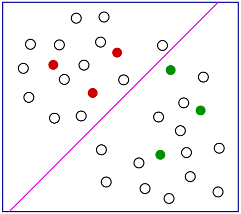
   Fit a classifier to the labels seen so far
   Query the unlabeled point that is closest to the boundary
   (or most uncertain, or most likely to decrease overall uncertainty,...)



Algorithms for active learning

1. **Use the current best classifier to choose the next query.**

    Fit a classifier to the labels seen so far
    Query the unlabeled point that is closest to the boundary
    (or most uncertain, or most likely to decrease overall uncertainty,...)



2. **Use the current version space to choose the next query.**
    E.g. Query-by-committee.

Sampling bias

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
(or most likely to decrease overall uncertainty, etc)

Sampling bias

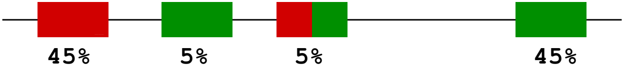
Start with a pool of unlabeled data

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
(or most likely to decrease overall uncertainty, etc)

Example: data in $\mathbb{R}$, $\mathcal{H} = \{\text{thresholds}\}$.



Sampling bias

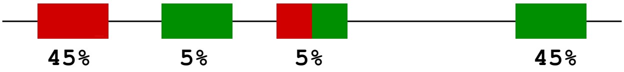
Start with a pool of unlabeled data

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
(or most likely to decrease overall uncertainty, etc)

Example: data in $\mathbb{R}$, $\mathcal{H} = \{\text{thresholds}\}$.



Even with infinitely many labels, converges to a classifier with 5% error instead of the best achievable, 2.5%. *Not consistent.*

Sampling bias

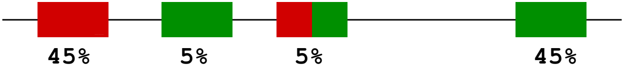
Start with a pool of unlabeled data

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
(or most likely to decrease overall uncertainty, etc)

Example: data in $\mathbb{R}$, $\mathcal{H} = \{\text{thresholds}\}$.



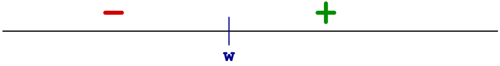
Even with infinitely many labels, converges to a classifier with 5% error instead of the best achievable, 2.5%. *Not consistent.*

Question: Is there a generic fix to uncertainty-based heuristics that makes them consistent?

Theory of active learning

1. Threshold functions on the real line ($\mathcal{X} = \mathbb{R}, \mathcal{Y} = \{+1, -1\}$)

$\mathcal{H} = \{h_w : w \in \mathbb{R}\}$
$h_w(x) = 1(x \geq w)$

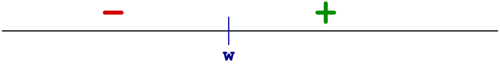


Supervised: for misclassification error $\leq \epsilon$, need $\approx 1/\epsilon$ labeled points.

Theory of active learning

1. Threshold functions on the real line ($\mathcal{X} = \mathbb{R}, \mathcal{Y} = \{+1, -1\}$)

$\mathcal{H} = \{h_w : w \in \mathbb{R}\}$
$h_w(x) = 1(x \geq w)$



Supervised: for misclassification error $\leq \epsilon$, need $\approx 1/\epsilon$ labeled points.

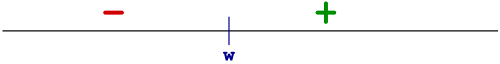
Active learning: instead, start with $1/\epsilon$ *unlabeled* points.

Binary search: need just $\log 1/\epsilon$ labels, from which the rest can be inferred. *Exponential improvement in label complexity*.

Theory of active learning

1. Threshold functions on the real line ($\mathcal{X} = \mathbb{R}, \mathcal{Y} = \{+1, -1\}$)

$\mathcal{H} = \{h_w : w \in \mathbb{R}\}$
$h_w(x) = 1(x \geq w)$



Supervised: for misclassification error $\leq \epsilon$, need $\approx 1/\epsilon$ labeled points.

Active learning: instead, start with $1/\epsilon$ *unlabeled* points.

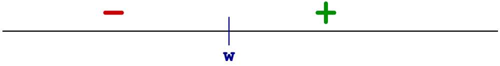
Binary search: need just $\log 1/\epsilon$ labels, from which the rest can be inferred. *Exponential improvement in label complexity*.

2. Various generalizations to other hypothesis classes

Theory of active learning

1. Threshold functions on the real line ($\mathcal{X} = \mathbb{R}, \mathcal{Y} = \{+1, -1\}$)

$\mathcal{H} = \{h_w : w \in \mathbb{R}\}$
$h_w(x) = 1(x \geq w)$



Supervised: for misclassification error $\leq \epsilon$, need $\approx 1/\epsilon$ labeled points.

Active learning: instead, start with $1/\epsilon$ *unlabeled* points.

Binary search: need just $\log 1/\epsilon$ labels, from which the rest can be inferred. *Exponential improvement in label complexity*.

2. Various generalizations to other hypothesis classes

But there's a basic problem with the whole model.

Active annotation

Input:

- Finite set of data points $\{x_1, \ldots, x_n\}$, each of which has an associated label y_i that is initially missing.
- Parameters $0 < \delta, \epsilon < 1$.
- Access to an oracle that can supply any label y_i, and perhaps other information as well.

Output:
A set of labels $\widehat{y}_1, \ldots, \widehat{y}_n$ such that with probability at least $1 - \delta$, at most an ϵ fraction of these labels are incorrect, that is,

$$\sum_i 1(y_i \neq \widehat{y}_i) \ \leq \ \epsilon n.$$

Goal: Minimize calls to the oracle.

Outline

1. Active annotation using label queries
   - Graph-based methods
   - Cluster-based methods

2. More general queries

Simple baseline: nearest neighbor

Naive but reasonable approach:

- Choose some points at random, get their labels
- Fill in the rest using nearest neighbor

Simple baseline: nearest neighbor

Naive but reasonable approach:

- Choose some points at random, get their labels
- Fill in the rest using nearest neighbor

What are some big ways in which we could improve upon this?

Simple baseline: nearest neighbor

Naive but reasonable approach:

- Choose some points at random, get their labels
- Fill in the rest using nearest neighbor

What are some big ways in which we could improve upon this?

1. More intelligent querying

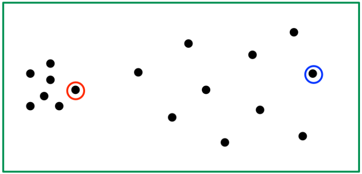
Simple baseline: nearest neighbor

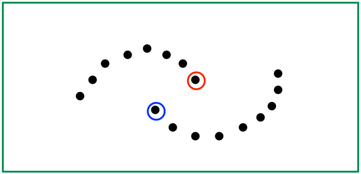
Naive but reasonable approach:

- Choose some points at random, get their labels
- Fill in the rest using nearest neighbor

What are some big ways in which we could improve upon this?

1. More intelligent querying
2. Something more attuned to underlying structure like clusters and manifolds

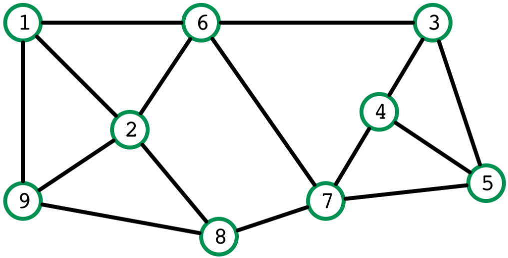




Active learning on graphs (Zhu-Ghahramani-Lafferty)

Given n unlabeled points, build neighborhood graph $G = (V, E)$:

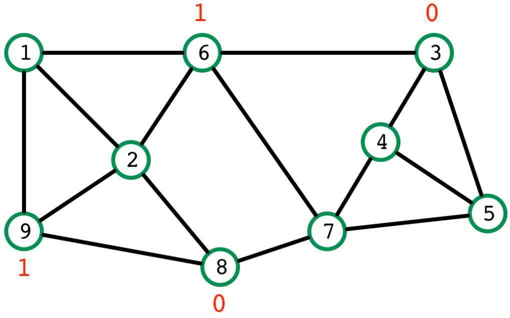
- One node per data point: $V = [n]$
- Edges between nearby points, with similarity weights w_{ij}



Active learning on graphs (Zhu-Ghahramani-Lafferty)

Given n unlabeled points, build neighborhood graph $G = (V, E)$:

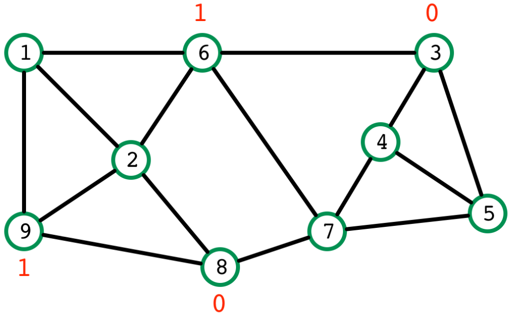
- One node per data point: $V = [n]$
- Edges between nearby points, with similarity weights w_{ij}



Active learning on graphs (Zhu-Ghahramani-Lafferty)

Given n unlabeled points, build neighborhood graph $G = (V, E)$:

- One node per data point: $V = [n]$
- Edges between nearby points, with similarity weights w_{ij}



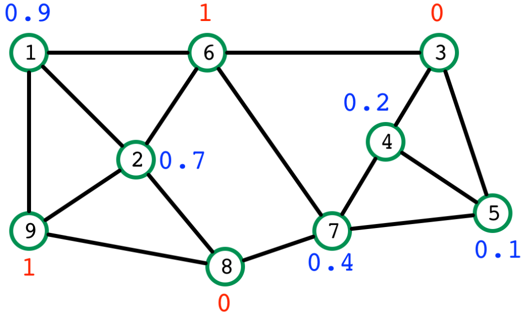
Given labels y_i on some subset of points $A \subset [n]$, find $f : [n] \to [0, 1]$:

Minimize: $\displaystyle\sum_{i,j} w_{ij}(f_i - f_j)^2$ subject to $f_i = y_i$ on $i \in A$.

Active learning on graphs (Zhu-Ghahramani-Lafferty)

Given n unlabeled points, build neighborhood graph $G = (V, E)$:

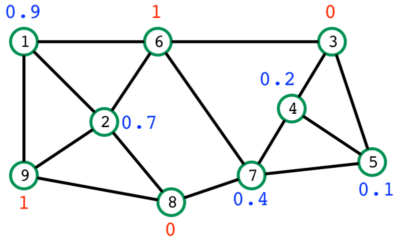
- One node per data point: $V = [n]$
- Edges between nearby points, with similarity weights w_{ij}



Given labels y_i on some subset of points $A \subset [n]$, find $f : [n] \to [0, 1]$:

$$\text{Minimize:} \quad \sum_{i,j} w_{ij}(f_i - f_j)^2 \quad \text{subject to } f_i = y_i \text{ on } i \in A.$$

Active querying



Query the point that most reduces overall uncertainty

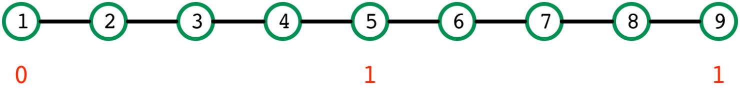
- Uncertainty in f:

$$U(f) = \sum_{i=1}^{n} \min(f_i, 1 - f_i).$$

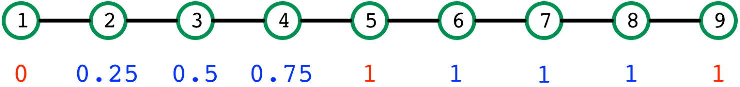
- To assess the effect of querying point i:
  - If its label is 1, then new f will be (say) f^+
  - It its label is 0, then new f will be (say) f^-
  - Estimated uncertainty after query: $f_i U(f^+) + (1 - f_i) U(f^-)$

Lack of consistency

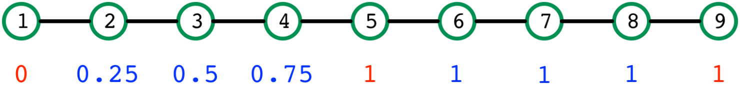
Lack of consistency



Lack of consistency



Lack of consistency

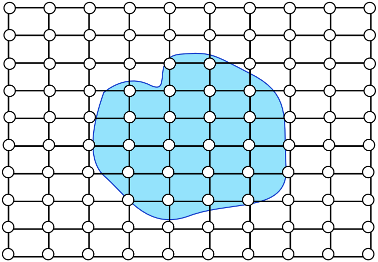


Will never query the right half of the points!

Another graph-based approach (Dasarthy-Nowak-Zhu)

Input: a **neighborhood graph** G whose nodes are the data points x.

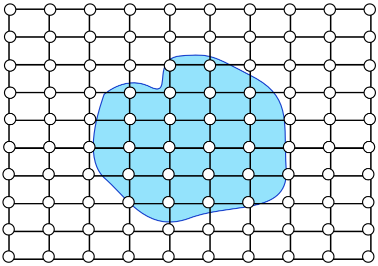
- Each node has an unknown label.
- Goal: find the *cut-edges* in this graph that separate two labels.



Another graph-based approach (Dasarthy-Nowak-Zhu)

Input: a **neighborhood graph** G whose nodes are the data points x.

- Each node has an unknown label.
- Goal: find the *cut-edges* in this graph that separate two labels.



What should label complexity depend upon?

- # cut edges
- log(diameter of graph)
- 1/(proportion of each class)

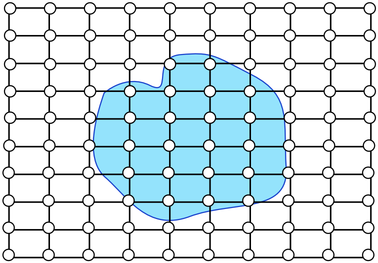
The S^2 algorithm (Dasarthy-Nowak-Zhu)

(For binary labels)

Keep going until budget runs out:

- If $\exists$ labeled nodes of opposite polarity that are connected in G:
  - Find the shortest path connecting nodes of opposite label.
  - Query its midpoint.

  Else:
  - Pick a random point and query it.
- Remove any newly-revealed cut edges from the graph G.



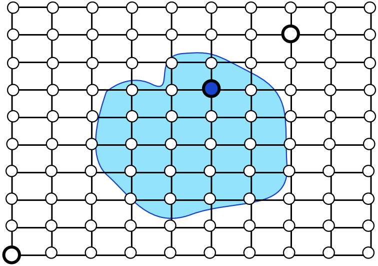
The S^2 algorithm (Dasarthy-Nowak-Zhu)

(For binary labels)

Keep going until budget runs out:

- If $\exists$ labeled nodes of opposite polarity that are connected in G:
  - Find the shortest path connecting nodes of opposite label.
  - Query its midpoint.

  Else:
  - Pick a random point and query it.
- Remove any newly-revealed cut edges from the graph G.



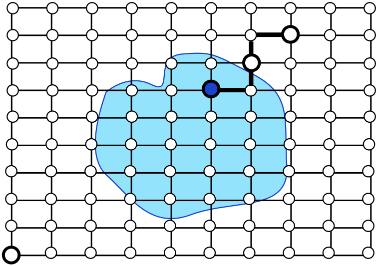
The S^2 algorithm (Dasarthy-Nowak-Zhu)

(For binary labels)

Keep going until budget runs out:

- If $\exists$ labeled nodes of opposite polarity that are connected in G:
  - Find the shortest path connecting nodes of opposite label.
  - Query its midpoint.

  Else:
  - Pick a random point and query it.
- Remove any newly-revealed cut edges from the graph G.



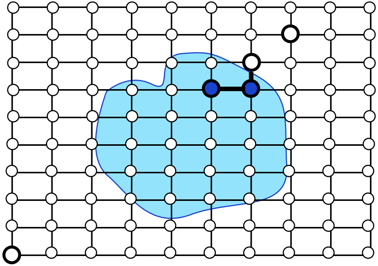
The S^2 algorithm (Dasarthy-Nowak-Zhu)

(For binary labels)

Keep going until budget runs out:

- If $\exists$ labeled nodes of opposite polarity that are connected in G:
  - Find the shortest path connecting nodes of opposite label.
  - Query its midpoint.

  Else:
  - Pick a random point and query it.
- Remove any newly-revealed cut edges from the graph G.



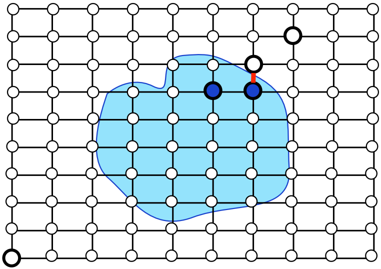
The S^2 algorithm (Dasarthy-Nowak-Zhu)

(For binary labels)

Keep going until budget runs out:

- If $\exists$ labeled nodes of opposite polarity that are connected in G:
  - Find the shortest path connecting nodes of opposite label.
  - Query its midpoint.

  Else:
  - Pick a random point and query it.
- Remove any newly-revealed cut edges from the graph G.



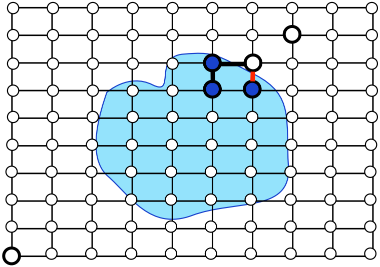
The S^2 algorithm (Dasarthy-Nowak-Zhu)

(For binary labels)

Keep going until budget runs out:

- If $\exists$ labeled nodes of opposite polarity that are connected in G:
  - Find the shortest path connecting nodes of opposite label.
  - Query its midpoint.

  Else:
  - Pick a random point and query it.
- Remove any newly-revealed cut edges from the graph G.



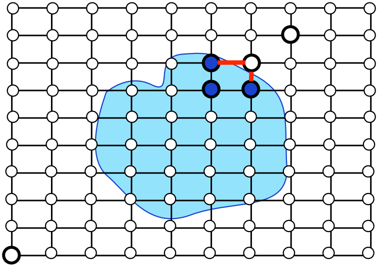
The S^2 algorithm (Dasarthy-Nowak-Zhu)

(For binary labels)

Keep going until budget runs out:

- If $\exists$ labeled nodes of opposite polarity that are connected in G:
  - Find the shortest path connecting nodes of opposite label.
  - Query its midpoint.

  Else:
  - Pick a random point and query it.
- Remove any newly-revealed cut edges from the graph G.



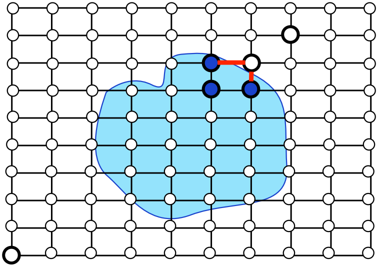
The S^2 algorithm (Dasarthy-Nowak-Zhu)

(For binary labels)

Keep going until budget runs out:

- If $\exists$ labeled nodes of opposite polarity that are connected in G:
  - Find the shortest path connecting nodes of opposite label.
  - Query its midpoint.

  Else:
  - Pick a random point and query it.
- Remove any newly-revealed cut edges from the graph G.



Graph-specific label complexity + nonparametric generalization bounds.

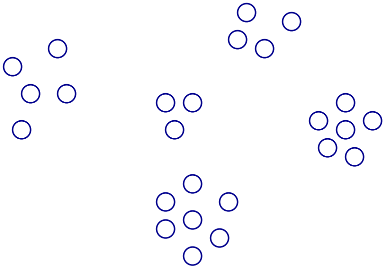
Outline

1. Active annotation using label queries
   - Graph-based methods
   - Cluster-based methods

2. More general queries

Exploiting cluster structure in data

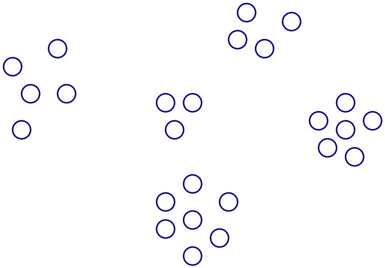
Suppose the unlabeled data looks like this.



Then perhaps we just need five labels.

Exploiting cluster structure in data

Suppose the unlabeled data looks like this.



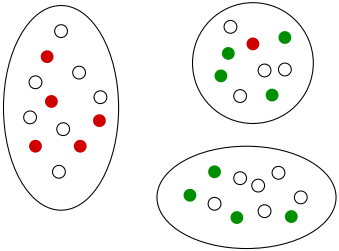
Then perhaps we just need five labels.

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

Exploiting cluster structure in data [D-Hsu]

Basic primitive:

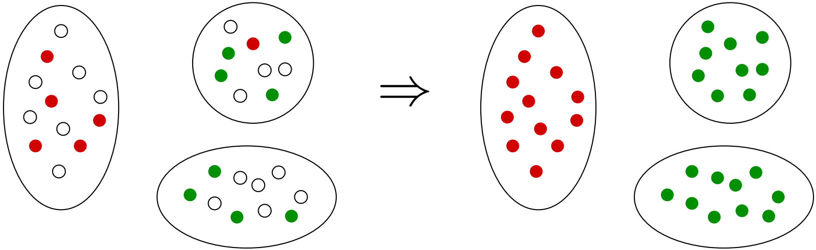
- Find a clustering of the data
- Sample a few *randomly-chosen* points in each cluster
- Assign each cluster its majority label



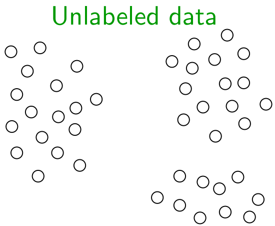
Exploiting cluster structure in data [D-Hsu]

Basic primitive:

- Find a clustering of the data
- Sample a few *randomly-chosen* points in each cluster
- Assign each cluster its majority label

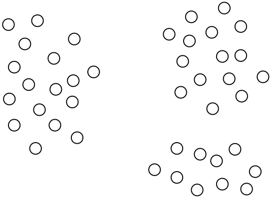


Finding the right granularity

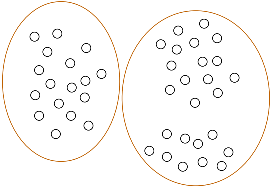


Finding the right granularity

Unlabeled data

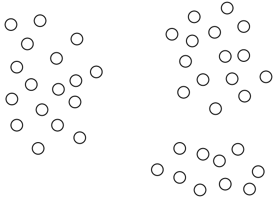


Find a clustering

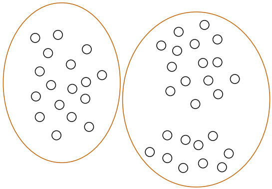


Finding the right granularity

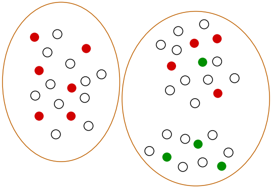
Unlabeled data



Find a clustering



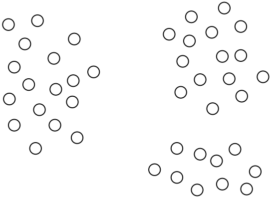
Ask for some labels



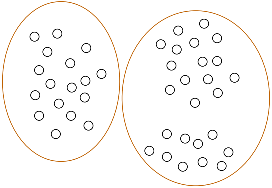
(random sampling within clusters)

Finding the right granularity

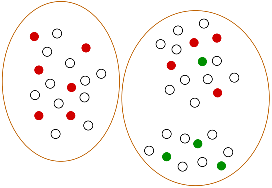
Unlabeled data



Find a clustering



Ask for some labels

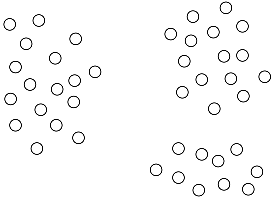


(random sampling within clusters)

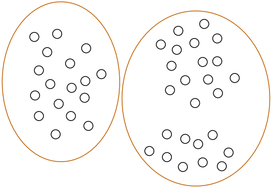
Now what?

Finding the right granularity

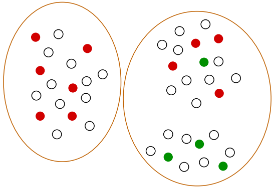
Unlabeled data



Find a clustering



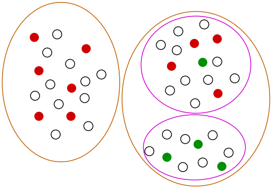
Ask for some labels



(random sampling within clusters)

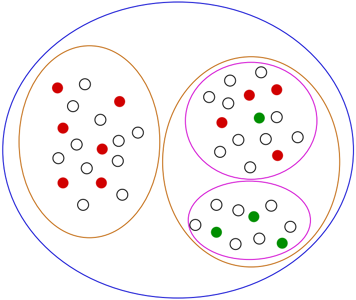
Now what?

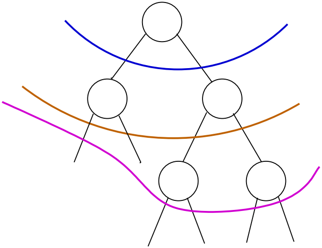
Refine the clustering



Queried points are also randomly distributed within the new clusters.

Using a hierarchical clustering





Rules:

- Always work with some pruning of the hierarchy: a clustering induced by the tree.
- To make a query, pick a cluster, whereupon a random point in that cluster will be chosen and its label will be queried.
- As time progresses, the current pruning can only move down the tree, not back up.

Hierarchical sampling framework

So far: a framework for sampling that avoids bias. Still need to specify:

1. How the initial hierarchical clustering is built.
2. Rule for deciding which cluster to query.
3. Rule for deciding when to move down from a cluster to its children.

Hierarchical sampling framework

So far: a framework for sampling that avoids bias. Still need to specify:

1. How the initial hierarchical clustering is built.
2. Rule for deciding which cluster to query.
3. Rule for deciding when to move down from a cluster to its children.

D-Hsu:

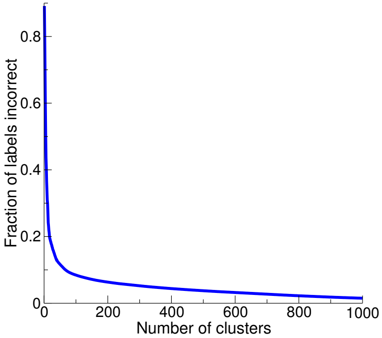
- Tree: Ward's agglomerative hierarchical clustering.
- Query least-pure cluster.
- Move down when confidence intervals indicate cluster's purity is below some threshold.

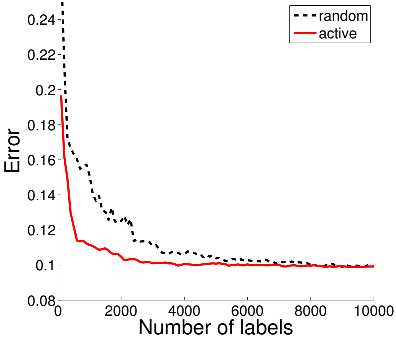
Urner-Wulff-Ben-David, Ben-David-Kpotufe-Urner:

- Tree: k-d tree or RP tree.
- Query fixed number of points in each cluster.
- Move down if there is any disagreement in the labels obtained for a cluster.

Example: MNIST digits

Hierarchy built using Ward's agglomerative clustering (k-means cost function) with Euclidean distance.





Outline

1. Active annotation using label queries
   - Graph-based methods
   - Cluster-based methods

2. More general queries

Explanation-based learning

In addition to labels, the human might provide an explanation, for instance in the form of relevant features.

Explanation-based learning

In addition to labels, the human might provide an explanation, for instance in the form of relevant features.

HOW FINISHES ENGLAND

NZ JUST TOO GOOD

England's hopes of a Test World Medal title are all but over. Following New Zealand's convincing win at Lords, England now have to hope India beat the Black Caps in the first test in New Zealand, and then crush India in one match at Kolkata. Given how well India have been playing lately, it will be a tough task for England to have a comprehensive victory, particularly in India.

The reason for this is New Zealand's resolute, assured performance against England which saw them take the match inside four days. Martin Guptill, Tim Southee, Kane Williamson and Jamie How were the outstanding performers- Southee picking up the Npower Man of the Match Award. Only Pietersen made something of a start for England, with no bowler taking more than 4 wickets for England.

It was a match in which several players put up their hands when the experienced players did not perform as well as they would have liked to. The likes of Vettori, Oram, Bond and McCullum played their part in the match, but it was players like Guptill, How, Southee and Williamson who were the stars. Having always looked the part in test cricket, How and Guptill finally have some runs to show for their talent. Southee has always enjoyed bowling in and against England, and here he made the world sit up and take notice with a fine bowling performance in the 2nd Innings, showing resilience after a disappointing 1st innings.

As for Williamson, expect big things from this boys in the future- he played with outstanding flair and yet great maturity.

Unfortunately, Pietersen was the only performer for England, although Prior did keep and bat solidly. Again, it was a game of missed opportunities.

Explanation-based learning

In addition to labels, the human might provide an explanation, for instance in the form of relevant features.

HOW FINISHES ENGLAND

NZ JUST TOO GOOD

England's hopes of a Test World Medal title are all but over. Following New Zealand's convincing win at Lords, England now have to hope India beat the Black Caps in the first test in New Zealand, and then crush India in one match at Kolkata. Given how well India have been playing lately, it will be a tough task for England to have a comprehensive victory, particularly in India.

The reason for this is New Zealand's resolute, assured performance against England which saw them take the match inside four days. Martin Guptill, Tim Southee, Kane Williamson and Jamie How were the outstanding performers- Southee picking up the Npower Man of the Match Award. Only Pietersen made something of a start for England, with no bowler taking more than 4 wickets for England.

It was a match in which several players put up their hands when the experienced players did not perform as well as they would have liked to. The likes of Vettori, Oram, Bond and McCullum played their part in the match, but it was players like Guptill, How, Southee and Williamson who were the stars. Having always looked the part in test cricket, How and Guptill finally have some runs to show for their talent. Southee has always enjoyed bowling in and against England, and here he made the world sit up and take notice with a fine bowling performance in the 2nd Innings, showing resilience after a disappointing 1st innings.

As for Williamson, expect big things from this boys in the future- he played with outstanding flair and yet great maturity.

Unfortunately, Pietersen was the only performer for England, although Prior did keep and bat solidly. Again, it was a game of missed opportunities.

Explanation-based learning

In addition to labels, the human might provide an explanation, for instance in the form of relevant features.

HOW FINISHES ENGLAND

NZ JUST TOO GOOD

England's hopes of a Test World Medal title are all but over. Following New Zealand's convincing win at Lords, England now have to hope India beat the Black Caps in the first test in New Zealand, and then crush India in one match at Kolkata. Given how well India have been playing lately, it will be a tough task for England to have a comprehensive victory, particularly in India.

The reason for this is New Zealand's resolute, assured performance against England which saw them take the match inside four days. Martin Guptill, Tim Southee, Kane Williamson and Jamie How were the outstanding performers- Southee picking up the Npower Man of the Match Award. Only Pietersen made something of a start for England, with no bowler taking more than 4 wickets for England.

It was a match in which several players put up their hands when the experienced players did not perform as well as they would have liked to. The likes of Vettori, Oram, Bond and McCullum played their part in the match, but it was players like Guptill, How, Southee and Williamson who were the stars. Having always looked the part in test cricket, How and Guptill finally have some runs to show for their talent. Southee has always enjoyed bowling in and against England, and here he made the world sit up and take notice with a fine bowling performance in the 2nd Innings, showing resilience after a disappointing 1st innings.

As for Williamson, expect big things from this boys in the future- he played with outstanding flair and yet great maturity.

Unfortunately, Pietersen was the only performer for England, although Prior did keep and bat solidly. Again, it was a game of missed opportunities.

- Benefit of explanations over labels alone?

Explanation-based learning

In addition to labels, the human might provide an explanation, for instance in the form of relevant features.

HOW FINISHES ENGLAND

NZ JUST TOO GOOD

England's hopes of a Test World Medal title are all but over. Following New Zealand's convincing win at Lords, England now have to hope India beat the Black Caps in the first test in New Zealand, and then crush India in one match at Kolkata. Given how well India have been playing lately, it will be a tough task for England to have a comprehensive victory, particularly in India.

The reason for this is New Zealand's resolute, assured performance against England which saw them take the match inside four days. Martin Guptill, Tim Southee, Kane Williamson and Jamie How were the outstanding performers- Southee picking up the Npower Man of the Match Award. Only Pietersen made something of a start for England, with no bowler taking more than 4 wickets for England.

It was a match in which several players put up their hands when the experienced players did not perform as well as they would have liked to. The likes of Vettori, Oram, Bond and McCullum played their part in the match, but it was players like Guptill, How, Southee and Williamson who were the stars. Having always looked the part in test cricket, How and Guptill finally have some runs to show for their talent. Southee has always enjoyed bowling in and against England, and here he made the world sit up and take notice with a fine bowling performance in the 2nd Innings, showing resilience after a disappointing 1st innings.

As for Williamson, expect big things from this boys in the future- he played with outstanding flair and yet great maturity.

Unfortunately, Pietersen was the only performer for England, although Prior did keep and bat solidly. Again, it was a game of missed opportunities.

- Benefit of explanations over labels alone?
- How to deal with ambiguity of feedback?

Predictive feature feedback (D-Poulis)

HOW FINISHES ENGLAND

NZ JUST TOO GOOD

England's hopes of a Test World Medal title are all but over. Following New Zealand's convincing win at Lords, England now have to hope India beat the Black Caps in the first test in New Zealand, and then crush India in one match at Kolkata. Given how well India have been playing lately, it will be a tough task for England to have a comprehensive victory, particularly in India.

The reason for this is New Zealand's resolute, assured performance against England which saw them take the match inside four days. Martin Guptill, Tim Southee, Kane Williamson and Jamie How were the outstanding performers- Southee picking up the Npower Man of the Match Award. Only Pietersen made something of a start for England, with no bowler taking more than 4 wickets for England.

It was a match in which several players put up their hands when the experienced players did not perform as well as they would have liked to. The likes of Vettori, Oram, Bond and McCullum played their part in the match, but it was players like Guptill, How, Southee and Williamson who were the stars. Having always looked the part in test cricket, How and Guptill finally have some runs to show for their talent. Southee has always enjoyed bowling in and against England, and here he made the world sit up and take notice with a fine bowling performance in the 2nd Innings, showing resilience after a disappointing 1st innings.

As for Williamson, expect big things from this boys in the future- he played with outstanding flair and yet great maturity.

Unfortunately, Pietersen was the only performer for England, although Prior did keep and bat solidly. Again, it was a game of missed opportunities.

Label: “sports”. Highlighted word: “wicketkeeper”.

Predictive feature feedback (D-Poulis)

HOW FINISHES ENGLAND

NZ JUST TOO GOOD

England's hopes of a Test World Medal title are all but over. Following New Zealand's convincing win at Lords, England now have to hope India beat the Black Caps in the first test in New Zealand, and then crush India in one match at Kolkata. Given how well India have been playing lately, it will be a tough task for England to have a comprehensive victory, particularly in India.

The reason for this is New Zealand's resolute, assured performance against England which saw them take the match inside four days. Martin Guptill, Tim Southee, Kane Williamson and Jamie How were the outstanding performers- Southee picking up the Npower Man of the Match Award.

Only Pietersen made something of a start for England, with no bowler taking more than 4 wickets for England.

It was a match in which several players put up their hands when the experienced players did not perform as well as they would have liked to. The likes of Vettori, Oram, Bond and McCullum played their part in the match, but it was players like Guptill, How, Southee and Williamson who were the stars. Having always looked the part in test cricket, How and Guptill finally have some runs to show for their talent. Southee has always enjoyed bowling in and against England, and here he made the world sit up and take notice with a fine bowling performance in the 2nd Innings, showing resilience after a disappointing 1st innings.

As for Williamson, expect big things from this boys in the future- he played with outstanding flair and yet great maturity.

Unfortunately, Pietersen was the only performer for England, although Prior did keep and bat solidly. Again, it was a game of missed opportunities.

Label: "sports". Highlighted word: "wicketkeeper".

Two ways to deal with this:

1. Add a weak rule:

$$\texttt{wicketkeeper} \Longrightarrow \texttt{sports}$$

Predictive feature feedback (D-Poulis)

HOW FINISHES ENGLAND

NZ JUST TOO GOOD

England's hopes of a Test World Medal title are all but over. Following New Zealand's convincing win at Lords, England now have to hope India beat the Black Caps in the first test in New Zealand, and then crush India in one match at Kolkata. Given how well India have been playing lately, it will be a tough task for England to have a comprehensive victory, particularly in India.
The reason for this is New Zealand's resolute, assured performance against England which saw them take the match inside four days. Martin Guptill, Tim Southee, Kane Williamson and Jamie How were the outstanding performers- Southee picking up the Npower Man of the Match Award.
Only Pietersen made something of a start for England, with no bowler taking more than 4 wickets for England.

It was a match in which several players put up their hands when the experienced players did not perform as well as they would have liked to. The likes of Vettori, Oram, Bond and McCullum played their part in the match, but it was players like Guptill, How, Southee and Williamson who were the stars. Having always looked the part in test cricket, How and Guptill finally have some runs to show for their talent. Southee has always enjoyed bowling in and against England, and here he made the world sit up and take notice with a fine bowling performance in the 2nd Innings, showing resilience after a disappointing 1st innings.
As for Williamson, expect big things from this boys in the future- he played with outstanding flair and yet great maturity.
Unfortunately, Pietersen was the only performer for England, although Prior did keep and bat solidly. Again, it was a game of missed opportunities.

Label: "sports". Highlighted word: "wicketkeeper".

Two ways to deal with this:

1. Add a weak rule:

$$\texttt{wicketkeeper} \implies \texttt{sports}$$

2. Suppose topic modeling is used on the corpus, and this instance of "wicketkeeper" is assigned to topic 143. Add weak rule:

$$\texttt{topic 143} \implies \texttt{sports}$$

Predictive feature feedback: results

A generative model for the labels:

- There are k topics, and an unknown map:

  $$\ell : [k] \rightarrow \mathcal{Y} \cap \{?\}.$$

  Topics t with $\ell(t) =?$ are *uninformative*.
- If a document has topic distribution $(\theta_1, \ldots, \theta_k)$:
  - Pick an informative topic t with probability $\propto \theta_t$
  - Assign label $\ell(t)$ to the document

Goal: given interaction with human, determine the mapping $\ell(\cdot)$.

Predictive feature feedback: results

A generative model for the labels:

- There are k topics, and an unknown map:

$$\ell : [k] \rightarrow \mathcal{Y} \cap \{?\}.$$

  Topics t with $\ell(t) = ?$ are *uninformative*.
- If a document has topic distribution $(\theta_1, \ldots, \theta_k)$:
  - Pick an informative topic t with probability $\propto \theta_t$
  - Assign label $\ell(t)$ to the document

Goal: given interaction with human, determine the mapping $\ell(\cdot)$.

1. This is NP-hard given only label feedback.

Predictive feature feedback: results

A generative model for the labels:

- There are k topics, and an unknown map:

$$\ell : [k] \to \mathcal{Y} \cap \{?\}.$$

  Topics t with $\ell(t) = ?$ are *uninformative*.
- If a document has topic distribution $(\theta_1, \ldots, \theta_k)$:
  - Pick an informative topic t with probability $\propto \theta_t$
  - Assign label $\ell(t)$ to the document

Goal: given interaction with human, determine the mapping $\ell(\cdot)$.

1. This is NP-hard given only label feedback.
2. With feature feedback, it is easy, using $O(k \log |\mathcal{Y}|)$ interactions.

Combining weak rules

The annotation problem:

- We have an unlabeled data set of n points
- We are able to query individual labels

Combining weak rules

The annotation problem:

- We have an unlabeled data set of n points
- We are able to query individual labels

Suppose we also have a collection of **weak rules** $h_1, \ldots, h_m$ that each make predictions on some of the points and abstain on others. E.g.:

- Rules-of-thumb from a human: wicketkeeper $\Longrightarrow$ `sports`
- Weak classifiers from other sources
- Each h_i could be a crowd-sourced worker

How can we make use of these?

Combining weak rules

The annotation problem:

- We have an unlabeled data set of n points
- We are able to query individual labels

Suppose we also have a collection of **weak rules** $h_1, \ldots, h_m$ that each make predictions on some of the points and abstain on others. E.g.:

- Rules-of-thumb from a human: wicketkeeper $\implies$ `sports`
- Weak classifiers from other sources
- Each h_i could be a crowd-sourced worker

How can we make use of these?

Common approach, e.g. Snorkel (Ratner-Bach-Ehrenberg-Re):

- Query a few labels at random
- Use these to estimate how accurate each h_i is, and possibly correlations between them
- Probabilistically combine the h_i

A game-theoretic approach (Balsubramani-Freund)

Goal: find weights $\alpha_1, \ldots, \alpha_m \geq 0$ for the weak rules $h_1, \ldots, h_m$ and predict using a weighted majority.

A game-theoretic approach (Balsubramani-Freund)

Goal: find weights $\alpha_1, \ldots, \alpha_m \geq 0$ for the weak rules $h_1, \ldots, h_m$ and predict using a weighted majority.

1. Using a few labeled points, obtain upper bounds on the error rate of each of the weak rules $h_1, \ldots, h_m$

A game-theoretic approach (Balsubramani-Freund)

Goal: find weights $\alpha_1, \ldots, \alpha_m \geq 0$ for the weak rules $h_1, \ldots, h_m$ and predict using a weighted majority.

1. Using a few labeled points, obtain upper bounds on the error rate of each of the weak rules $h_1, \ldots, h_m$
2. Let $V \subset \{-1, 1\}^n$ denote the set of all labelings of the unlabeled points that are consistent with these error rates.

A game-theoretic approach (Balsubramani-Freund)

Goal: find weights $\alpha_1, \ldots, \alpha_m \geq 0$ for the weak rules $h_1, \ldots, h_m$ and predict using a weighted majority.

1. Using a few labeled points, obtain upper bounds on the error rate of each of the weak rules $h_1, \ldots, h_m$
2. Let $V \subset \{-1, 1\}^n$ denote the set of all labelings of the unlabeled points that are consistent with these error rates.
3. Find the weighting α with smallest worst-case error on V:

$$\min_{\alpha} \max_{y \in V} \text{(error of weighted combination } \alpha \text{ on labeling } y\text{)}$$

A game-theoretic approach (Balsubramani-Freund)

Goal: find weights $\alpha_1, \ldots, \alpha_m \geq 0$ for the weak rules $h_1, \ldots, h_m$ and predict using a weighted majority.

1. Using a few labeled points, obtain upper bounds on the error rate of each of the weak rules $h_1, \ldots, h_m$
2. Let $V \subset \{-1, 1\}^n$ denote the set of all labelings of the unlabeled points that are consistent with these error rates.
3. Find the weighting α with smallest worst-case error on V:

$$\min_{\alpha} \max_{y \in V} \text{ (error of weighted combination } \alpha \text{ on labeling } y)$$

4. Has a nice game-theoretic interpretation and solution

Open problems

1. Graph-based annotation

Open problems

1. Graph-based annotation
2. Cluster-based annotation

Open problems

1. Graph-based annotation
2. Cluster-based annotation
3. Richer feedback