Learning with Deficient Supervision-II

CS771: Introduction to Machine Learning

Purushottam Kar

Announcements

- End sem examination: November 17, 1600-1900, L18,19,20, ERES
  - Open notes (handwritten only)
  - Please bring a pencil and eraser with you
  - Seating arrangement same as midsem – will announce again
- Project presentation schedule open for slot choice
  - 15 minute presentations for each group
  - Please choose a slot by November 18 1159 hrs
  - Please do not overwrite each other's slots
- Final project report due November 26 1159PM IST (premidnight)
  - No late submission deadline since grade submission is overdue by then
  - Please refer to Piazza post on how to write and structure your report

Recap

- We saw how label imbalance impacts classification problems
  - Binary, multi-class and multi-label
- For binary classification problems we saw three solutions
  - Resampling
    - Oversample rare class or undersample popular class
    - Not nice as either throws away data or bloats up training set
    - Not extendable to multi-class and multi-label settings elegantly
  - Reweighted learning
    - Assign different weights to point of different classes
    - Nicely extends to multi-class settings
  - Changing the evaluation/training loss function
    - F-measure, AUC etc
    - Algorithms more involved but the extra work does pay off!
- Today: learning with weak, active, and semi-supervision

Learning with Weak Supervision

Weak Supervision

Weak Supervision

- Refers to supervision at a high level

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance learning

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance learning

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance learning

Man, Horse, Dog

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance learning

Man, Horse, Dog

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance learning

Man, Horse, Dog

Man, Horse, Dog

Weak Supervision

Weak supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance learning

Man, Horse, Dog

Man, Horse, Dog

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance l

Weak supervision

Less effort to label

Man, Horse, Dog

Man, Horse, Dog

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance learning

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance learning


Not logged in Talk Contributions Create account Log in

WIKIPEDIA
The Free Encyclopedia

Article Talk Read View source View history Search Wikipedia

WIKIPEDIA ASIAN MONTH This November is the Wikipedia Asian Month. Come join us.

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export

Walt Disney

From Wikipedia, the free encyclopedia

This article is about the man. For the company he co-founded, see The Walt Disney Company. For other uses, see Walt Disney (disambiguation).

Walter Elias Disney (/ˈdɪzni/;[1] December 5, 1901 – December 15, 1966) was an American entrepreneur, animator, voice actor and film producer. A pioneer of the American animation industry, he introduced several developments in the production of cartoons. As a film producer, Disney holds the record for most Academy Awards earned by an individual, having won 22 Oscars from 59 nominations. He was presented with two Golden Globe Special Achievement Awards and an Emmy Award, among other honors. Several of his films are included in the National Film Registry by the Library of Congress.

Born in Chicago in 1901, Disney developed an early interest in drawing. He took art classes as a boy and got a job as a commercial illustrator at the age of 18. He moved to California in the early 1920s and set up the Disney Brothers Studio with his brother Roy. With Ub Iwerks, Walt developed the character Mickey Mouse in 1928, his first highly popular success; he also provided the voice for his creation in the early years. As the studio grew, Disney became more adventurous, introducing synchronized sound, full-color three-strip Technicolor, feature-length cartoons and technical developments in cameras. The results, seen in features such as *Snow White and the Seven Dwarfs* (1937), *Fantasia*, *Pinocchio* (both 1940), *Dumbo* (1941) and *Bambi* (1942), furthered the development of animated film. New animated and live-action films followed after World War II, including the critically successful *Cinderella* (1950) and *Mary Poppins* (1964), the latter of which received five Academy Awards.

In the 1950s, Disney expanded into the amusement park industry, and in 1955 he opened Disneyland. To fund the project he diversified into television programs, such as *Walt Disney's Disneyland* and *The Mickey Mouse Club*; he was also

Walt Disney

Disney in 1946

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance learning


Not logged in Talk Contributions Create account Log in

WIKIPEDIA
The Free Encyclopedia

Article Talk Read View source View history Search Wikipedia

WIKIPEDIA ASIAN MONTH This November is the Wikipedia Asian Month. Come join us.

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export

Walt Disney

From Wikipedia, the free encyclopedia

This article is about the man. For the company he co-founded, see The Walt Disney Company. For other uses, see Walt Disney (disambiguation).

Walter Elias Disney (/ˈdɪzni/;[1] December 5, 1901 – December 15, 1966) was an American entrepreneur, animator, voice actor and film producer. A pioneer of the American animation industry, he introduced several developments in the production of cartoons. As a film producer, Disney holds the record for most Academy Awards earned by an individual, having won 22 Oscars from 59 nominations. He was presented with two Golden Globe Special Achievement Awards and an Emmy Award, among other honors. Several of his films are included in the National Film Registry by the Library of Congress.

Born in Chicago in 1901, Disney developed an early interest in drawing. He took art classes as a boy and got a job as a commercial illustrator at the age of 18. He moved to California in the early 1920s and set up the Disney Brothers Studio with his brother Roy. With Ub Iwerks, Walt developed the character Mickey Mouse in 1928, his first highly popular success; he also provided the voice for his creation in the early years. As the studio grew, Disney became more adventurous, introducing synchronized sound, full-color three-strip Technicolor, feature-length cartoons and technical developments in cameras. The results, seen in features such as *Snow White and the Seven Dwarfs* (1937), *Fantasia*, *Pinocchio* (both 1940), *Dumbo* (1941) and *Bambi* (1942), furthered the development of animated film. New animated and live-action films followed after World War II, including the critically successful *Cinderella* (1950) and *Mary Poppins* (1964), the latter of which received five Academy Awards.

In the 1950s, Disney expanded into the amusement park industry, and in 1955 he opened Disneyland. To fund the project he diversified into television programs, such as *Walt Disney's Disneyland* and *The Mickey Mouse Club*; he was also

Walt Disney

Disney in 1946

CREATE(WALT DISNEY, MICKEY MOUSE)

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance learning

Not logged in Talk Contributions Create account Log in

WIKIPEDIA
The Free Encyclopedia

Article Talk Read View source View history Search Wikipedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export

WIKIPEDIA ASIAN MONTH — This November is the Wikipedia Asian Month. Come join us.

Walt Disney

From Wikipedia, the free encyclopedia

This article is about the man. For the company he co-founded, see The Walt Disney Company. For other uses, see Walt Disney (disambiguation).

Walter Elias Disney (/ˈdɪzni/;[1] December 5, 1901 – December 15, 1966) was an American entrepreneur, animator, voice actor and film producer. A pioneer of the American animation industry, he introduced several developments in the production of cartoons. As a film producer, Disney holds the record for most Academy Awards earned by an individual, having won 22 Oscars from 59 nominations. He was presented with two Golden Globe Special Achievement Awards and an Emmy Award, among other honors. Several of his films are included in the National Film Registry by the Library of Congress.

Born in Chicago in 1901, Disney developed an early interest in drawing. He took art classes as a boy and got a job as a commercial illustrator at the age of 18. He moved to California in the early 1920s and set up the Disney Brothers Studio with his brother Roy. With Ub Iwerks, Walt developed the character Mickey Mouse in 1928, his first highly popular success; he also provided the voice for his creation in the early years. As the studio grew, Disney became more adventurous, introducing synchronized sound, full-color three-strip Technicolor, feature-length cartoons and technical developments in cameras. The results, seen in features such as *Snow White and the Seven Dwarfs* (1937), *Fantasia*, *Pinocchio* (both 1940), *Dumbo* (1941) and *Bambi* (1942), furthered the development of animated film. New animated and live-action films followed after World War II, including the critically successful *Cinderella* (1950) and *Mary Poppins* (1964), the latter of which received five Academy Awards.

In the 1950s, Disney expanded into the amusement park industry, and in 1955 he opened Disneyland. To fund the project he diversified into television programs, such as *Walt Disney's Disneyland* and *The Mickey Mouse Club*; he was also

Walt Disney

Disney in 1946

CREATE(WALT DISNEY, MICKEY MOUSE)

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance learning


Not logged in Talk Contributions Create account Log in

Article Talk Read View source View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

WIKIPEDIA ASIAN MONTH This November is the Wikipedia Asian Month. Come join us.

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export

Walt Disney

From Wikipedia, the free encyclopedia

This article is about the man. For the company he co-founded, see The Walt Disney Company. For other uses, see Walt Disney (disambiguation).

Walter Elias Disney (/ˈdɪzni/;[1] December 5, 1901 – December 15, 1966) was an American entrepreneur, animator, voice actor and film producer. A pioneer of the American animation industry, he introduced several developments in the production of cartoons. As a film producer, Disney holds the record for most Academy Awards earned by an individual, having won 22 Oscars from 59 nominations. He was presented with two Golden Globe Special Achievement Awards and an Emmy Award, among other honors. Several of his films are included in the National Film Registry by the Library of Congress.

Born in Chicago in 1901, Disney developed an early interest in drawing. He took art classes as a boy and got a job as a commercial illustrator at the age of 18. He moved to California in the early 1920s and set up the Disney Brothers Studio with his brother Roy. With Ub Iwerks, Walt developed the character Mickey Mouse in 1928, his first highly popular success; he also provided the voice for his creation in the early years. As the studio grew, Disney became more adventurous, introducing synchronized sound, full-color three-strip Technicolor, feature-length cartoons and technical developments in cameras. The results, seen in features such as *Snow White and the Seven Dwarfs* (1937), *Fantasia*, *Pinocchio* (both 1940), *Dumbo* (1941) and *Bambi* (1942), furthered the development of animated film. New animated and live-action films followed after World War II, including the critically successful *Cinderella* (1950) and *Mary Poppins* (1964), the latter of which received five Academy Awards.

In the 1950s, Disney expanded into the amusement park industry, and in 1955 he opened Disneyland. To fund the project he diversified into television programs, such as *Walt Disney's Disneyland* and *The Mickey Mouse Club*; he was also

Walt Disney

Disney in 1946


Not logged in Talk Contributions Create account Log in

Article Talk Read View source View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

WIKIPEDIA ASIAN MONTH This November is the Wikipedia Asian Month. Come join us.

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export

Walt Disney

From Wikipedia, the free encyclopedia

This article is about the man. For the company he co-founded, see The Walt Disney Company. For other uses, see Walt Disney (disambiguation).

Walter Elias Disney (/ˈdɪzni/;[1] December 5, 1901 – December 15, 1966) was an American entrepreneur, animator, voice actor and film producer. A pioneer of the American animation industry, he introduced several developments in the production of cartoons. As a film producer, Disney holds the record for most Academy Awards earned by an individual, having won 22 Oscars from 59 nominations. He was presented with two Golden Globe Special Achievement Awards and an Emmy Award, among other honors. Several of his films are included in the National Film Registry by the Library of Congress.

Born in Chicago in 1901, Disney developed an early interest in drawing. He took art classes as a boy and got a job as a commercial illustrator at the age of 18. He moved to California in the early 1920s and set up the Disney Brothers Studio with his brother Roy. With Ub Iwerks, Walt developed the character Mickey Mouse in 1928, his first highly popular success; he also provided the voice for his creation in the early years. As the studio grew, Disney became more adventurous, introducing synchronized sound, full-color three-strip Technicolor, feature-length cartoons and technical developments in cameras. The results, seen in features such as Snow White and the Seven Dwarfs (1937), Fantasia, Pinocchio (both 1940), Dumbo (1941) and Bambi (1942), furthered the development of animated film. New animated and live-action films followed after World War II, including the critically successful Cinderella (1950) and Mary Poppins (1964), the latter of which received five Academy Awards.

In the 1950s, Disney expanded into the amusement park industry, and in 1955 he opened Disneyland. To fund the project he diversified into television programs, such as Walt Disney's Disneyland and The Mickey Mouse Club; he was also

Walt Disney

Disney in 1946

CREATE(WALT DISNEY, MICKEY MOUSE)

wikipedia.com

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance learning

Not logged in Talk Contributions Create account Log in

Article Talk Read View source View history Search Wikipedia

WIKIPEDIA The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export

WIKIPEDIA ASIAN MONTH This November is the Wikipedia Asian Month. Come join us.

Walt Disney

From Wikipedia, the free encyclopedia

This article is about the man. For the company he co-founded, see The Walt Disney Company. For other uses, see Walt Disney (disambiguation).

Walter Elias Disney (/ˈdɪzni/;[1] December 5, 1901 – December 15, 1966) was an American entrepreneur, animator, voice actor and film producer. A pioneer of the American animation industry, he introduced several developments in the production of cartoons. As a film producer, Disney holds the record for most Academy Awards earned by an individual, having won 22 Oscars from 59 nominations. He was presented with two Golden Globe Special Achievement Awards and an Emmy Award, among other honors. Several of his films are included in the National Film Registry by the Library of Congress.

Born in Chicago in 1901, Disney developed an early interest in drawing. He took art classes as a boy and got a job as a commercial illustrator at the age of 18. He moved to California in the early 1920s and set up the Disney Brothers Studio with his brother Roy. With Ub Iwerks, Walt developed the character Mickey Mouse in 1928, his first highly popular success; he also provided the voice for his creation in the early years. As the studio grew, Disney became more adventurous, introducing synchronized sound, full-color three-strip Technicolor, feature-length cartoons and technical developments in cameras. The results, seen in features such as *Snow White and the Seven Dwarfs* (1937), *Fantasia*, *Pinocchio* (both 1940), *Dumbo* (1941) and *Bambi* (1942), furthered the development of animated film. New animated and live-action films followed after World War II, including the critically successful *Cinderella* (1950) and *Mary Poppins* (1964), the latter of which received five Academy Awards.

In the 1950s, Disney expanded into the amusement park industry, and in 1955 he opened Disneyland. To fund the project he diversified into television programs, such as *Walt Disney's Disneyland* and *The Mickey Mouse Club*; he was also

Walt Disney

Disney in 1946

CREATE(WALT DISNEY, MICKEY MOUSE)

Not logged in Talk Contributions Create account Log in

Article Talk Read View source View history Search Wikipedia

WIKIPEDIA The Free Encyclopedia

WIKIPEDIA ASIAN MONTH This November is the Wikipedia Asian Month. Come join us.

Walt Disney

From Wikipedia, the free encyclopedia

Walter Elias Disney (/ˈdɪzni/;[1] December 5, 1901 – December 15, 1966) was an American entrepreneur, animator, voice actor and film producer. ...

Walt Disney

Disney in 1946

CREATE(WALT DISNEY, MICKEY MOUSE)

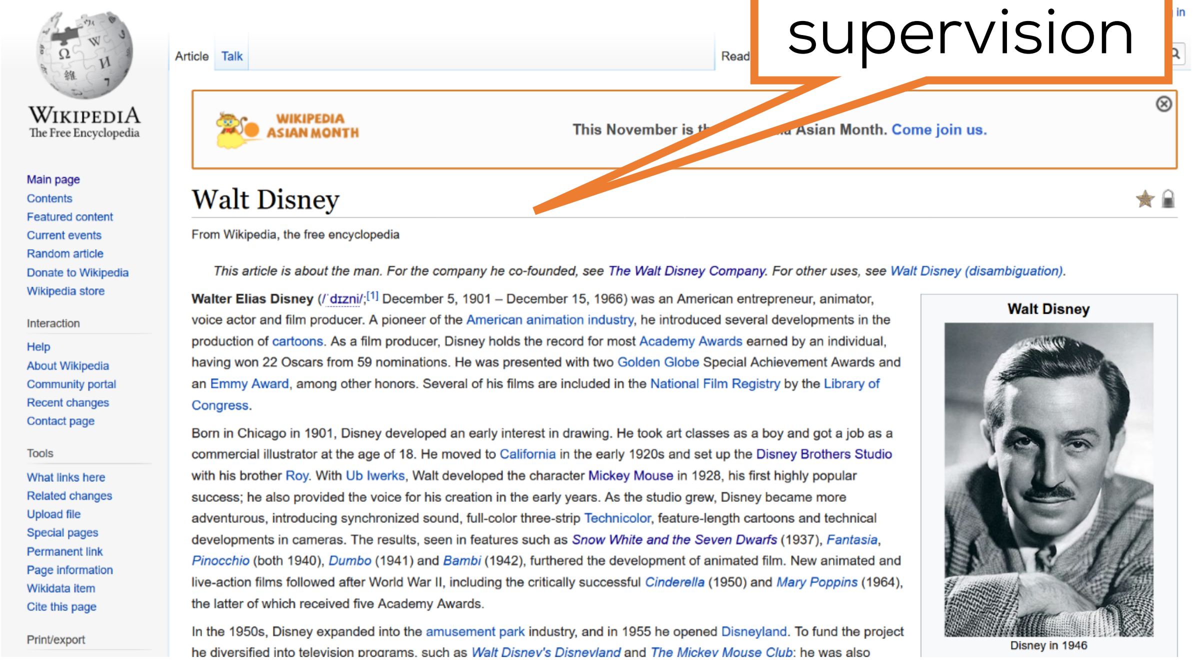
wikipedia.com

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance

Weak supervision

CREATE(WALT DISNEY, MICKEY MOUSE)

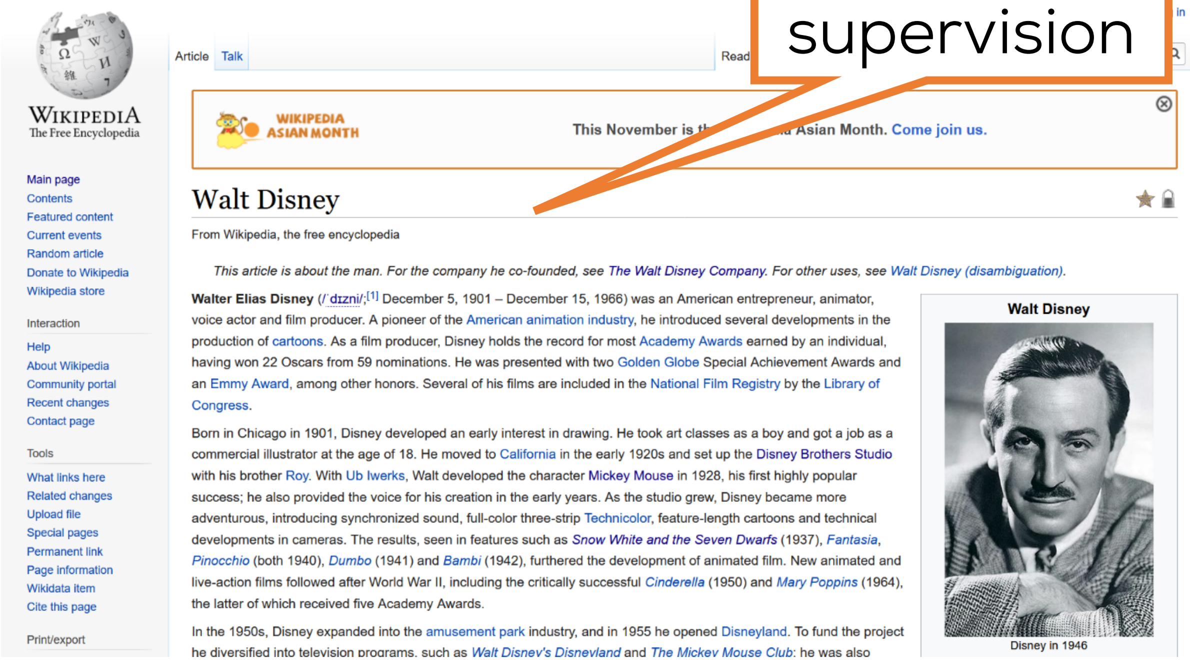


CREATE(WALT DISNEY, MICKEY MOUSE)

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance

CREATE(WALT DISNEY, MICKEY MOUSE)



Weak supervision

CREATE(WALT DISNEY, MICKEY MOUSE)

No human labeller needed if encyclopedia present

wikipedia.com

Weak Supervision

- Refers to supervision at a high level
- Related concepts – distant supervision, multi-instance

Weak supervision


Not logged in Talk Contributions Create account Log in

WIKIPEDIA The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export

Article Talk Read View source View history Search Wikipedia

WIKIPEDIA ASIAN MONTH This November is the Wikipedia Asian Month. Come join us.

Walt Disney

From Wikipedia, the free encyclopedia

This article is about the man. For the company he co-founded, see The Walt Disney Company. For other uses, see Walt Disney (disambiguation).

Walter Elias Disney (/ˈdɪzni/;[1] December 5, 1901 – December 15, 1966) was an American entrepreneur, animator, voice actor and film producer. A pioneer of the American animation industry, he introduced several developments in the production of cartoons. As a film producer, Disney holds the record for most Academy Awards earned by an individual, having won 22 Oscars from 59 nominations. He was presented with two Golden Globe Special Achievement Awards and an Emmy Award, among other honors. Several of his films are included in the National Film Registry by the Library of Congress.

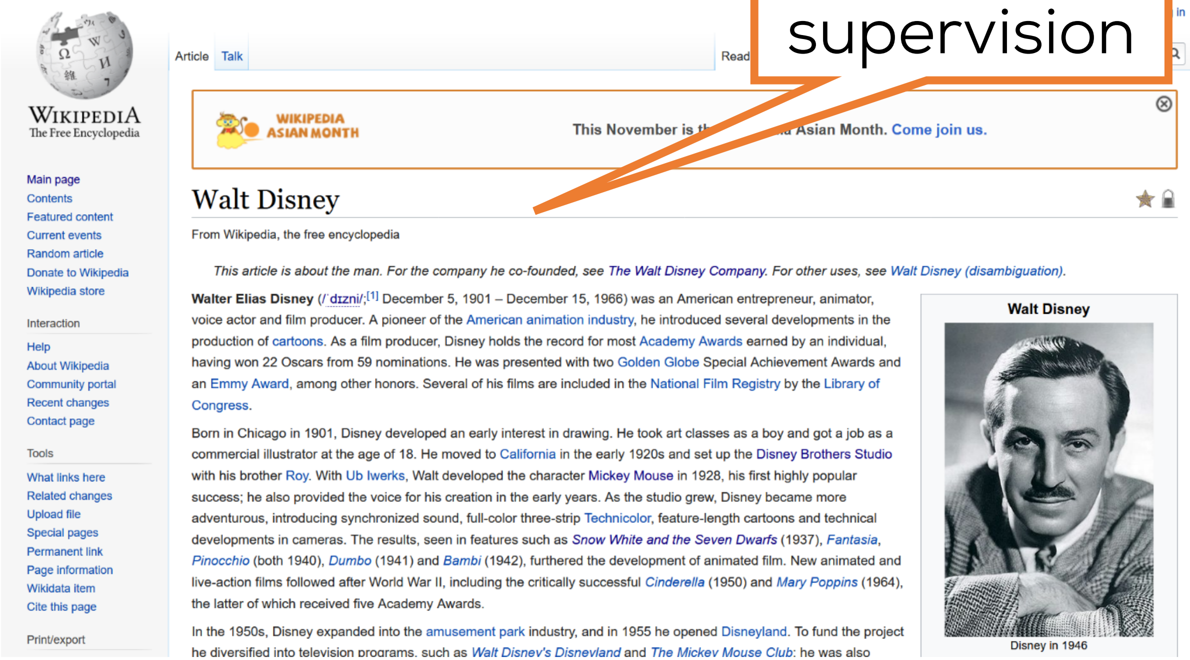
Born in Chicago in 1901, Disney developed an early interest in drawing. He took art classes as a boy and got a job as a commercial illustrator at the age of 18. He moved to California in the early 1920s and set up the Disney Brothers Studio with his brother Roy. With Ub Iwerks, Walt developed the character Mickey Mouse in 1928, his first highly popular success; he also provided the voice for his creation in the early years. As the studio grew, Disney became more adventurous, introducing synchronized sound, full-color three-strip Technicolor, feature-length cartoons and technical developments in cameras. The results, seen in features such as *Snow White and the Seven Dwarfs* (1937), *Fantasia*, *Pinocchio* (both 1940), *Dumbo* (1941) and *Bambi* (1942), furthered the development of animated film. New animated and live-action films followed after World War II, including the critically successful *Cinderella* (1950) and *Mary Poppins* (1964), the latter of which received five Academy Awards.

In the 1950s, Disney expanded into the amusement park industry, and in 1955 he opened Disneyland. To fund the project he diversified into television programs, such as *Walt Disney's Disneyland* and *The Mickey Mouse Club*; he was also

Walt Disney

Disney in 1946

CREATE(WALT DISNEY, MICKEY MOUSE)


WIKIPEDIA The Free Encyclopedia

WIKIPEDIA ASIAN MONTH This November is the Wikipedia Asian Month. Come join us.

Walt Disney

From Wikipedia, the free encyclopedia

Walter Elias Disney (/ˈdɪzni/;[1] December 5, 1901 – December 15, 1966) was an American entrepreneur, animator, voice actor and film producer. ...

Walt Disney

Disney in 1946

CREATE(WALT DISNEY, MICKEY MOUSE)

- Relation extraction problem

No human labeller needed if encyclopedia present

wikipedia.com

Weak Supervision

- Supervision at high level – distant superv., multi-instance learning
- **Multi-instance learning**: every data point is a "bag" of m "items"

$$x^i = \left\{x^{i,1}, x^{i,2}, \dots, x^{i,m}\right\}$$

  - Every possible bounding box is an item. Their collection is the image
  - Every sentence is an item. Their collection is the Wikipedia document
- The "bag" has a multi-label $\mathbf{y}^i \in \{0,1\}^L$
  - Labels in the image example could be man, horse, dog, cat, tree, river, …
  - A bit more tricky to define labels for relation extraction so skip it for now
- Not told which item(s) caused individual labels $\mathbf{y}_k^i$ to turn on or off
  - Which bounding box contains the dog?
  - Which sentence signals the relationship b/w Disney and Mickey?
- "Fine" supervision could have, for e.g., labelled every item in the bag

Solutions?

- Can often be elegantly cast as latent variable learning problems
- For every bounding box/sentence, latent vars. $z^{i,j,k}, j \in [m], k \in [L]$
- $z^{i,j,k} = +1$ if that item is of "interest" w.r.t. label k else $z^{i,j,k} = 0$
- Final label can be $\mathbf{y}_k^i = \mathrm{UNION}\left(z^{i,1,k}, z^{i,2,k}, \ldots, z^{i,m,k}\right)$
  - $\mathbf{y}_k^i = 1$ if any item declares that label to be present
  - Other "aggregation" techniques possible – ask at least 2 items to declare
- Learn models for joint inference of latent variables and final label
- Alternative is a two-stage process
  - First find item that are of interest i.e. for which $z^{i,j} \neq 0$
    - Use object detection techniques to find useful bounding boxes
    - Use entity detection techniques to find mentions
  - Label each interesting item and aggregate to get labels for bag

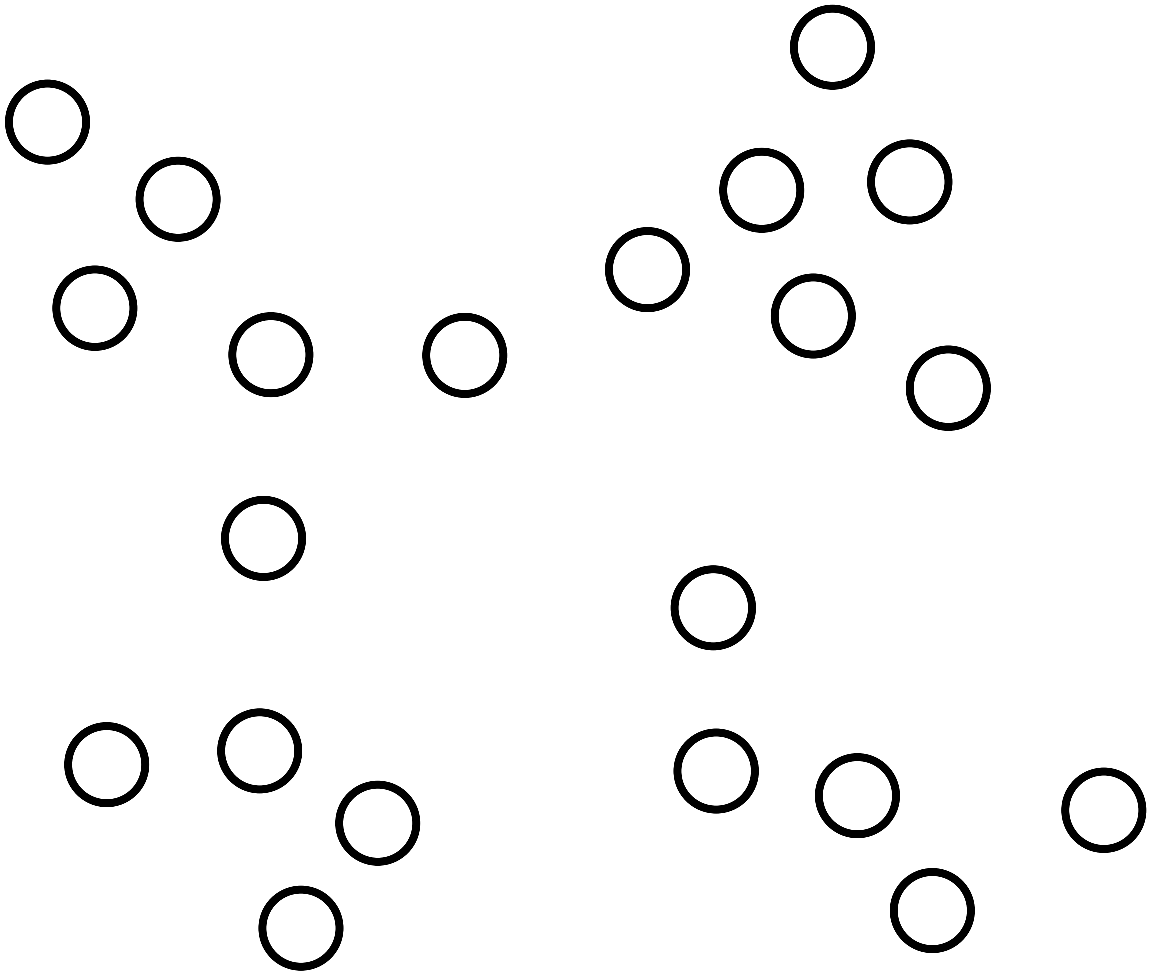
Active Learning

The active teacher

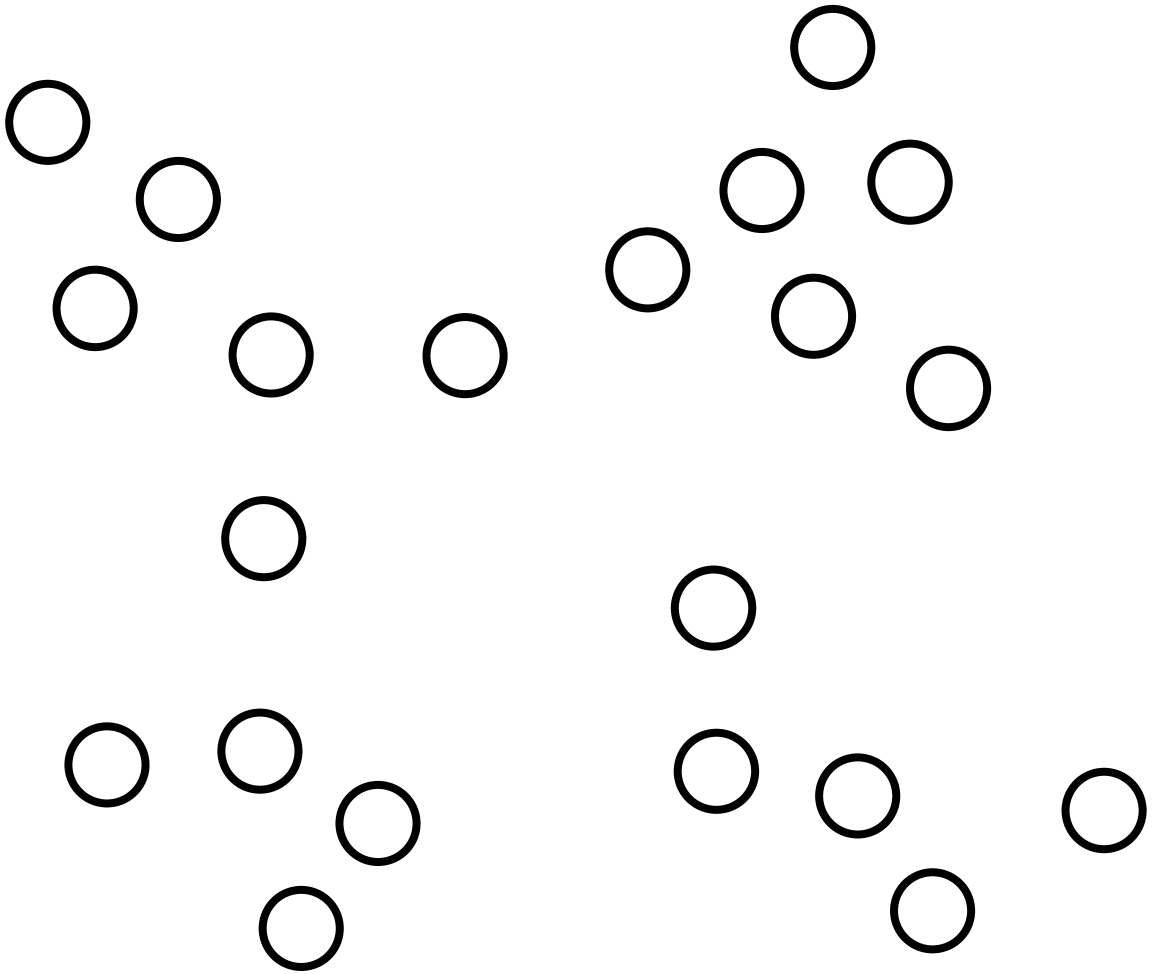
- The algorithm only gets unlabelled data points $S = \{x^1, \ldots, x^n\}$
- However, labels can be requested for $k \ll n$ points “on demand”
  - Need not ask all labels in a single go
  - Can ask for one label, process that, then ask for another label, and so on
- k must be small since most often, an actual human is involved
- We have only studied “passive” models till now where teacher retires after presenting the data
- Lots of techniques available for active learning
  - Most rely on some form of expected knowledge gain
  - Which unlabelled data point expected to surprise me the most?
  - Very “active” area (pun intended) – dedicated workshops at NIPS, ICML
  - Very useful in guiding consumer surveys - cannot have a lot of them!
  - Carefully choose which customers to entice into revealing more info

Active Learning Attempt I

Active Learning Attempt I

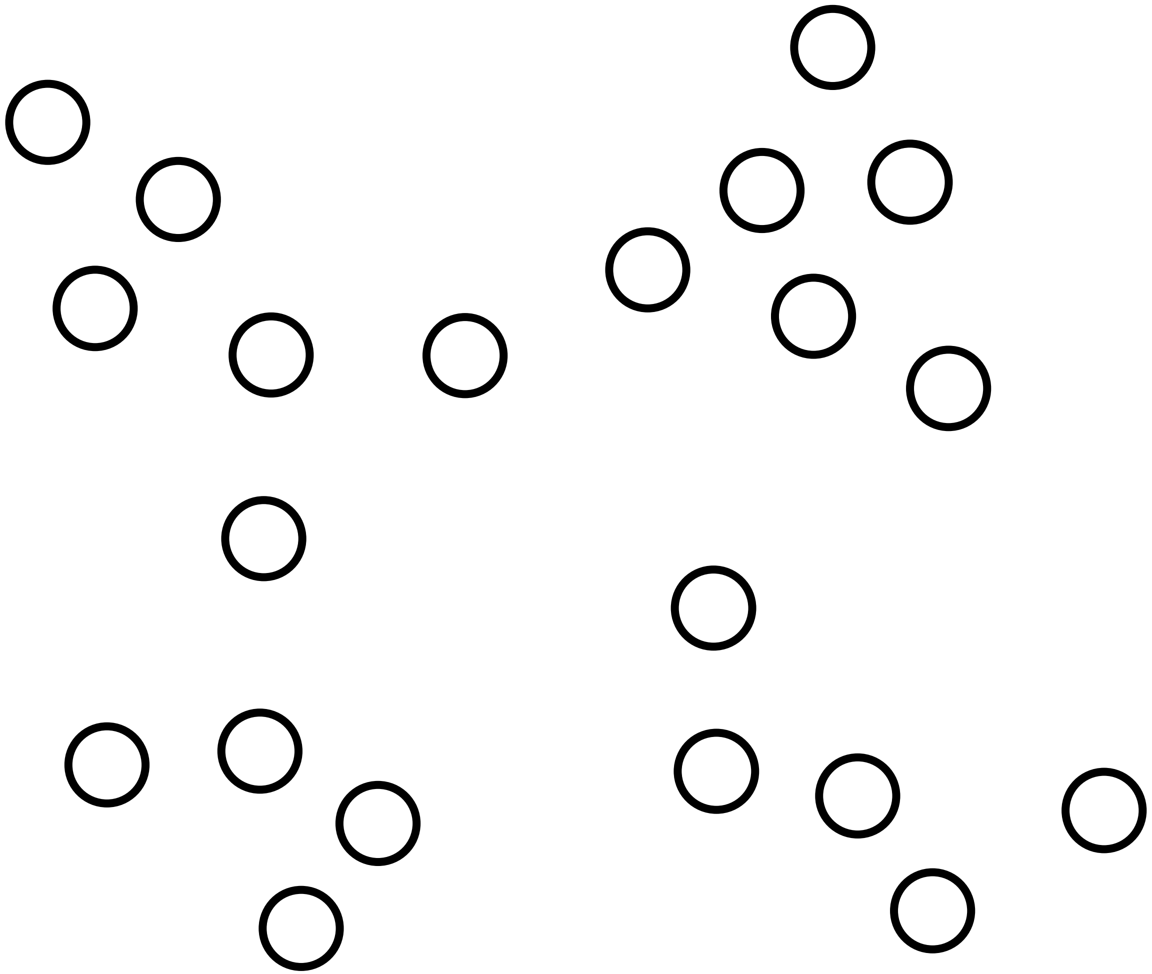


Active Learning Attempt I



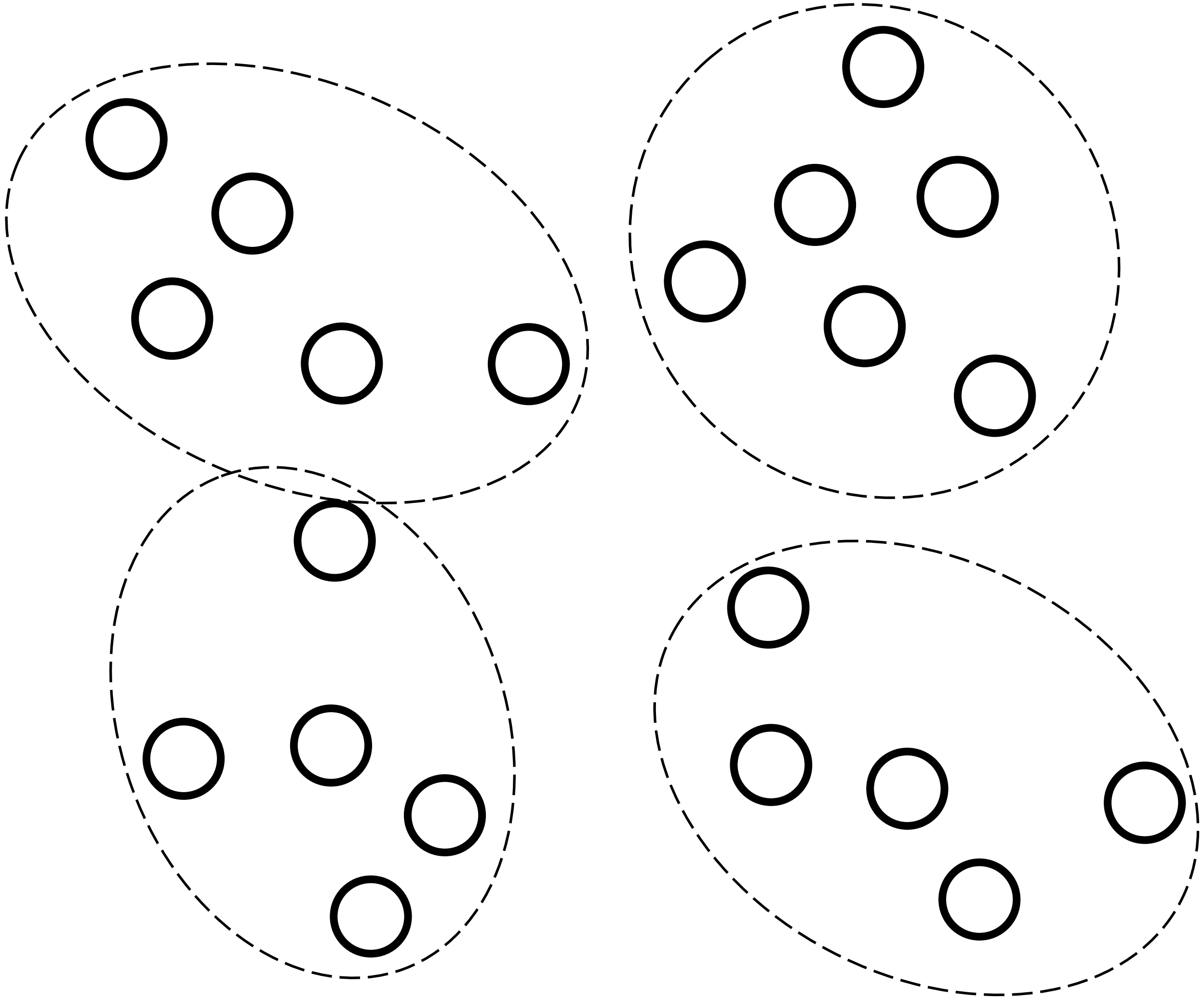
- Suppose data is has k clusters and each cluster is pure (only red or only green)

Active Learning Attempt I



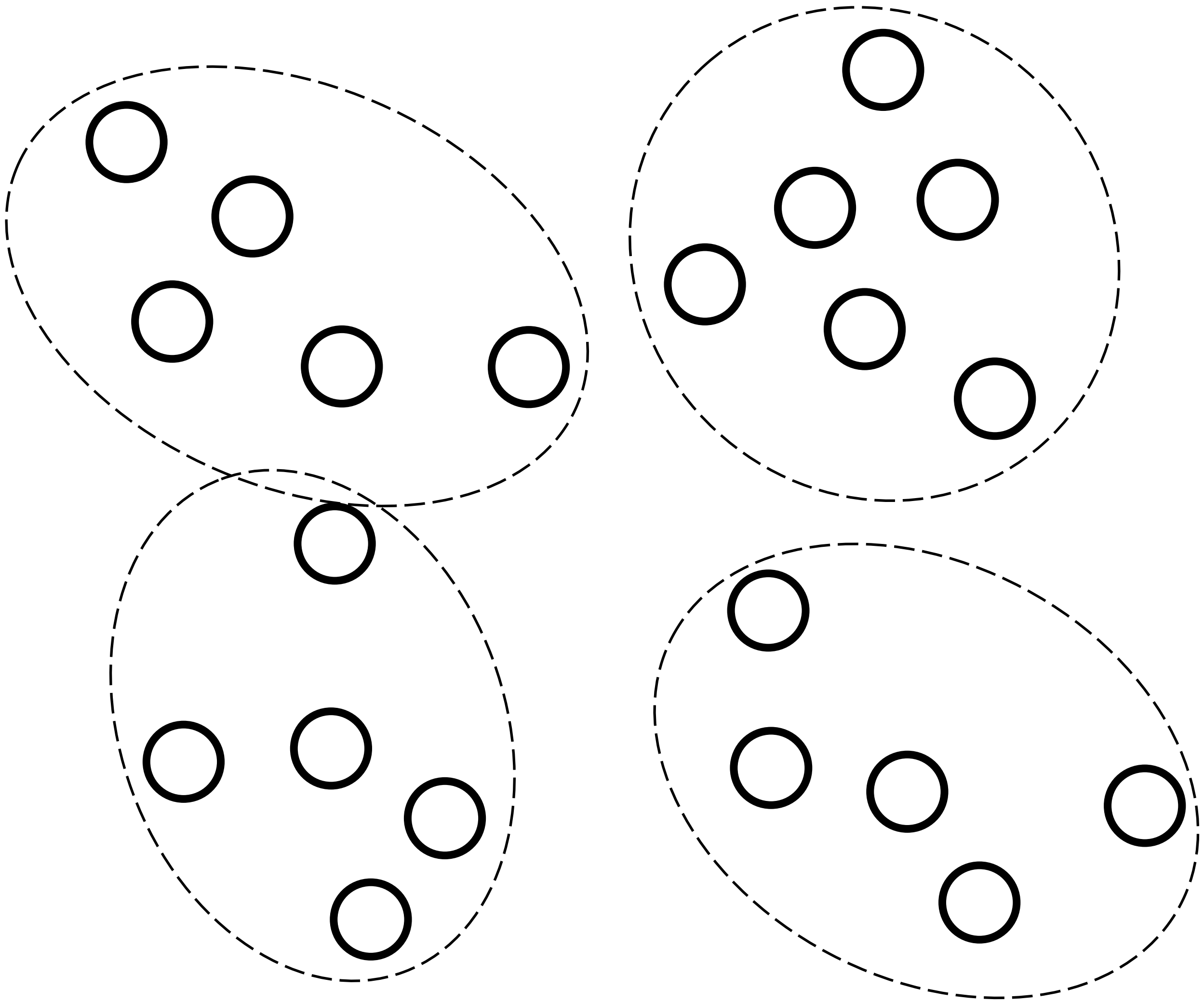
- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)

Active Learning Attempt I



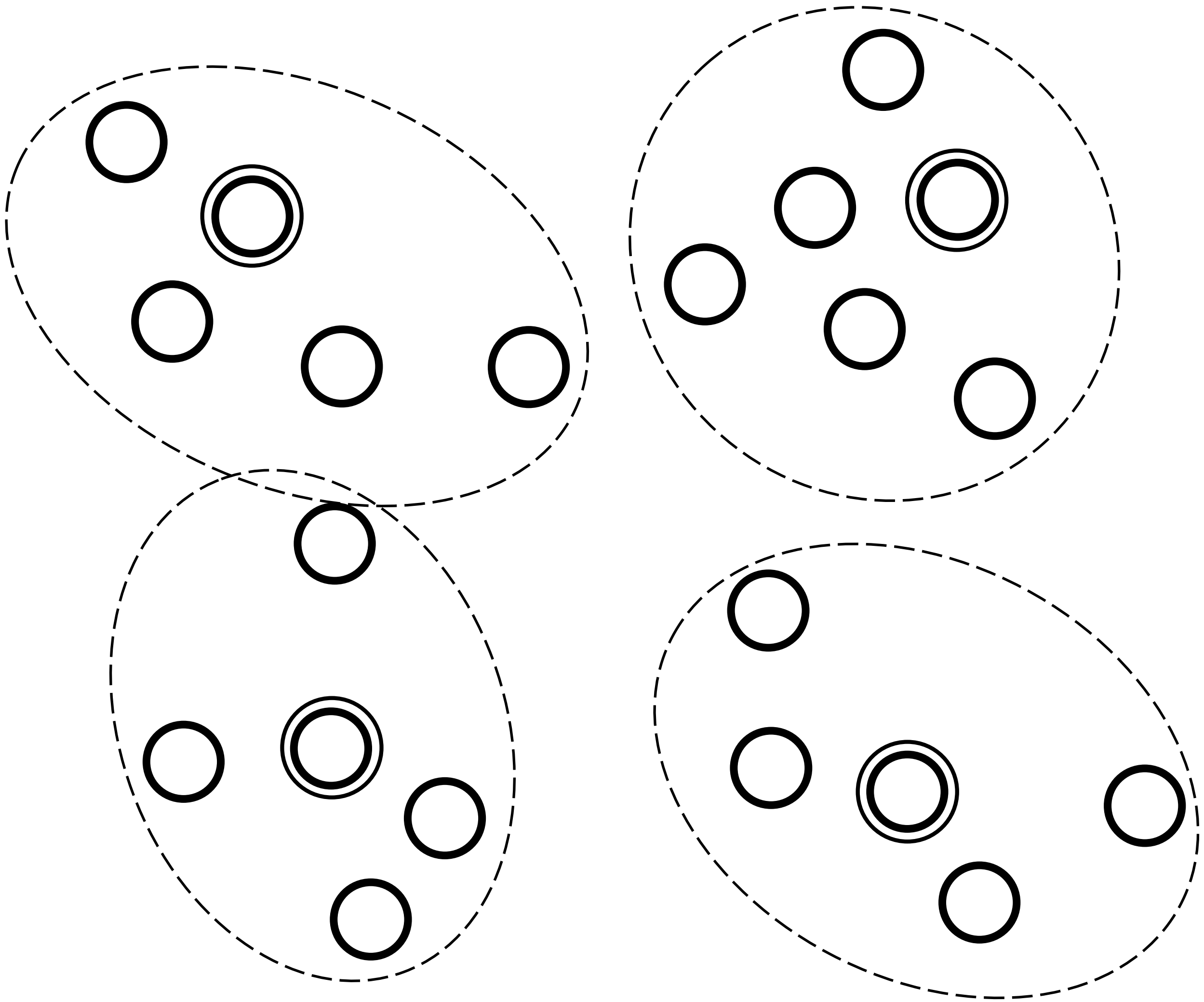
- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)

Active Learning Attempt I



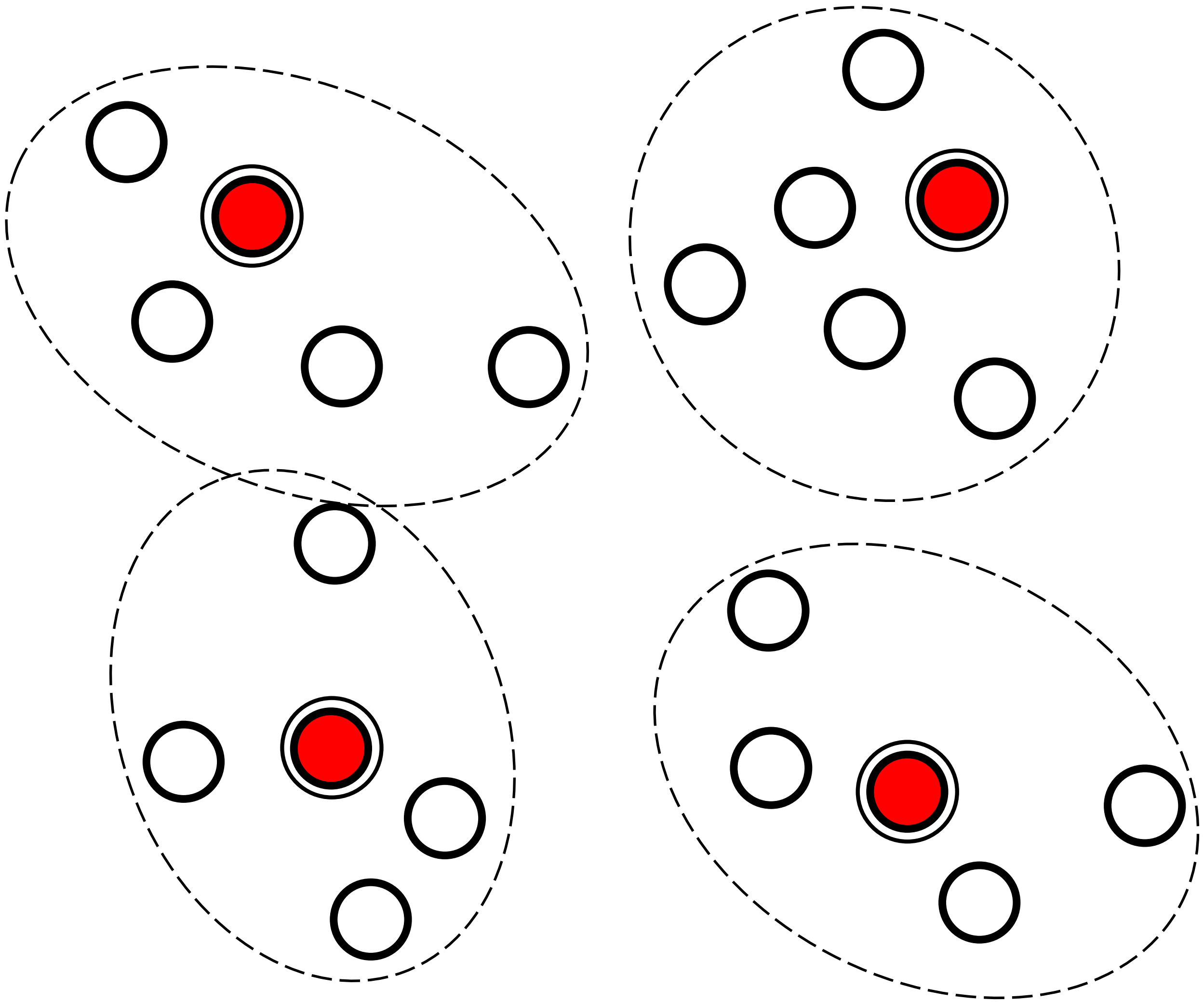
- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)
- Choose one data point from each cluster and query its label

Active Learning Attempt I



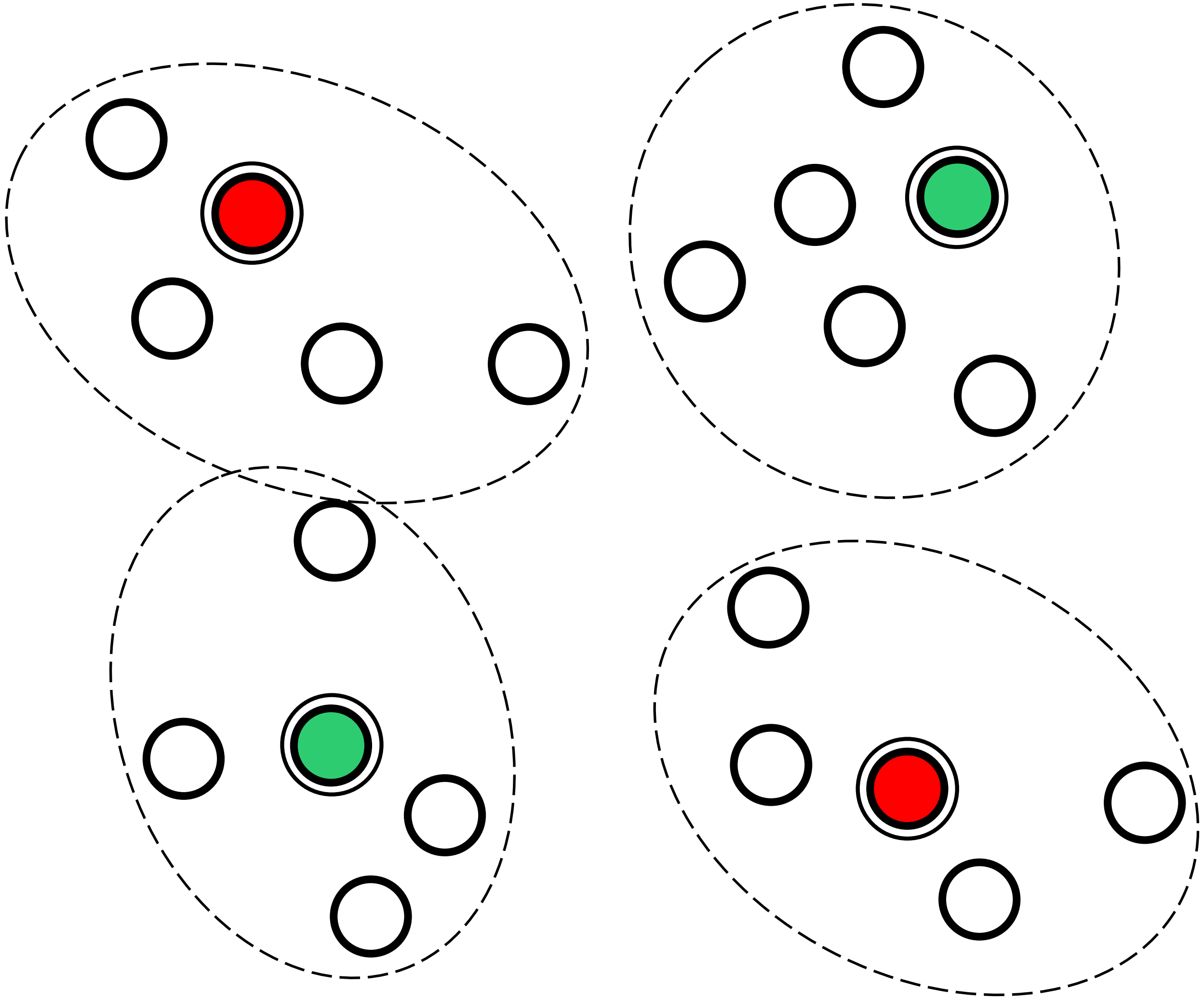
- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)
- Choose one data point from each cluster and query its label

Active Learning Attempt I



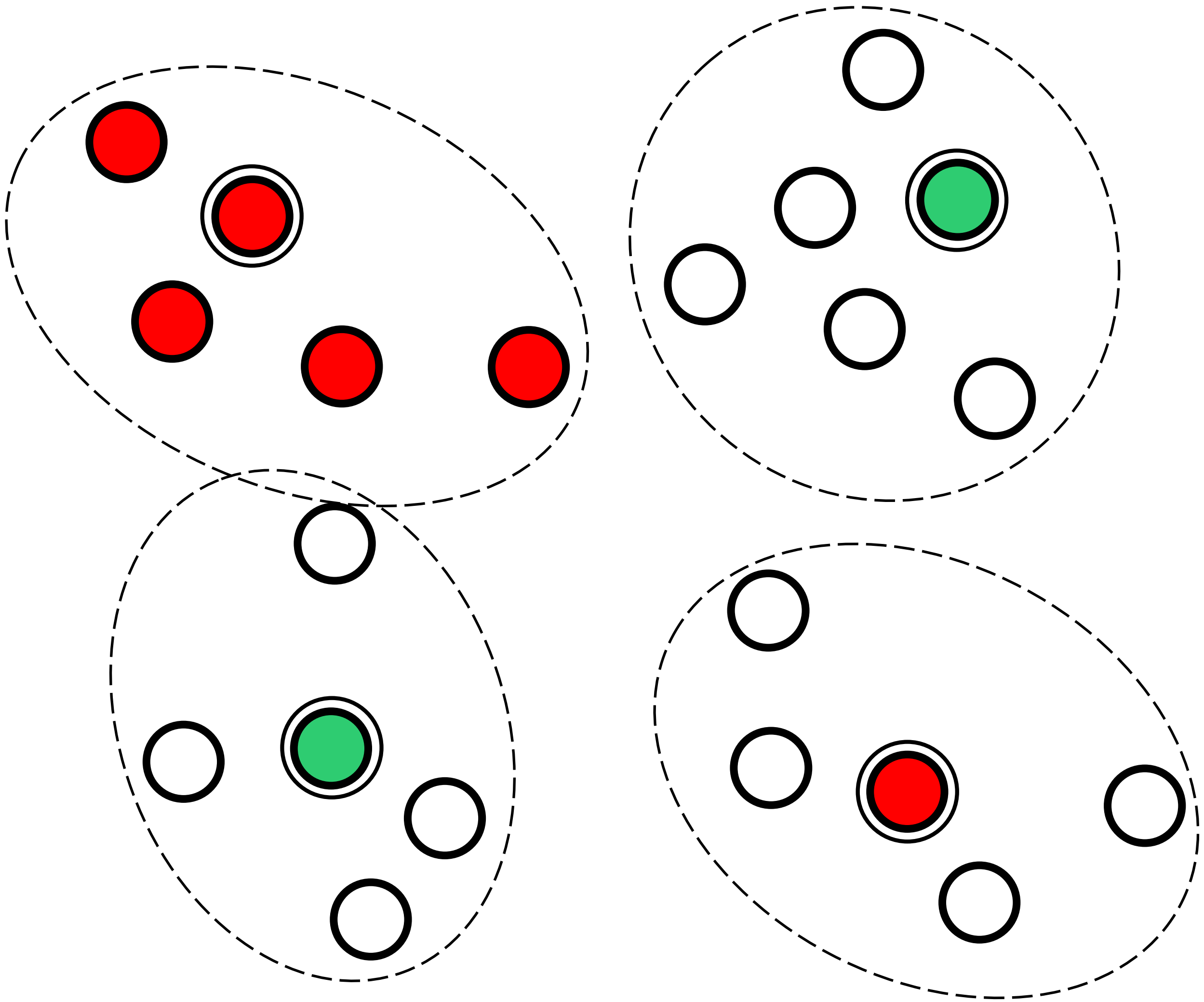
- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)
- Choose one data point from each cluster and query its label

Active Learning Attempt I



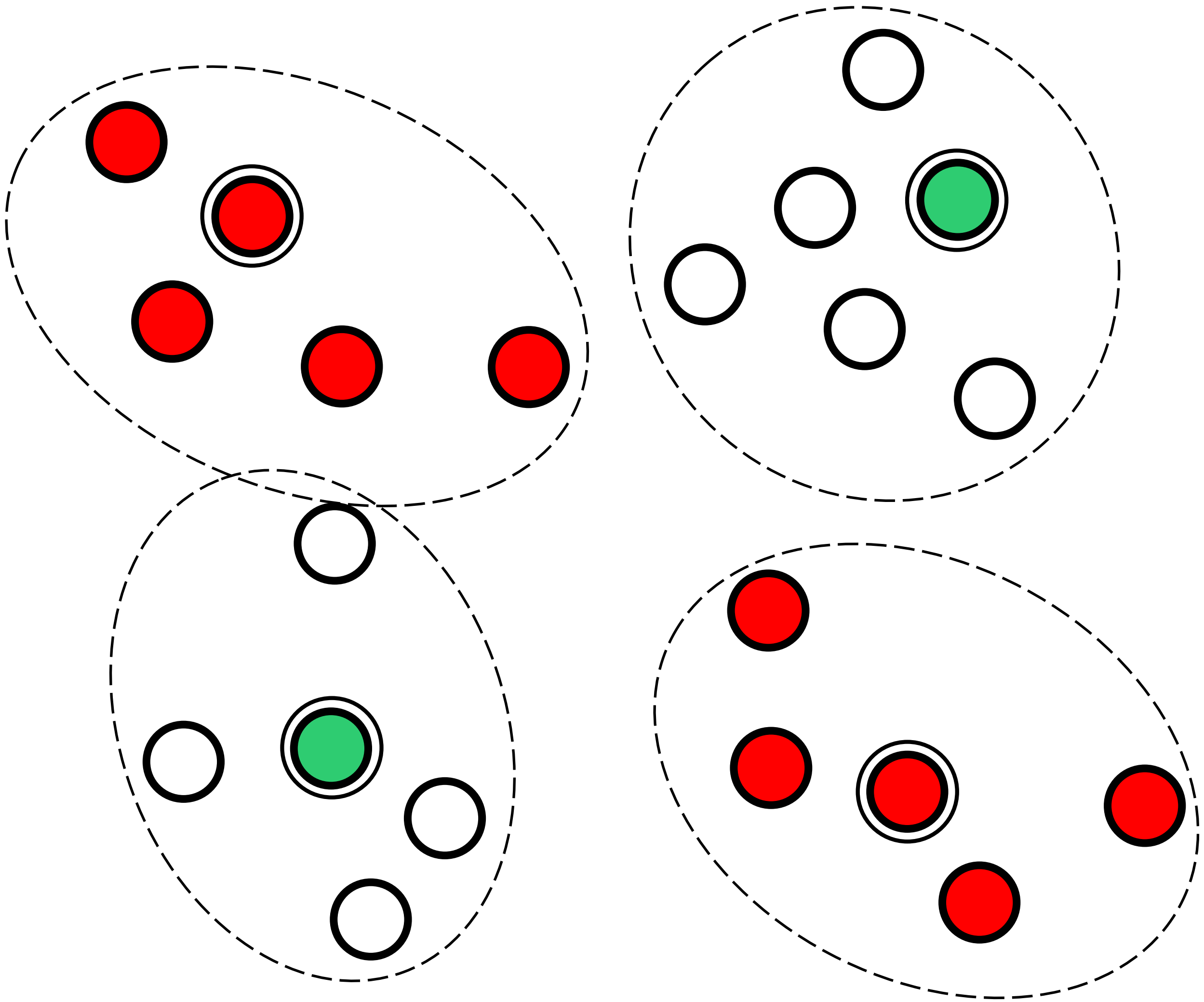
- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)
- Choose one data point from each cluster and query its label
- Label the cluster!

Active Learning Attempt I



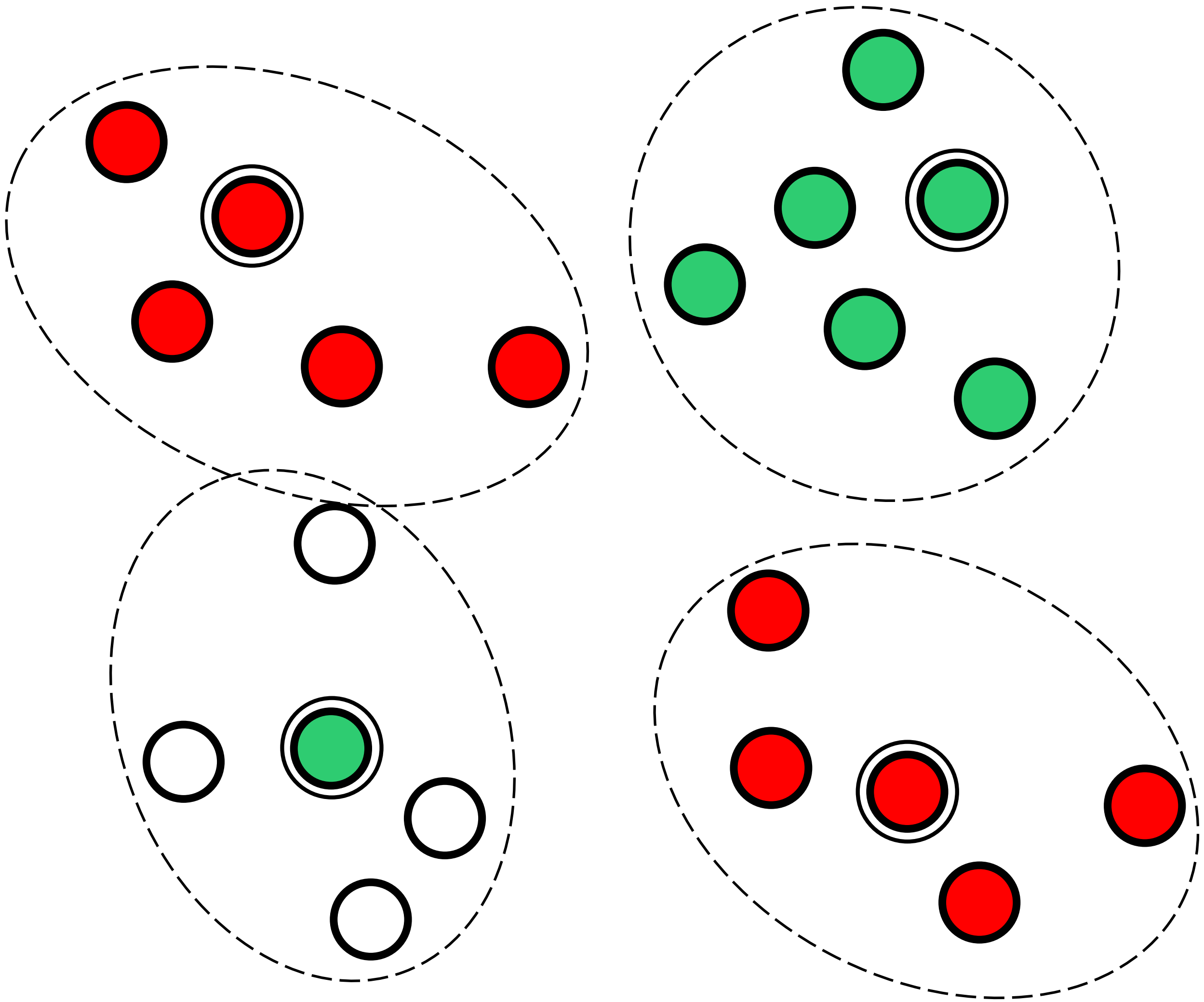
- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)
- Choose one data point from each cluster and query its label
- Label the cluster!

Active Learning Attempt I



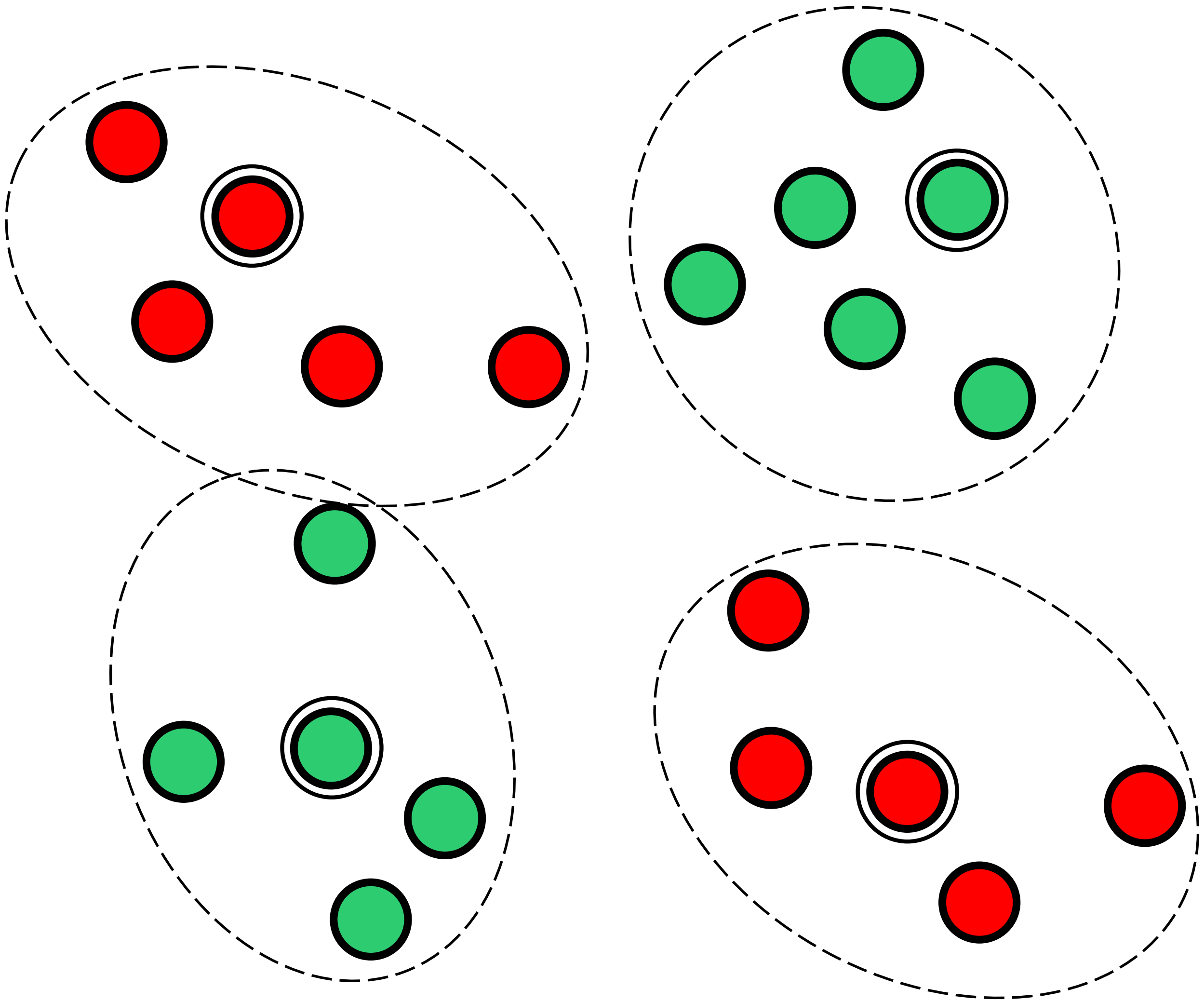
- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)
- Choose one data point from each cluster and query its label
- Label the cluster!

Active Learning Attempt I



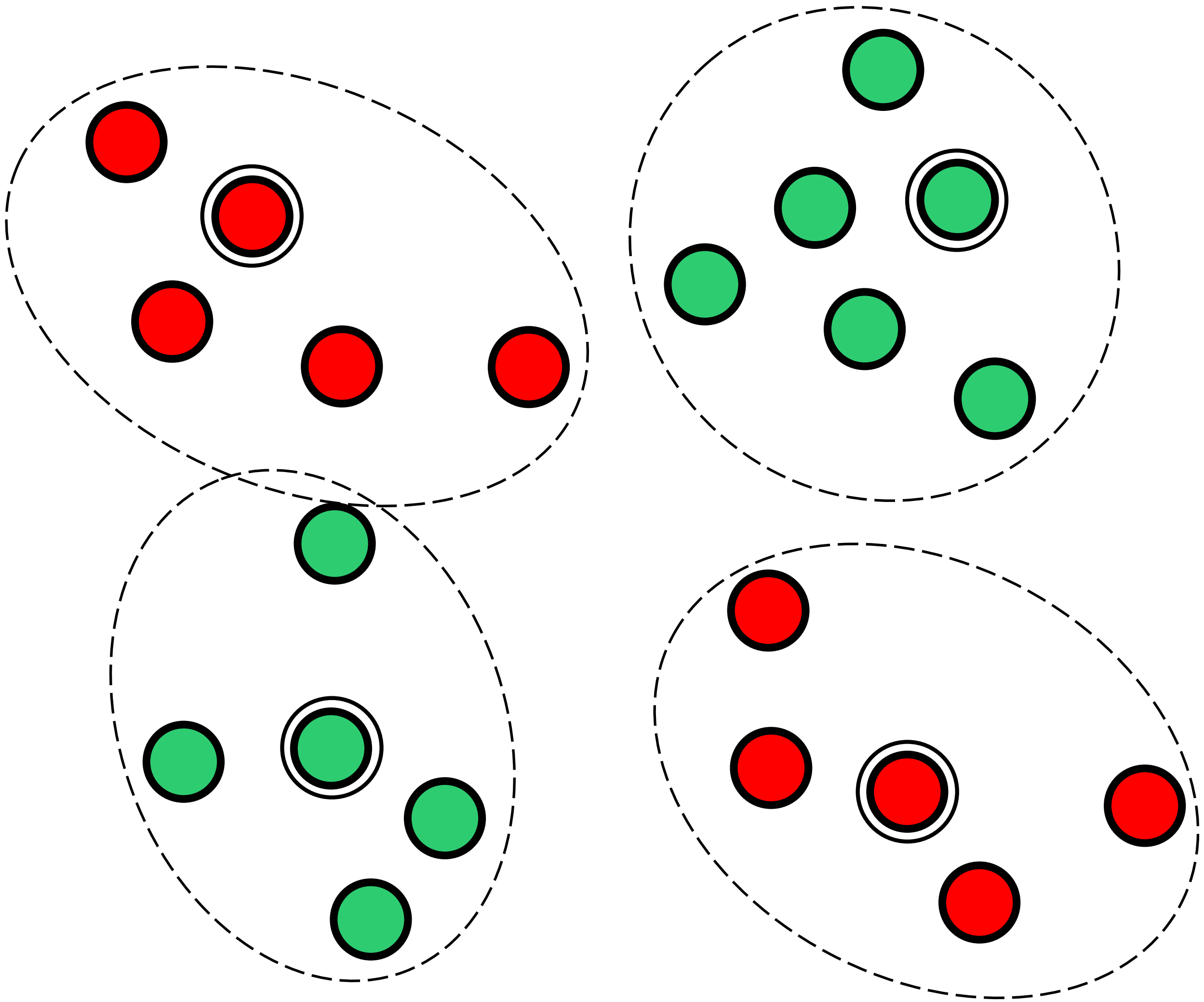
- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)
- Choose one data point from each cluster and query its label
- Label the cluster!

Active Learning Attempt I



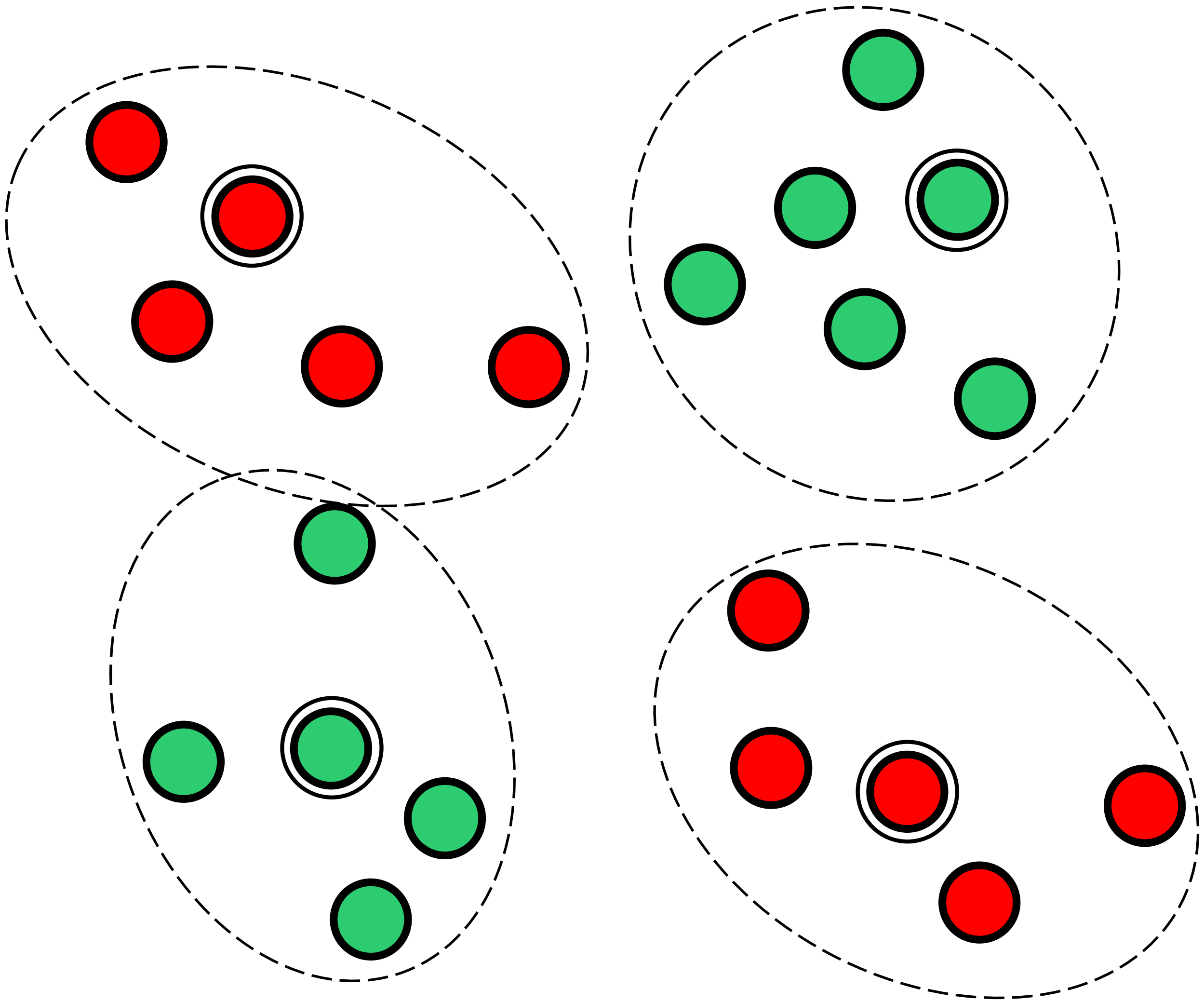
- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)
- Choose one data point from each cluster and query its label
- Label the cluster!

Active Learning Attempt I



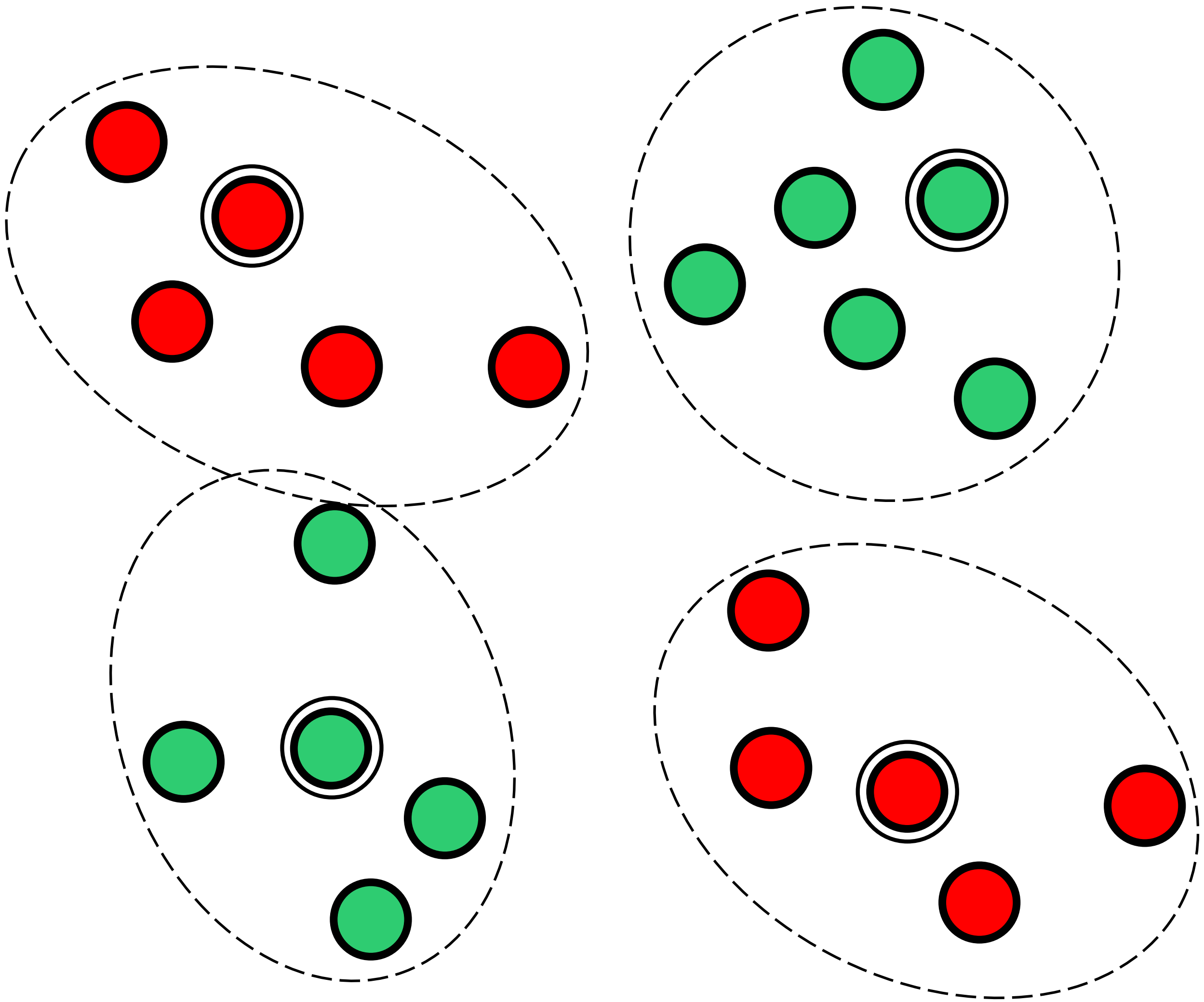
- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)
- Choose one data point from each cluster and query its label
- Label the cluster!
- Many challenges

Active Learning Attempt I



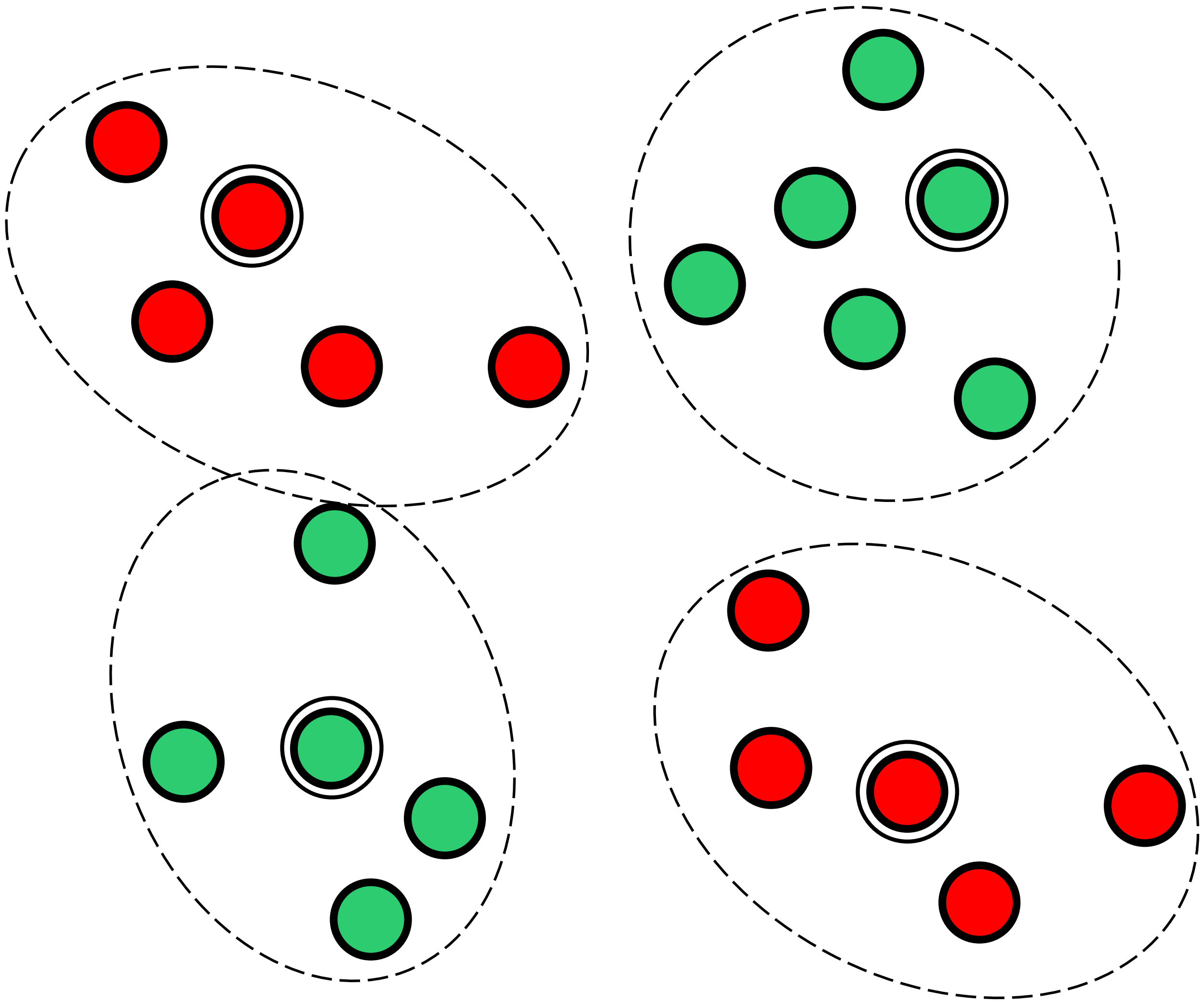
- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)
- Choose one data point from each cluster and query its label
- Label the cluster!
- Many challenges
  - Clusters not pure in general

Active Learning Attempt I



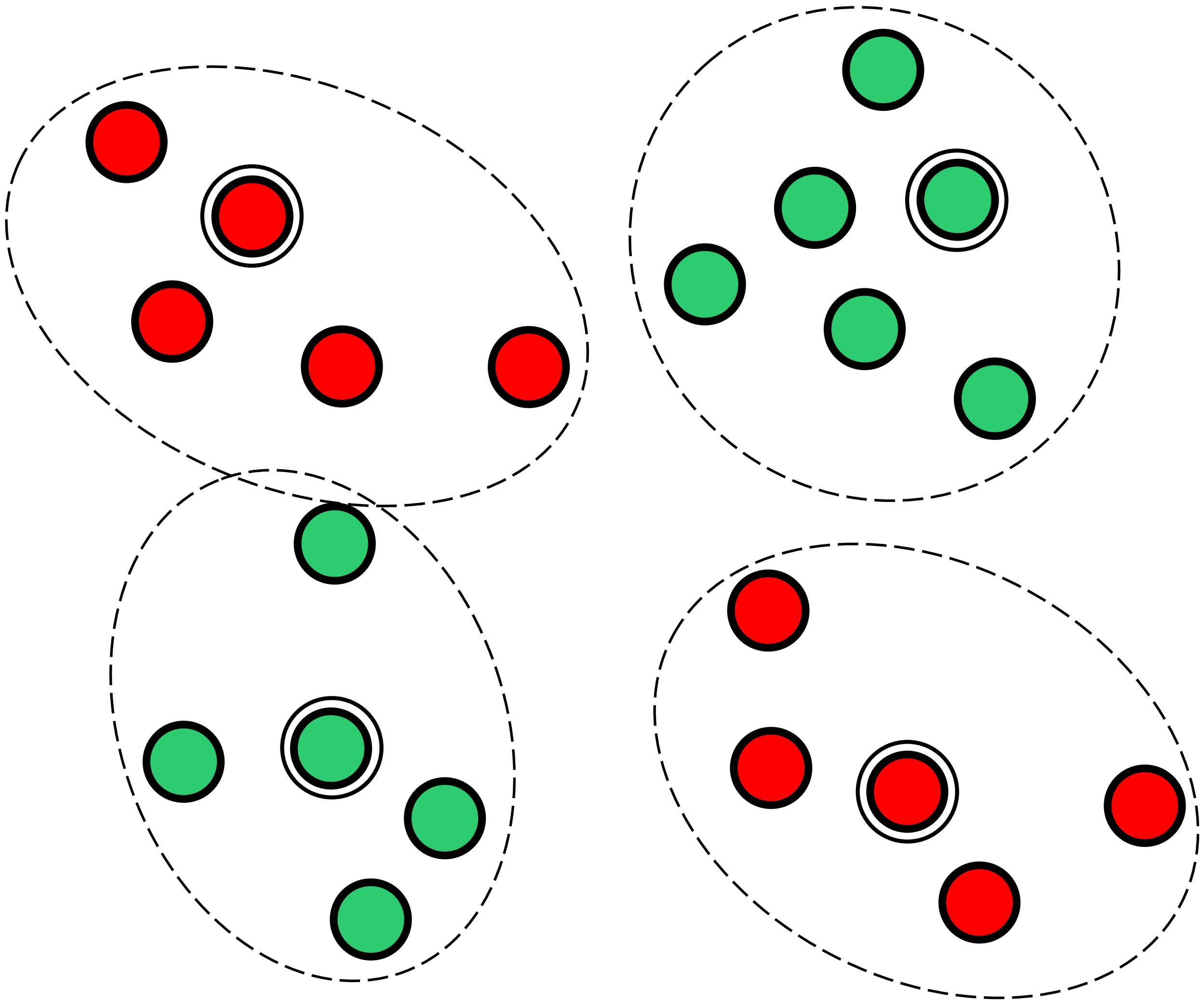
- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)
- Choose one data point from each cluster and query its label
- Label the cluster!
- Many challenges
  - Clusters not pure in general
  - Clustering is itself a tricky problem

Active Learning Attempt I



- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)
- Choose one data point from each cluster and query its label
- Label the cluster!
- Many challenges
  - Clusters not pure in general
  - Clustering is itself a tricky problem
  - Clustering mistakes disastrous

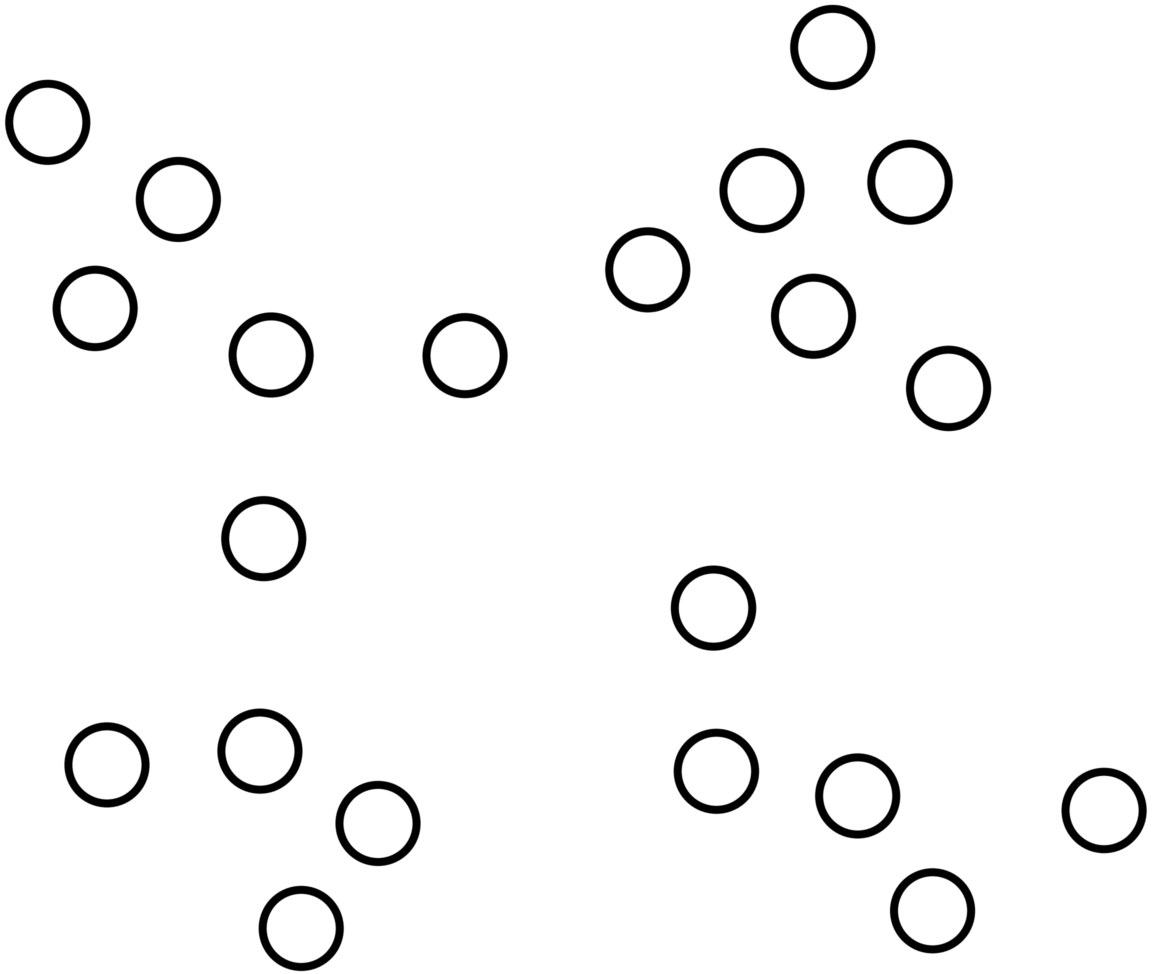
Active Learning Attempt I



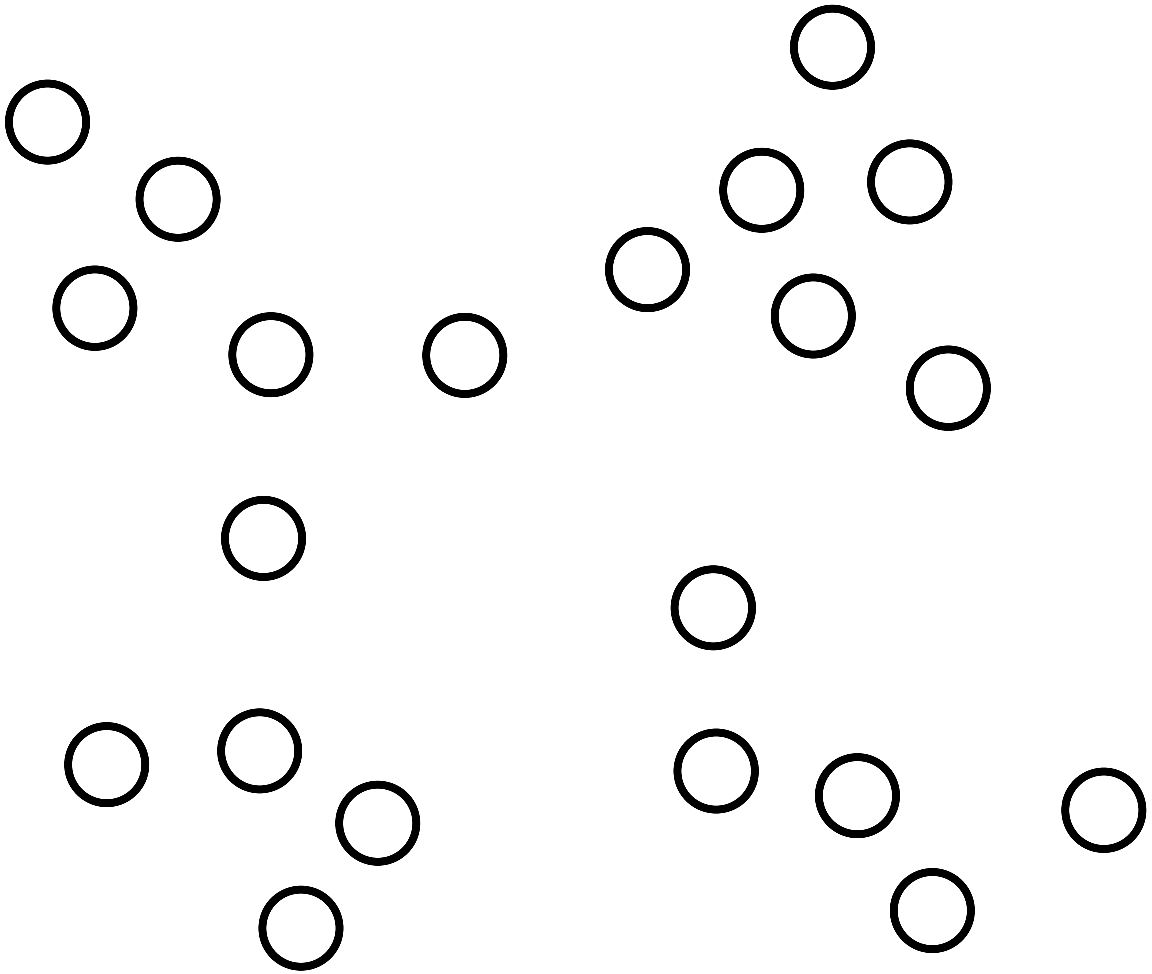
- Suppose data is has k clusters and each cluster is pure (only red or only green)
- Apply clustering (k-means, agglomerative, kernel k-means)
- Choose one data point from each cluster and query its label
- Label the cluster!
- Many challenges
  - Clusters not pure in general
  - Clustering is itself a tricky problem
  - Clustering mistakes disastrous
- Many solutions – discuss one

Active Learning Attempt I

Active Learning Attempt I

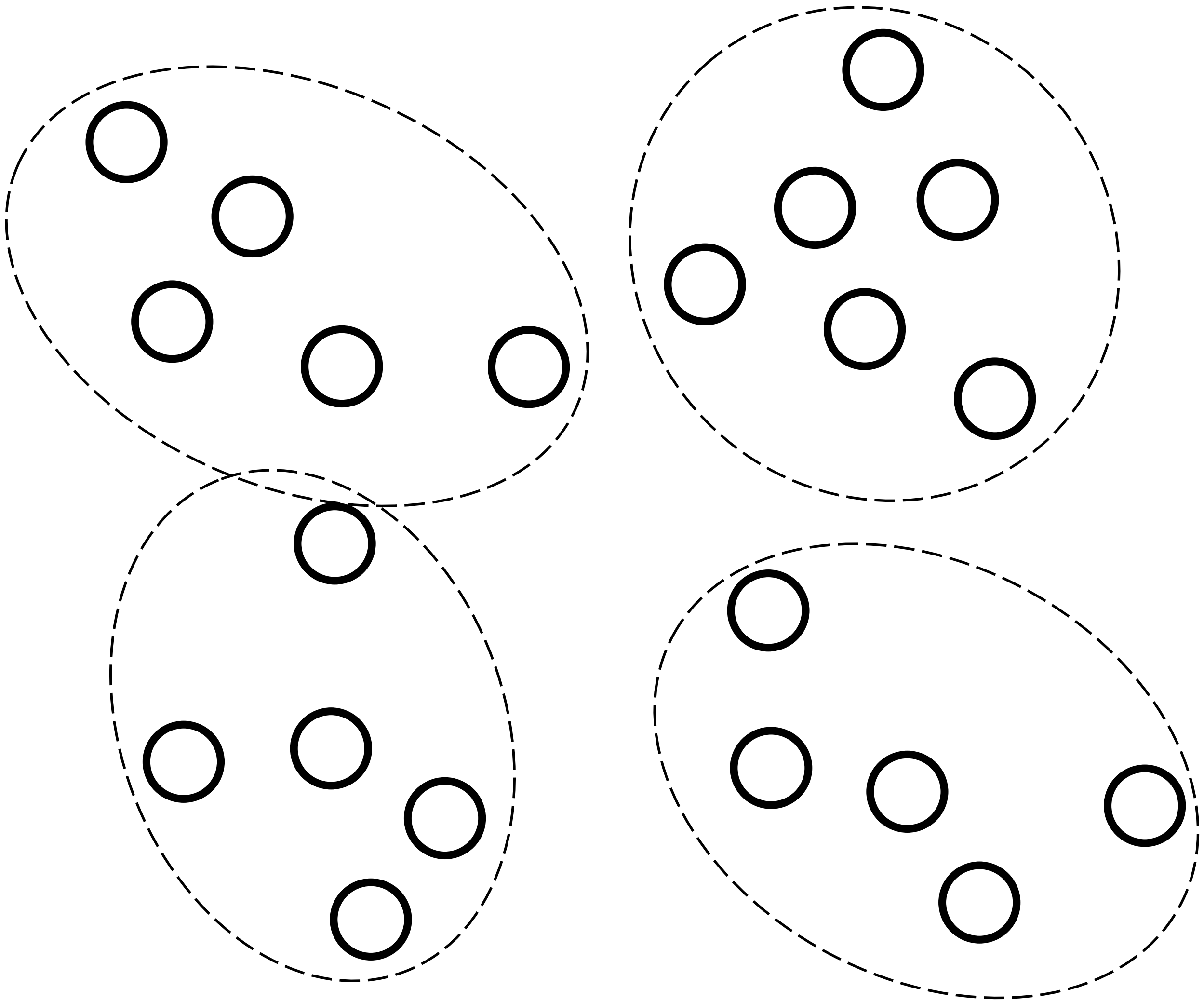


Active Learning Attempt I



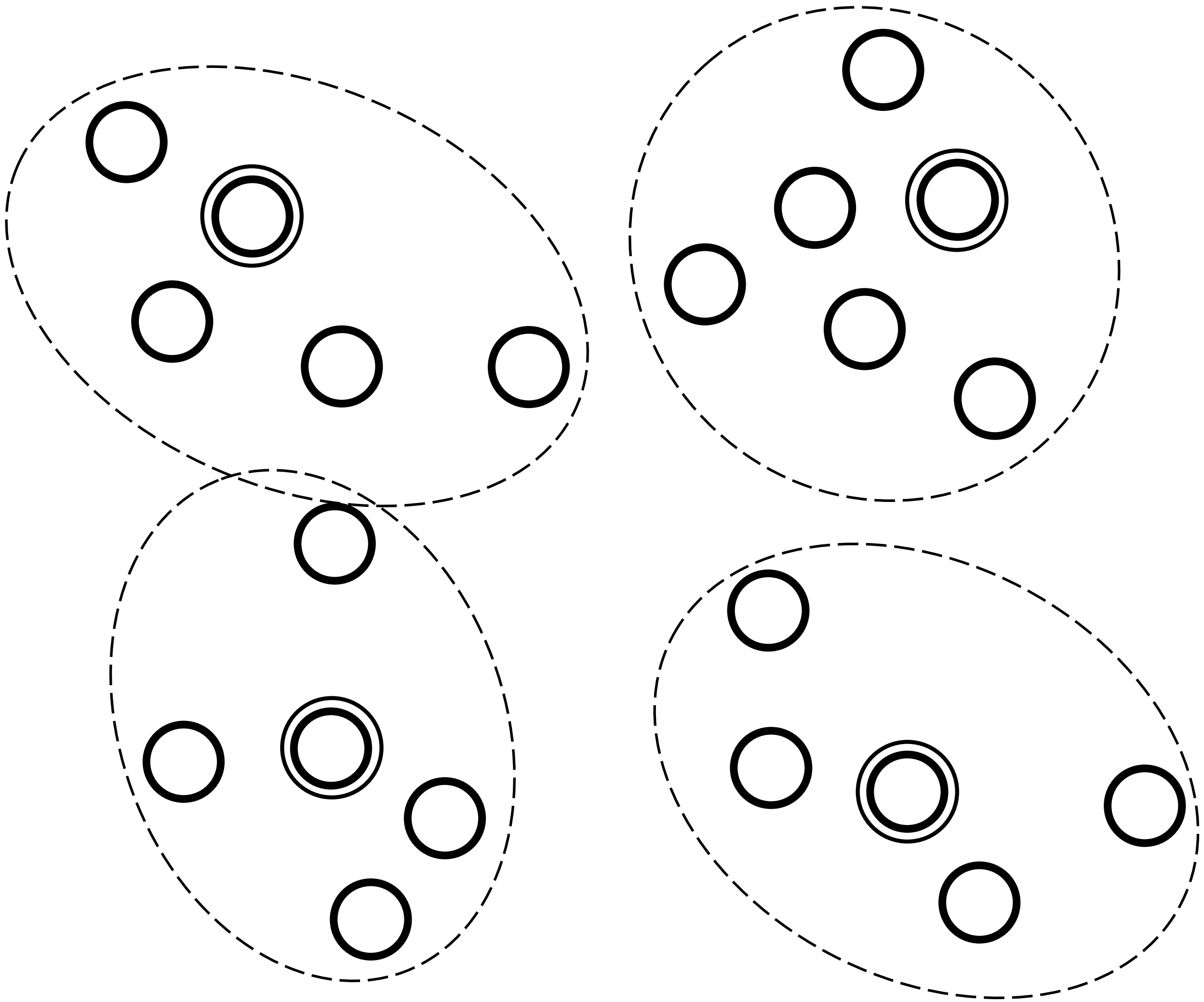
- Cluster and get labels for cluster representatives

Active Learning Attempt I



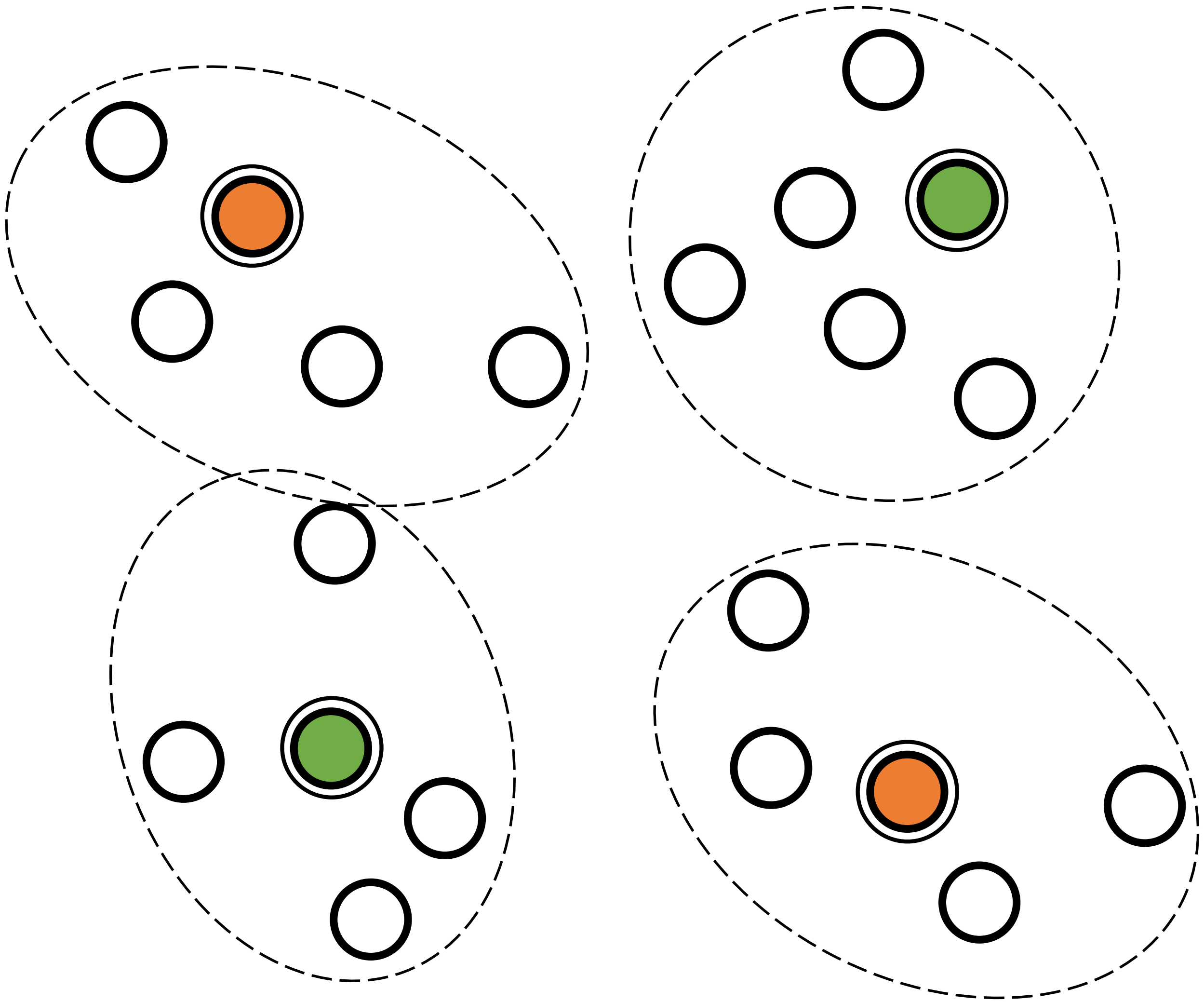
- Cluster and get labels for cluster representatives

Active Learning Attempt I



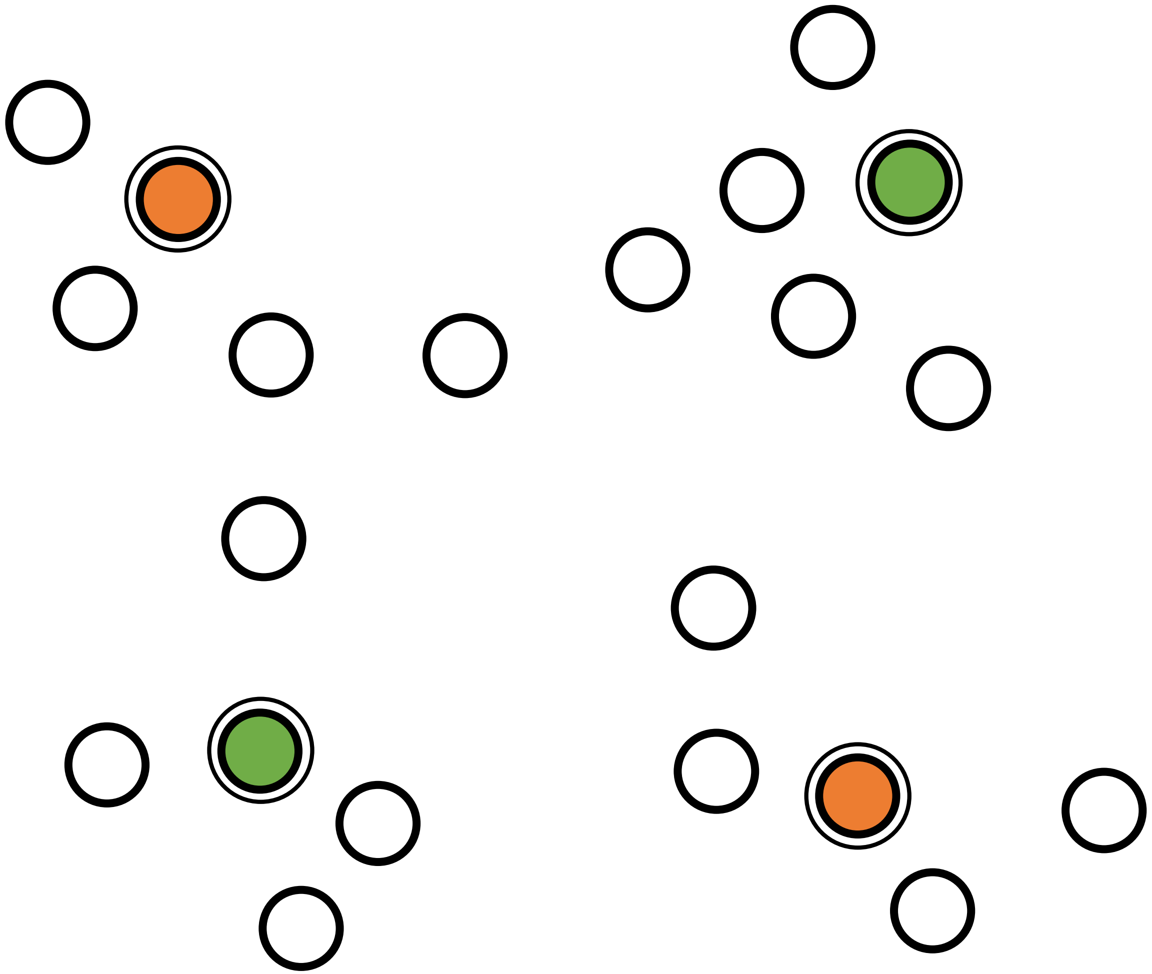
- Cluster and get labels for cluster representatives

Active Learning Attempt I



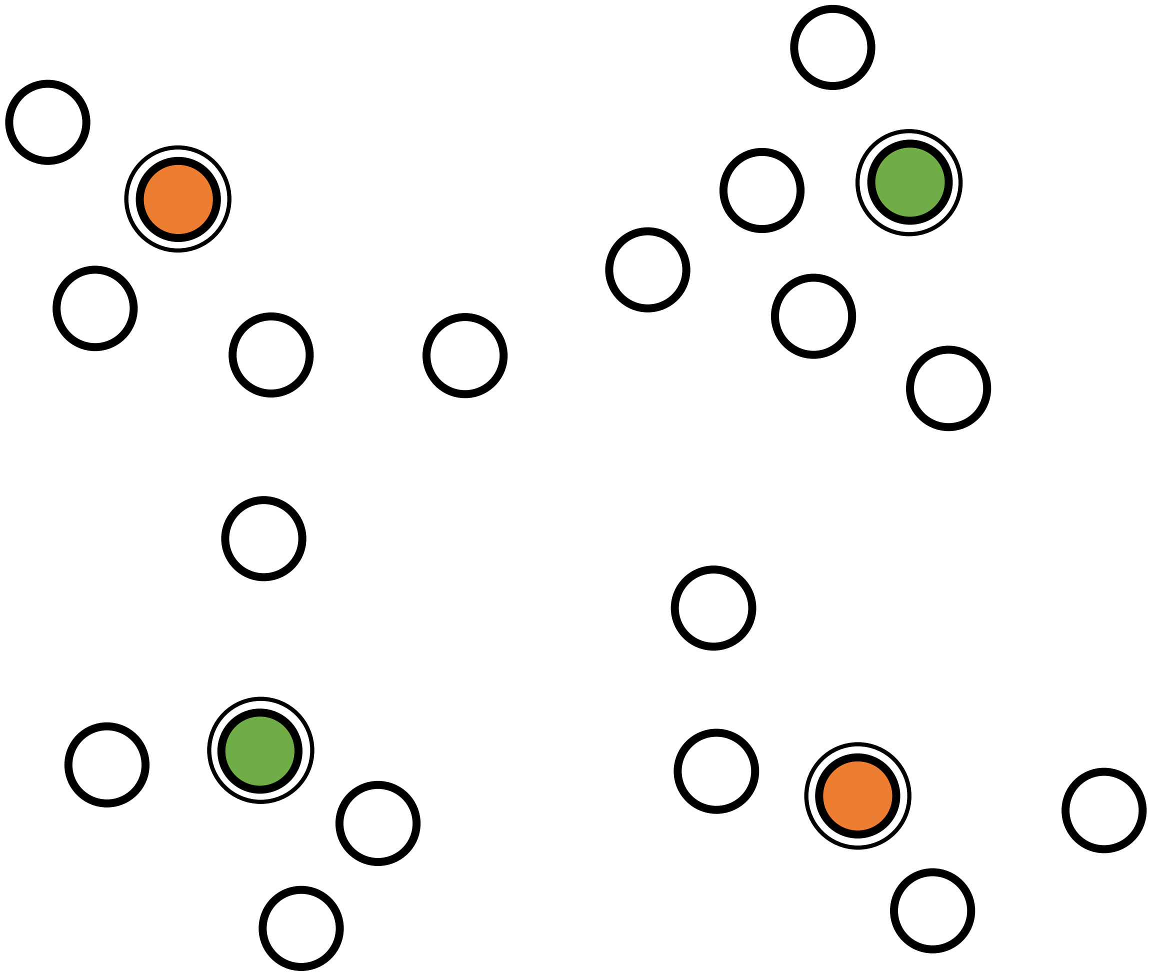
- Cluster and get labels for cluster representatives

Active Learning Attempt I



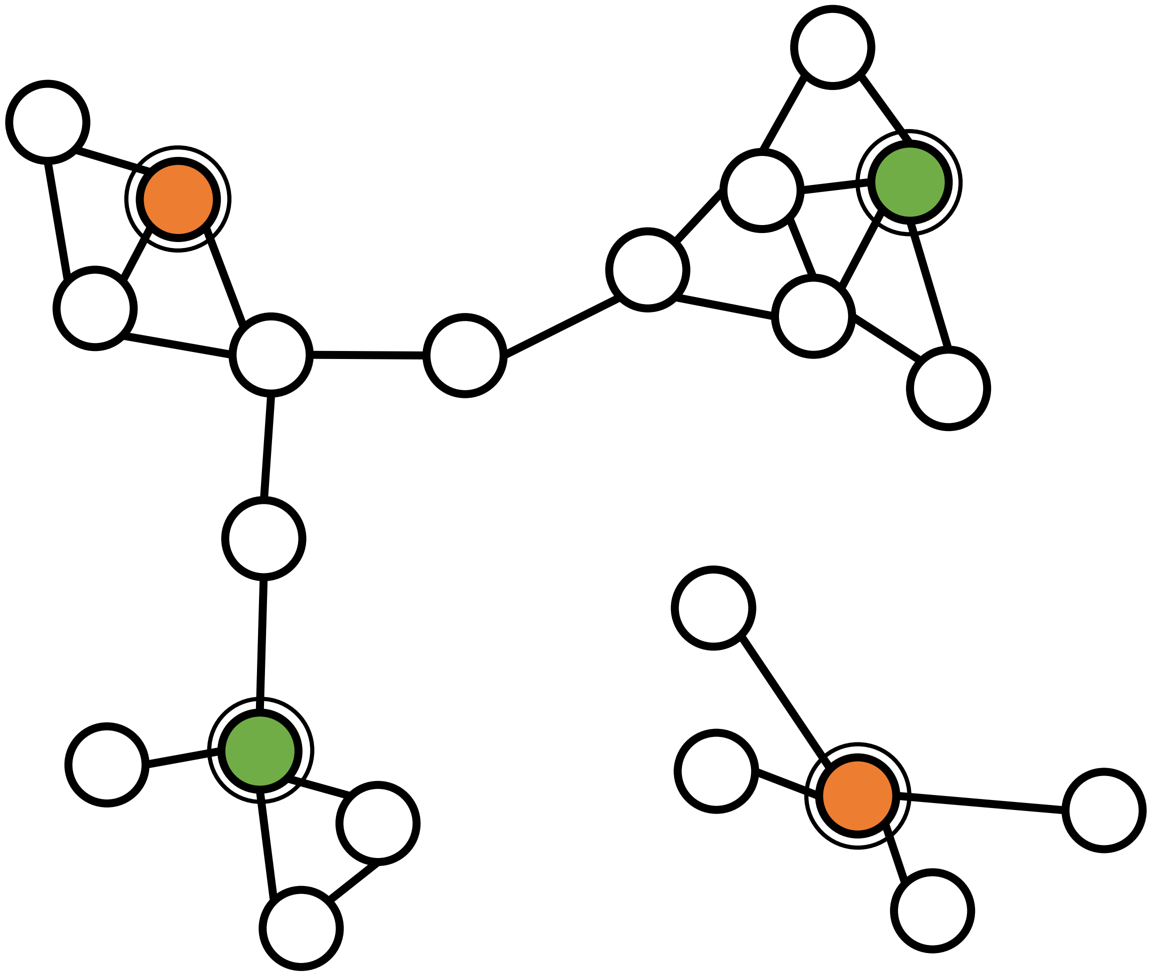
- Cluster and get labels for cluster representatives

Active Learning Attempt I



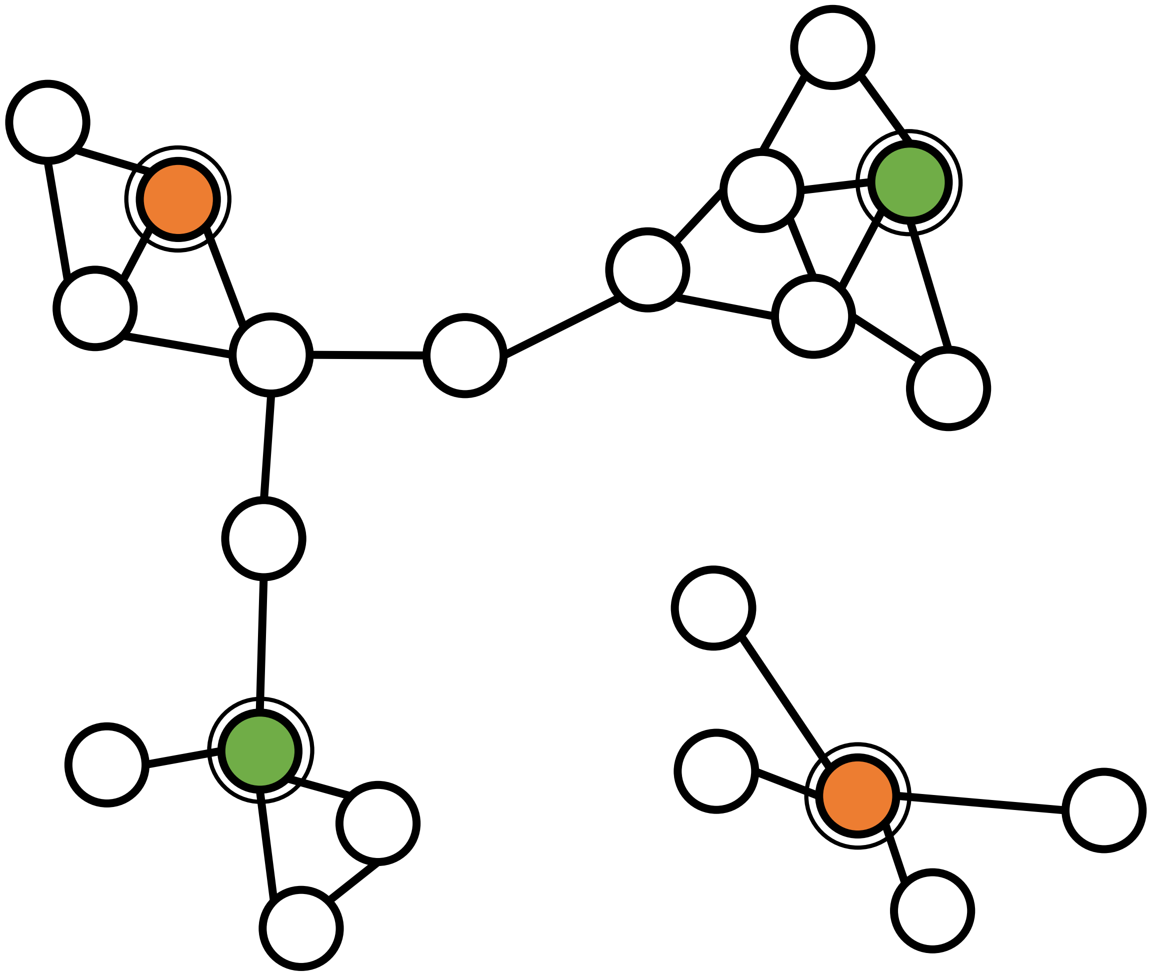
- Cluster and get labels for cluster representatives
- Create a "similarity graph" over data points

Active Learning Attempt I



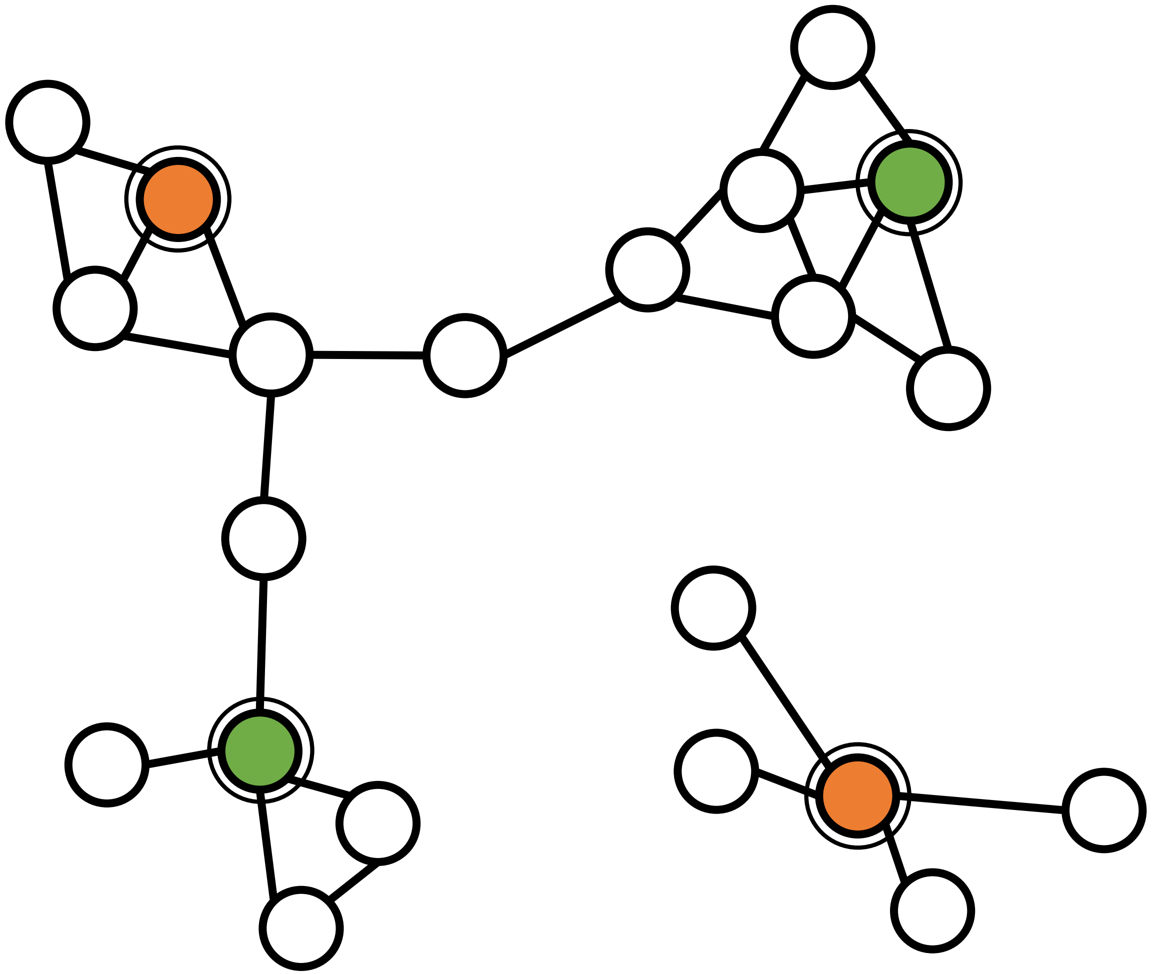
- Cluster and get labels for cluster representatives
- Create a "similarity graph" over data points

Active Learning Attempt I



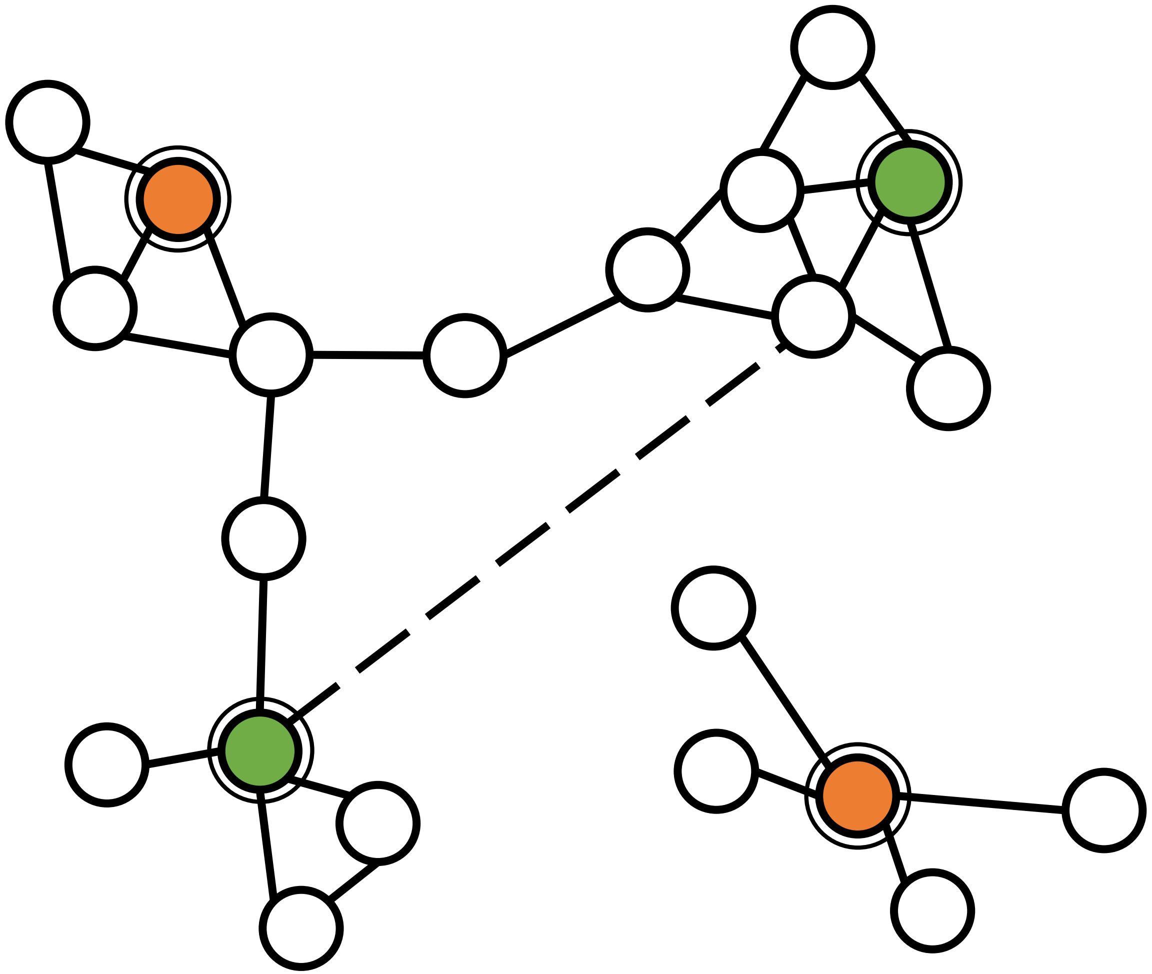
- Cluster and get labels for cluster representatives
- Create a “similarity graph” over data points
  - May use a Mercer kernel for edge weights

Active Learning Attempt I



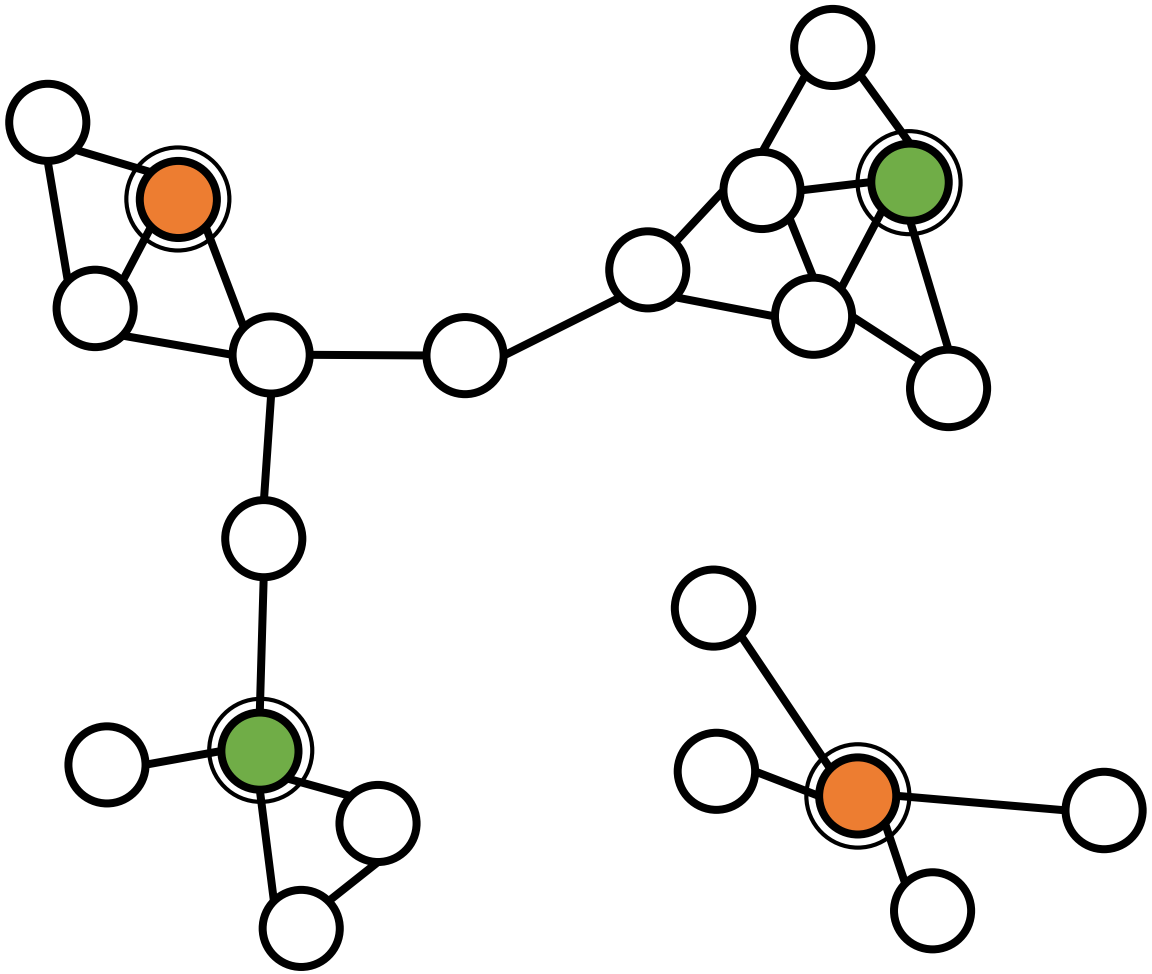
- Cluster and get labels for cluster representatives
- Create a "similarity graph" over data points
  - May use a Mercer kernel for edge weights
  - Don't include edges with very small weight – sparse graph!

Active Learning Attempt I



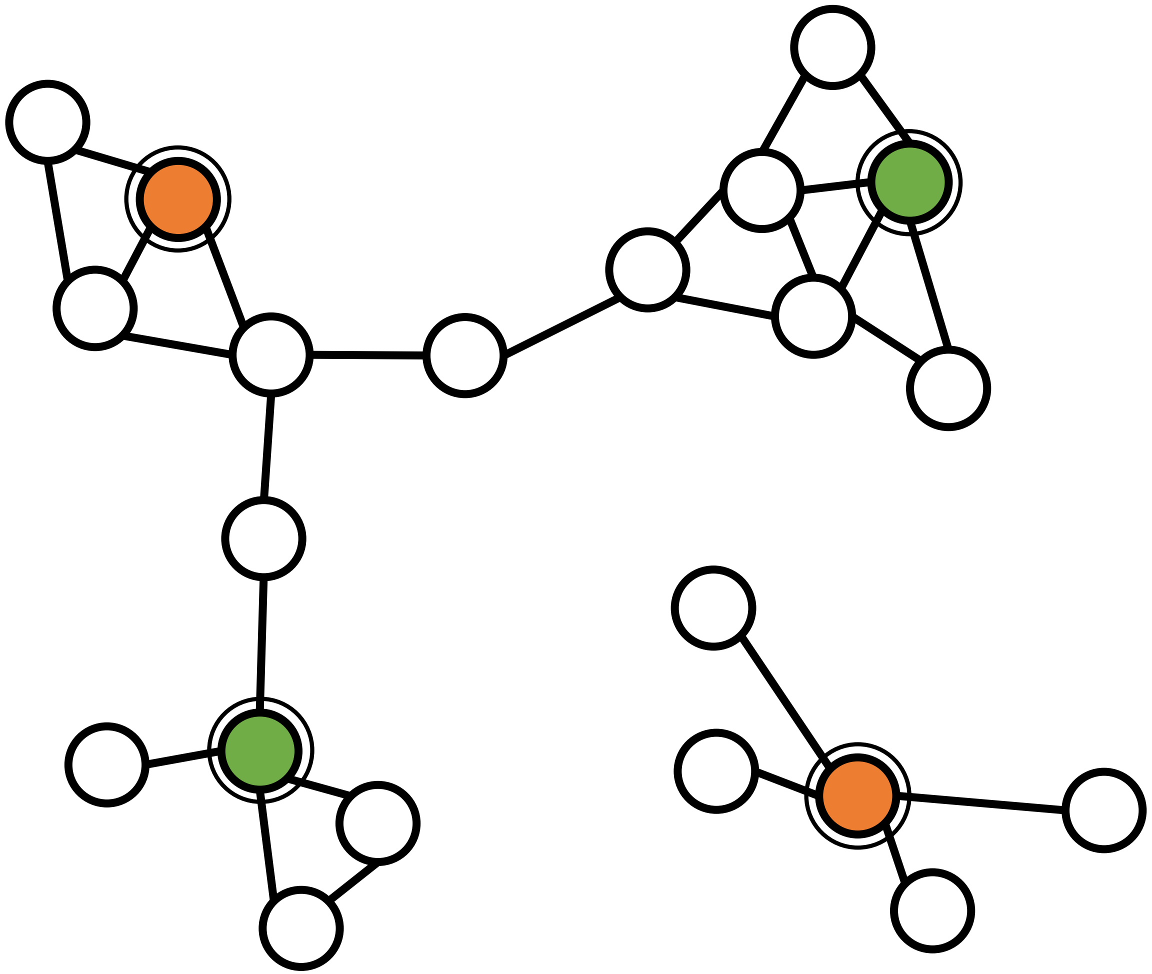
- Cluster and get labels for cluster representatives
- Create a "similarity graph" over data points
  - May use a Mercer kernel for edge weights
  - Don't include edges with very small weight – sparse graph!

Active Learning Attempt I



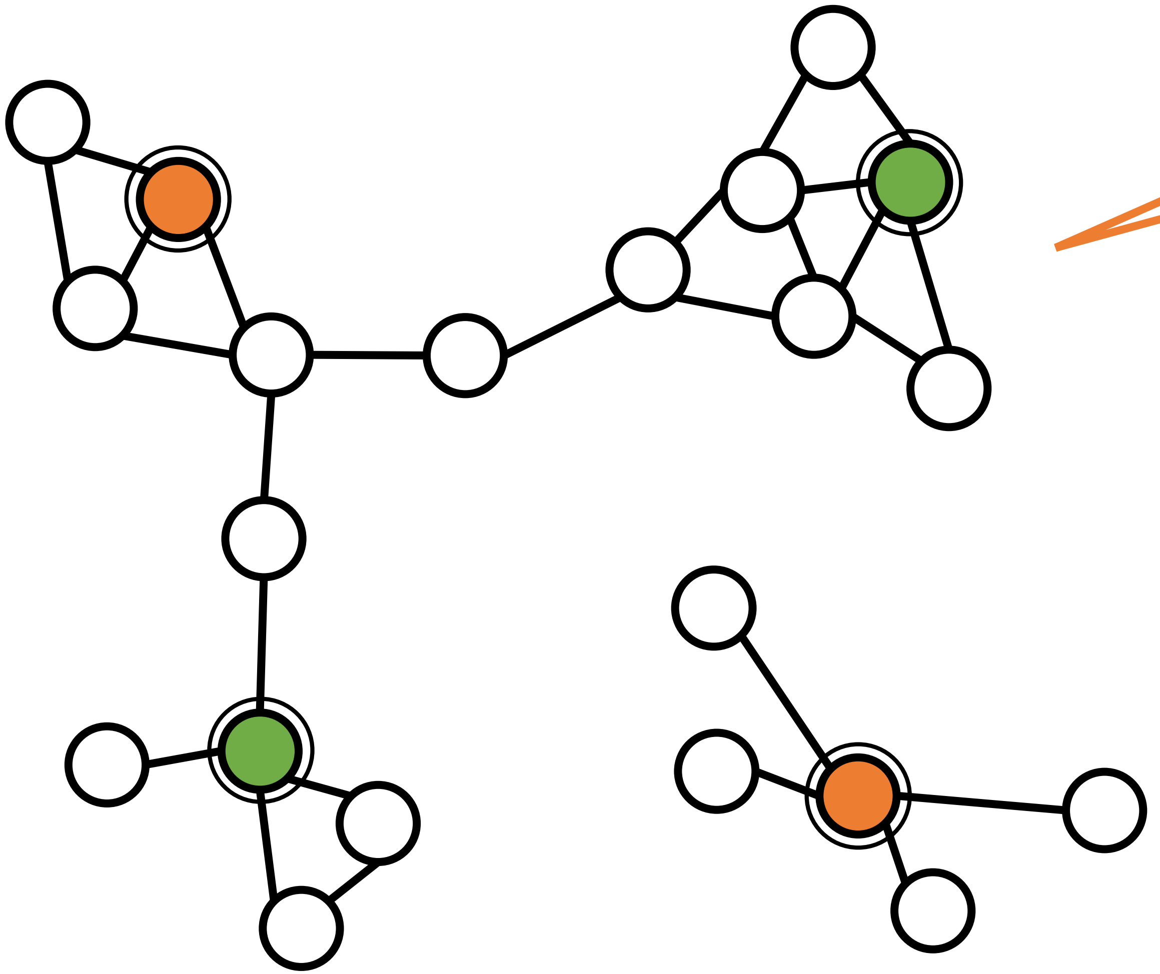
- Cluster and get labels for cluster representatives
- Create a "similarity graph" over data points
  - May use a Mercer kernel for edge weights
  - Don't include edges with very small weight – sparse graph!

Active Learning Attempt I



- Cluster and get labels for cluster representatives
- Create a "similarity graph" over data points
  - May use a Mercer kernel for edge weights
  - Don't include edges with very small weight – sparse graph!
- "Propagate" labels from labelled to unlabelled points

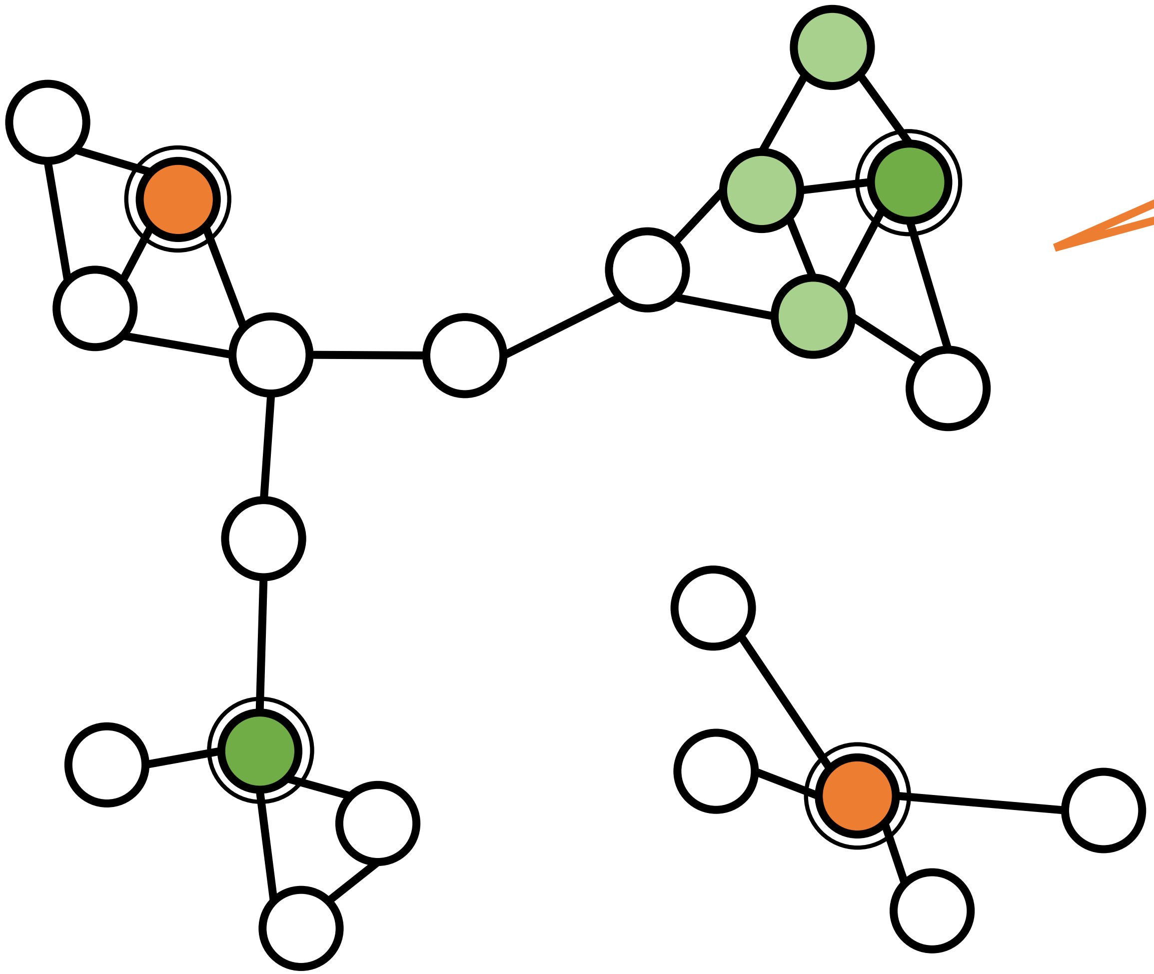
Active Learning Attempt I



Many possibilities e.g. each node passes its label ± 1 to its neighbours multiplied by weight of the edge

- Cl… cluster representatives
- Create a “similarity graph” over data points
  - May use a Mercer kernel for edge weights
  - Don’t include edges with very small weight – sparse graph!
- “Propagate” labels from labelled to unlabelled points

Active Learning Attempt I

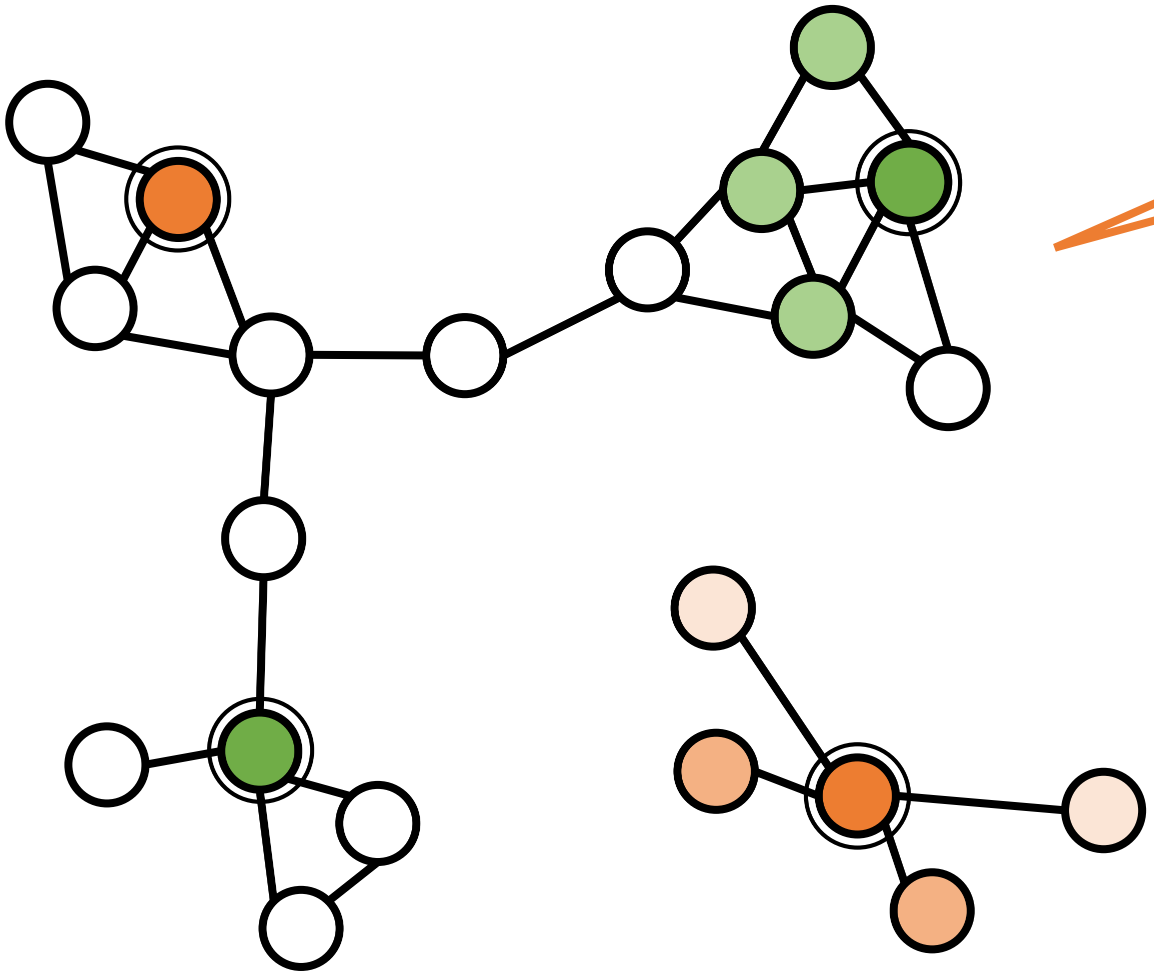


- C…
  clus… representatives

Many possibilities e.g. each node passes its label ± 1 to its neighbours multiplied by weight of the edge

- Create a "similarity graph" over data points
  - May use a Mercer kernel for edge weights
  - Don't include edges with very small weight – sparse graph!
- "Propagate" labels from labelled to unlabelled points

Active Learning Attempt I

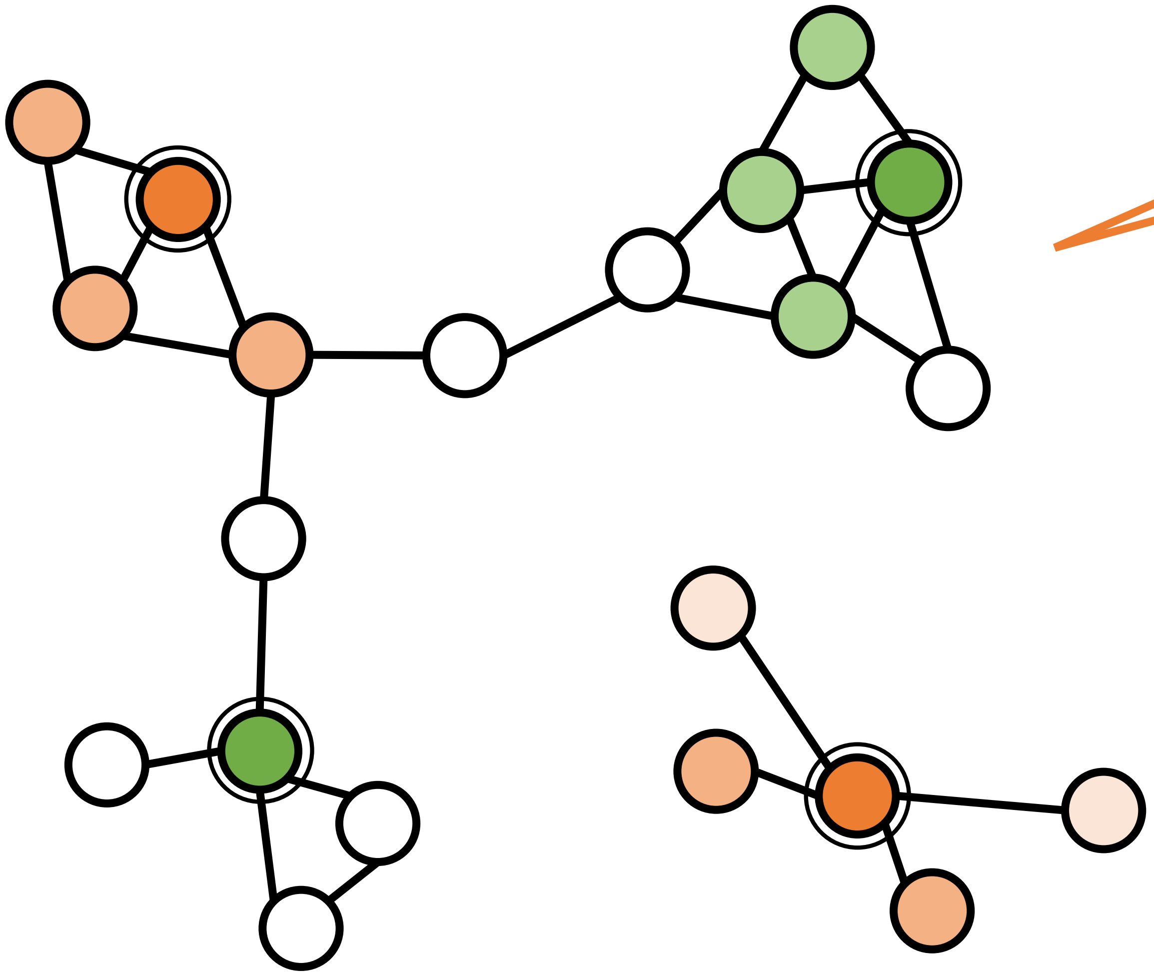


- Cl… cluster representatives

Many possibilities e.g. each node passes its label ± 1 to its neighbours multiplied by weight of the edge

- Create a "similarity graph" over data points
  - May use a Mercer kernel for edge weights
  - Don't include edges with very small weight – sparse graph!
- "Propagate" labels from labelled to unlabelled points

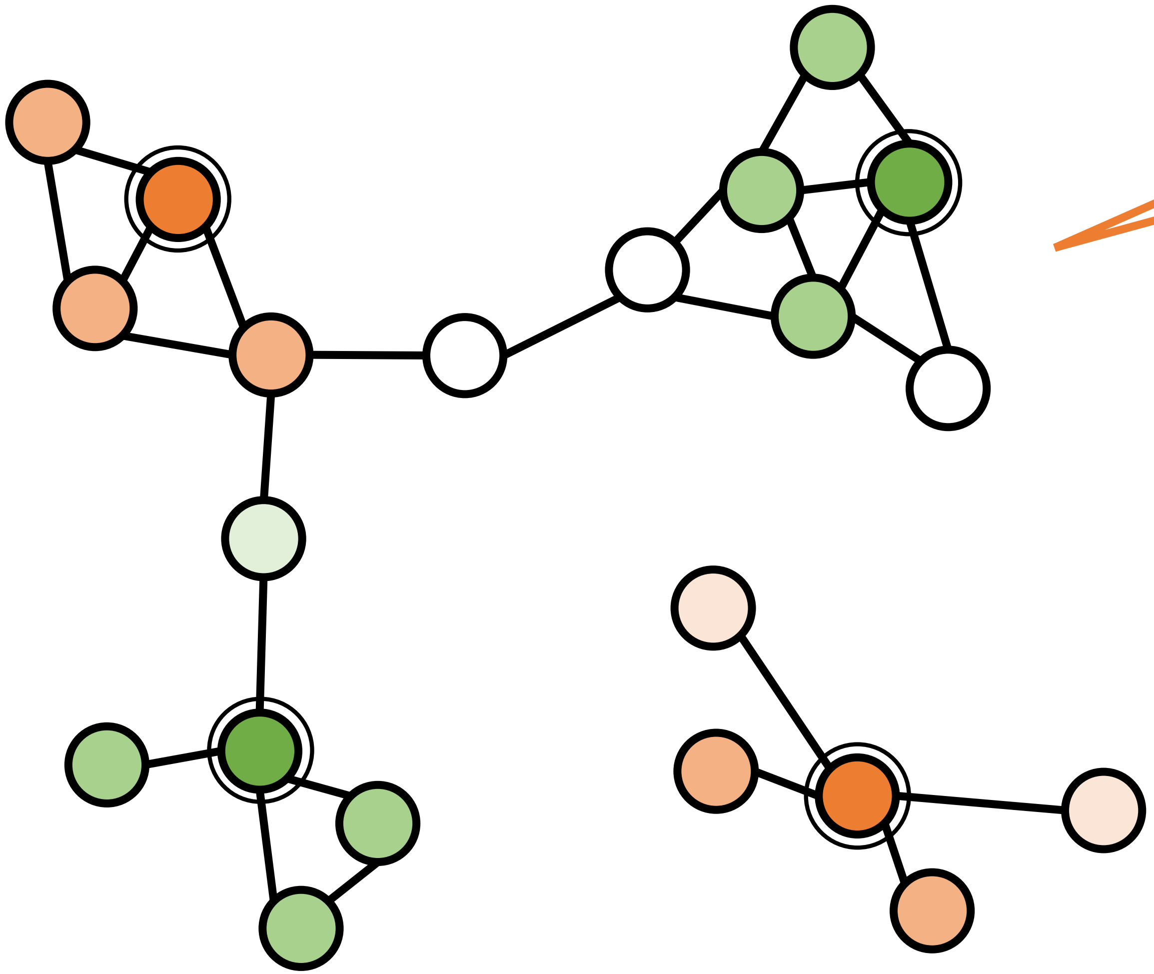
Active Learning Attempt I



Many possibilities e.g. each node passes its label ± 1 to its neighbours multiplied by weight of the edge

- Cl… cluster representatives
- Create a "similarity graph" over data points
  - May use a Mercer kernel for edge weights
  - Don't include edges with very small weight – sparse graph!
- "Propagate" labels from labelled to unlabelled points

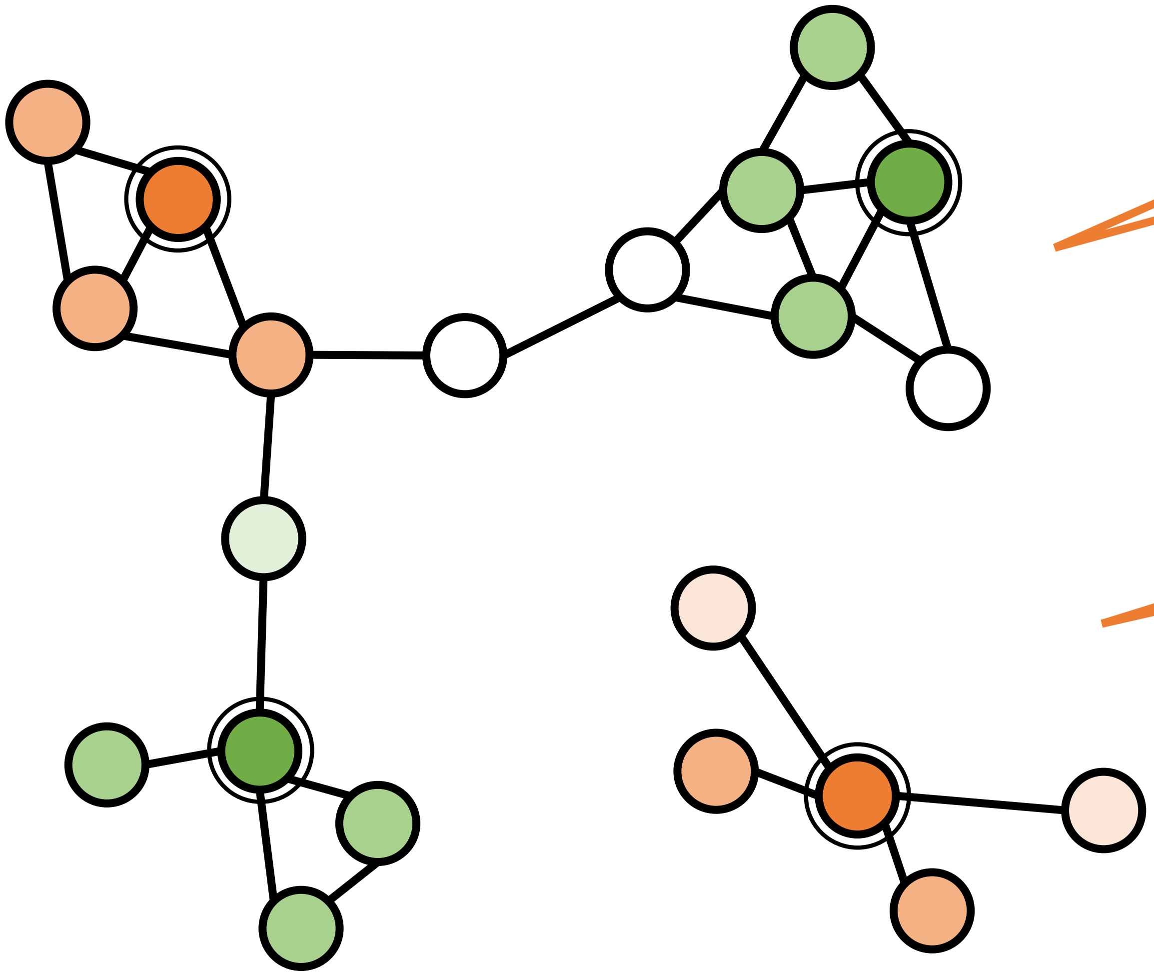
Active Learning Attempt I



Many possibilities e.g. each node passes its label ± 1 to its neighbours multiplied by weight of the edge

- Cl… clus… representatives
- Create a "similarity graph" over data points
  - May use a Mercer kernel for edge weights
  - Don't include edges with very small weight – sparse graph!
- "Propagate" labels from labelled to unlabelled points

Active Learning Attempt I

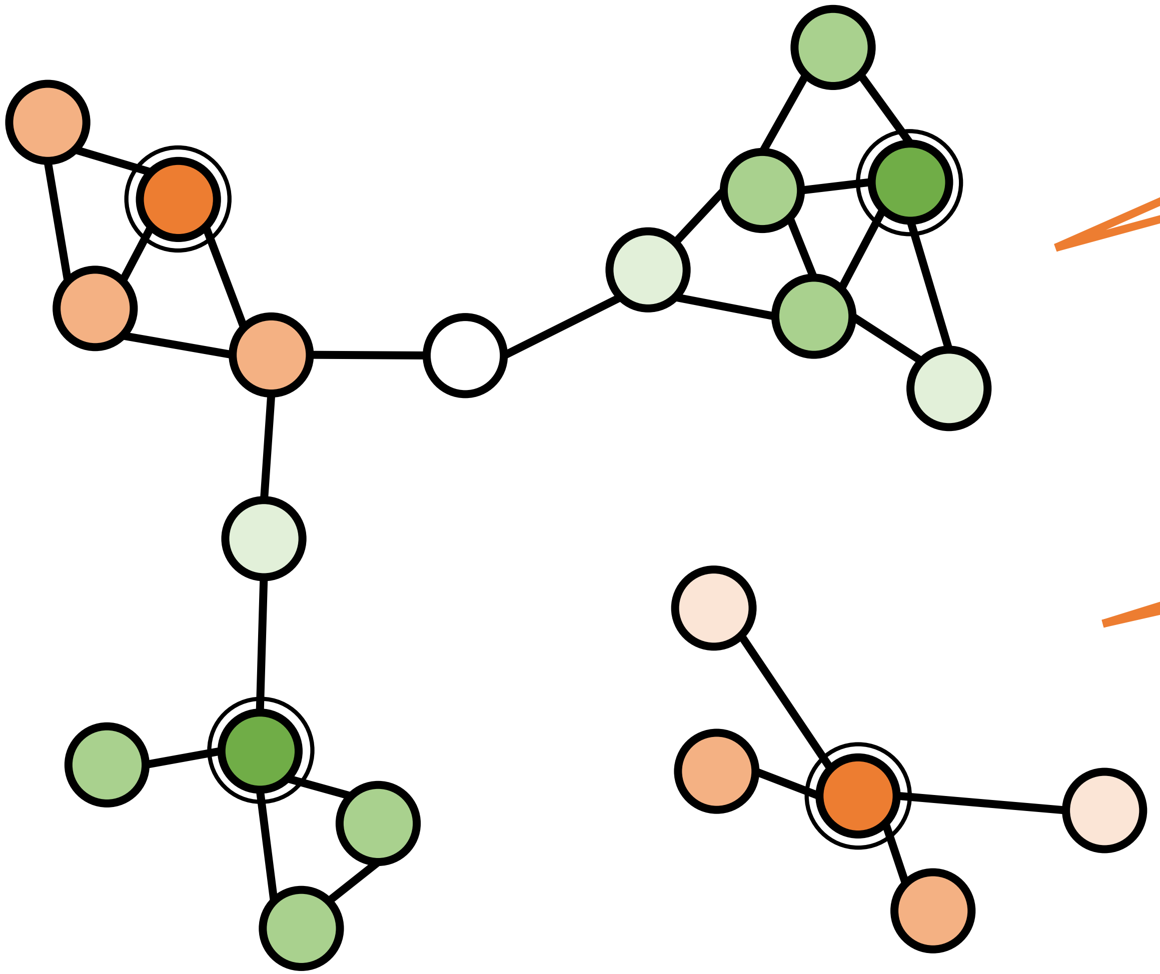


- Cl… clus… representatives
- Cre… over…
  - Ma… …ercer kernel for …ge weights
  - Don't include edges with very small weight – sparse graph!
- "Propagate" labels from labelled to unlabelled points

Many possibilities e.g. each node passes its label ± 1 to its neighbours multiplied by weight of the edge

Neighbors pass them on in successive iterations – "message passing"

Active Learning Attempt I

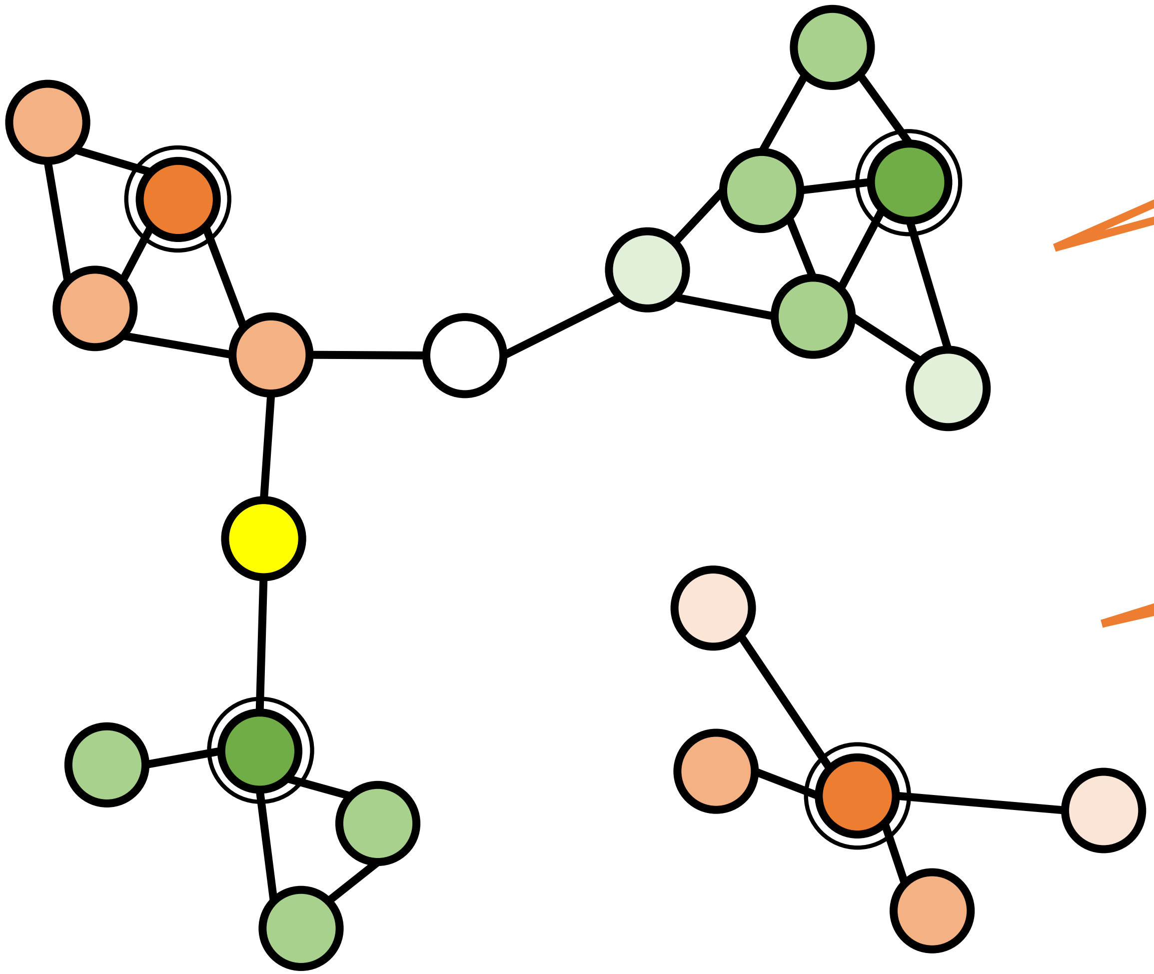


Many possibilities e.g. each node passes its label ±1 to its neighbours multiplied by weight of the edge

Neighbors pass them on in successive iterations – "message passing"

- Cl... clus... representatives
- Cre... over...
  - M... ...ercer kernel for ...ge weights
  - Don't include edges with very small weight – sparse graph!
- "Propagate" labels from labelled to unlabelled points

Active Learning Attempt I

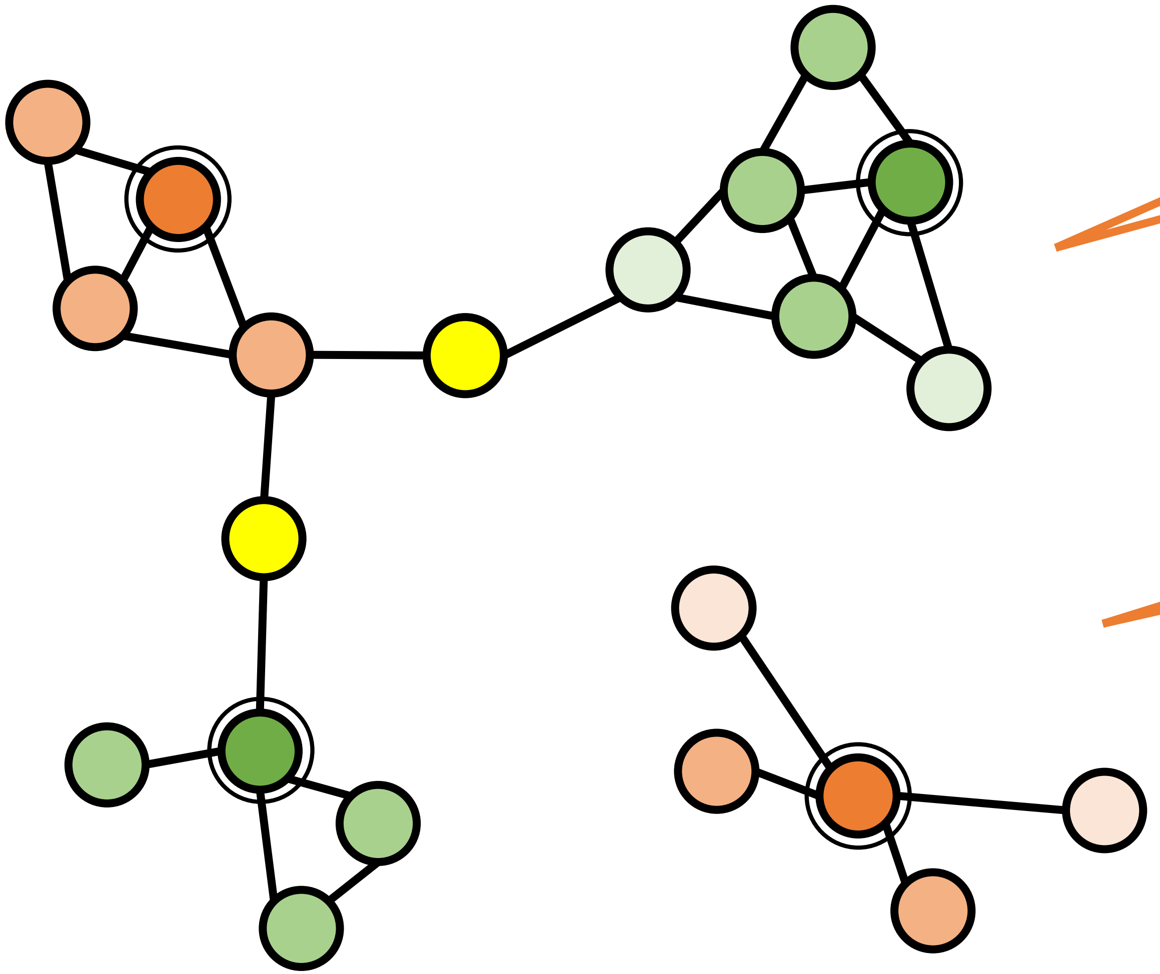


- Cl
  clus representatives
- Cre
  over
  - M
    ge weights
  - Don’t include edges with very small weight – sparse graph!
- “Propagate” labels from labelled to unlabelled points

Many possibilities e.g. each node passes its label ± 1 to its neighbours multiplied by weight of the edge

Neighbors pass them on in successive iterations – “message passing”

Active Learning Attempt I

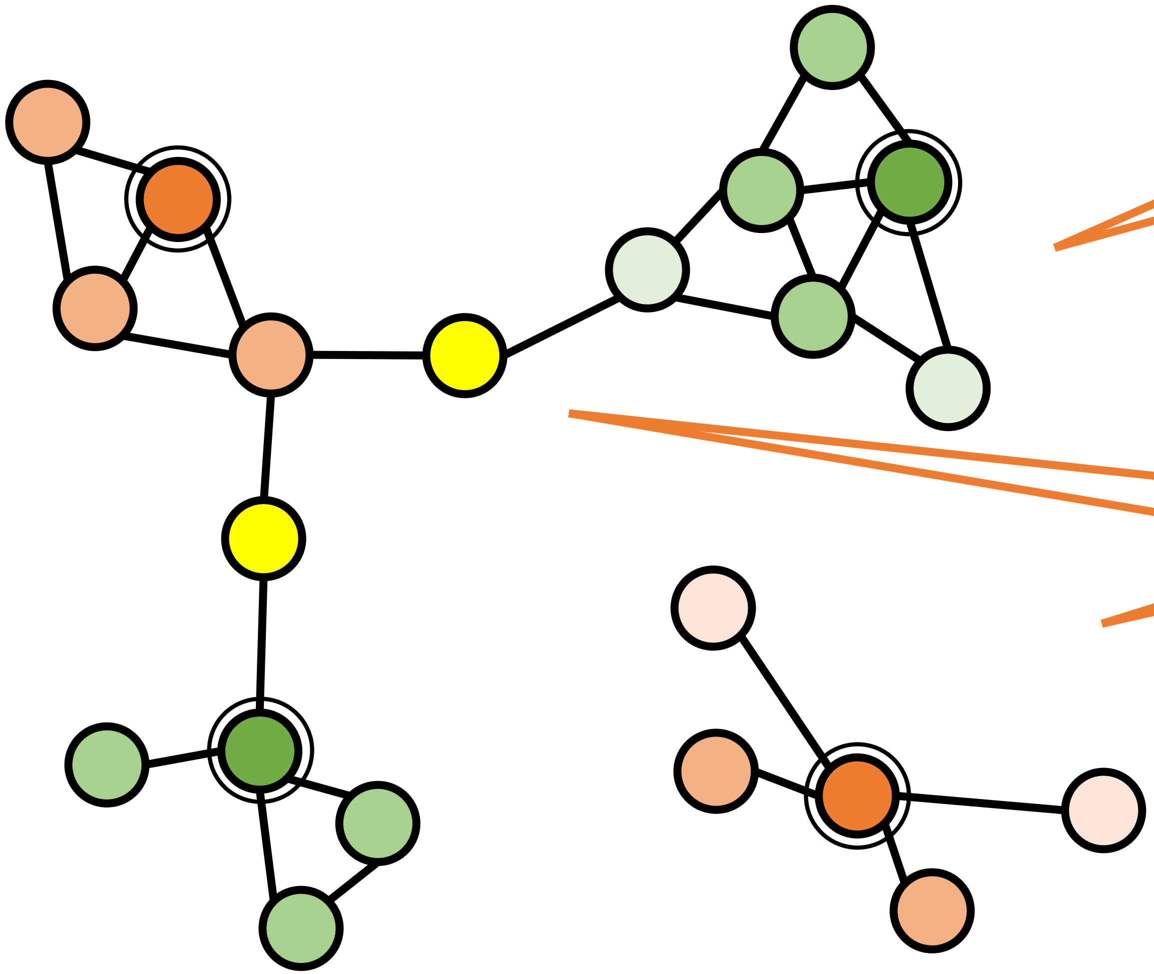


Many possibilities e.g. each node passes its label ±1 to its neighbours multiplied by weight of the edge

Neighbors pass them on in successive iterations – "message passing"

- Cl… clus… representatives
- Cre… over…
  - M… …ercer kernel for …ge weights
  - Don't include edges with very small weight – sparse graph!
- "Propagate" labels from labelled to unlabelled points

Active Learning Attempt I



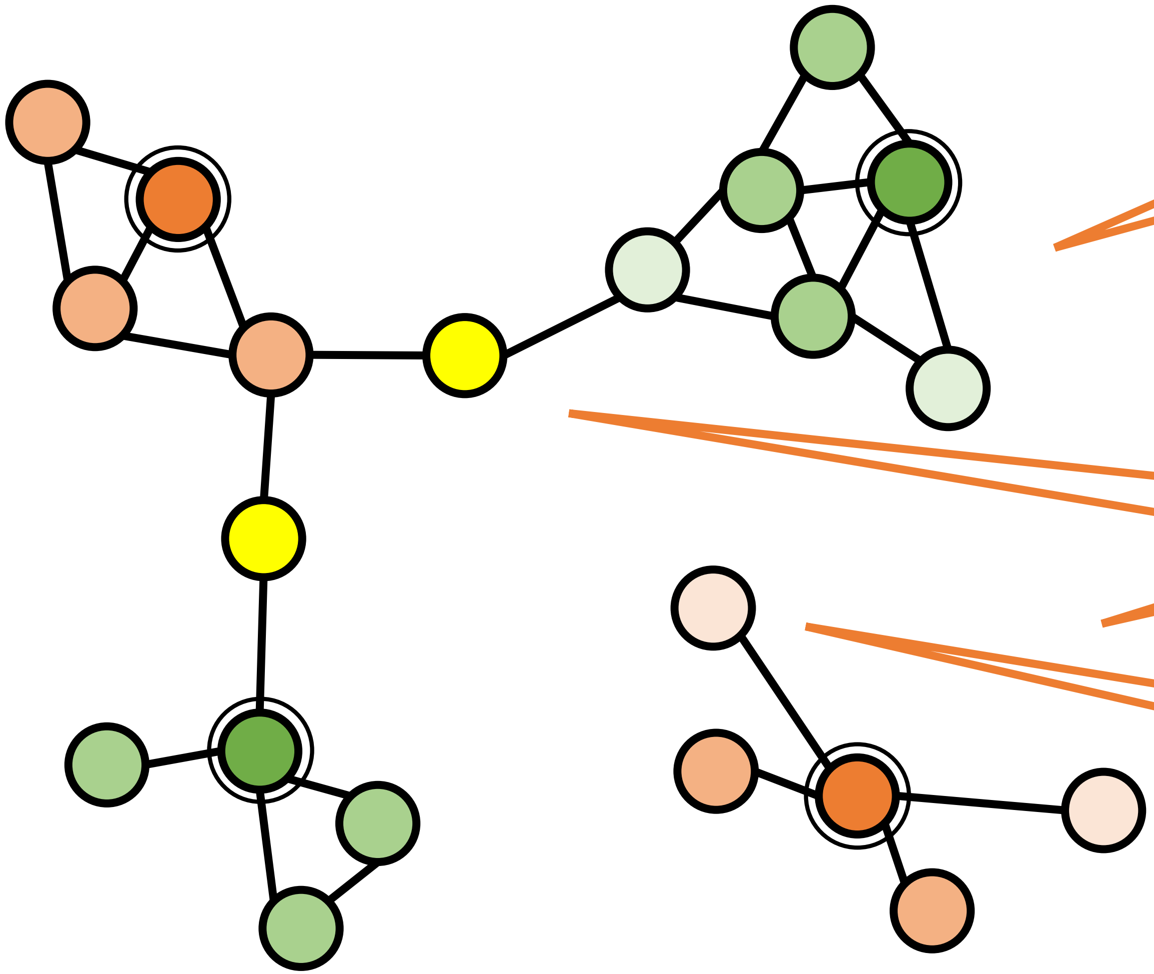
Many possibilities e.g. each node passes its label ± 1 to its neighbours multiplied by weight of the edge

Neighbors pass them on in successive iterations – "message passing"

Some nodes may end up getting mixed messages

- Cl
  clus
- Cre
  over
  - M
  - D
    s
- "Propagate" labels from labelled to unlabelled points

Active Learning Attempt I



- Cl…
  clus… representatives
- Cre… over…
  - May … ercer kernel for
  - D… s…
- … agg…
  labelled …

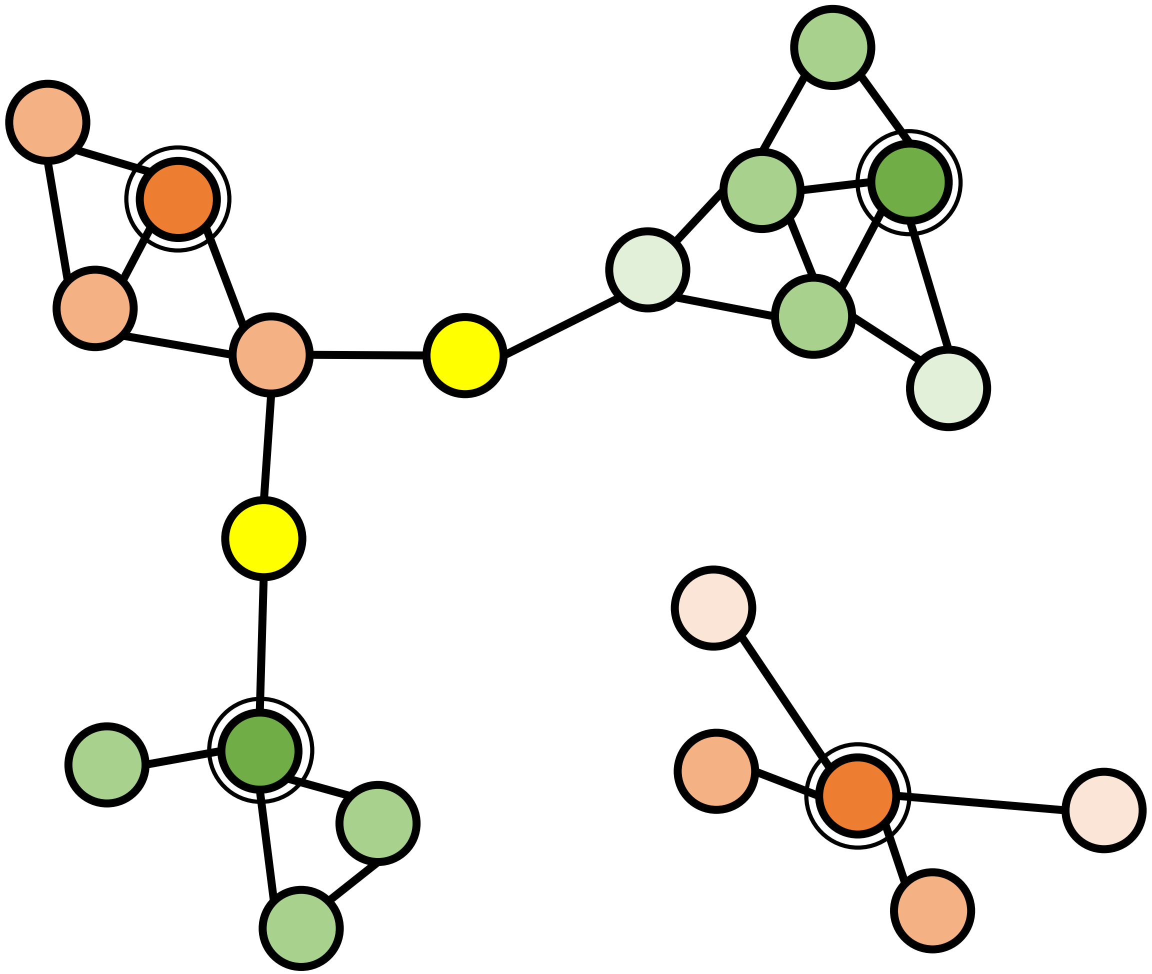
Many possibilities e.g. each node passes its label ±1 to its neighbours multiplied by weight of the edge

Neighbors pass them on in successive iterations – "message passing"

Some nodes may end up getting mixed messages

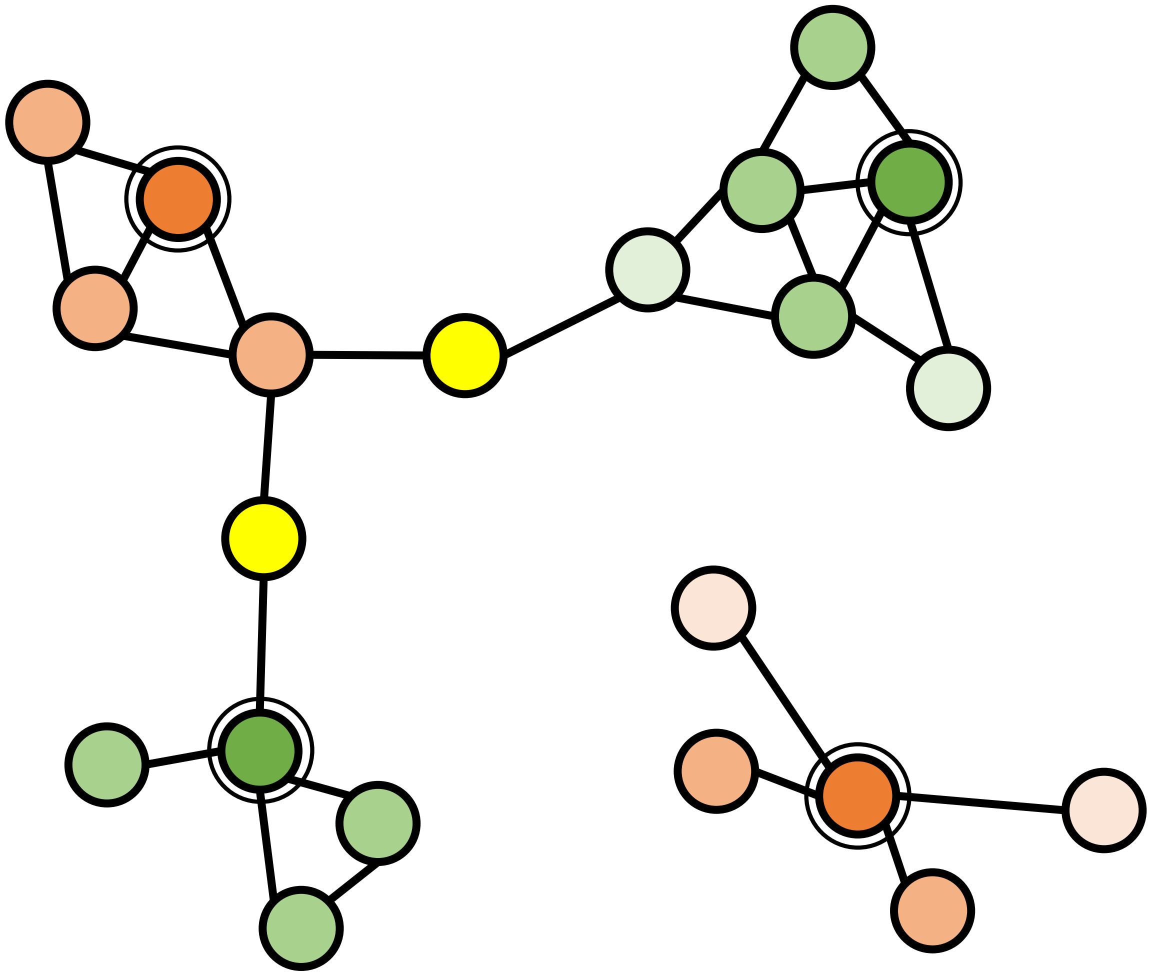
… or be very unsure of their label

Active Learning Attempt I



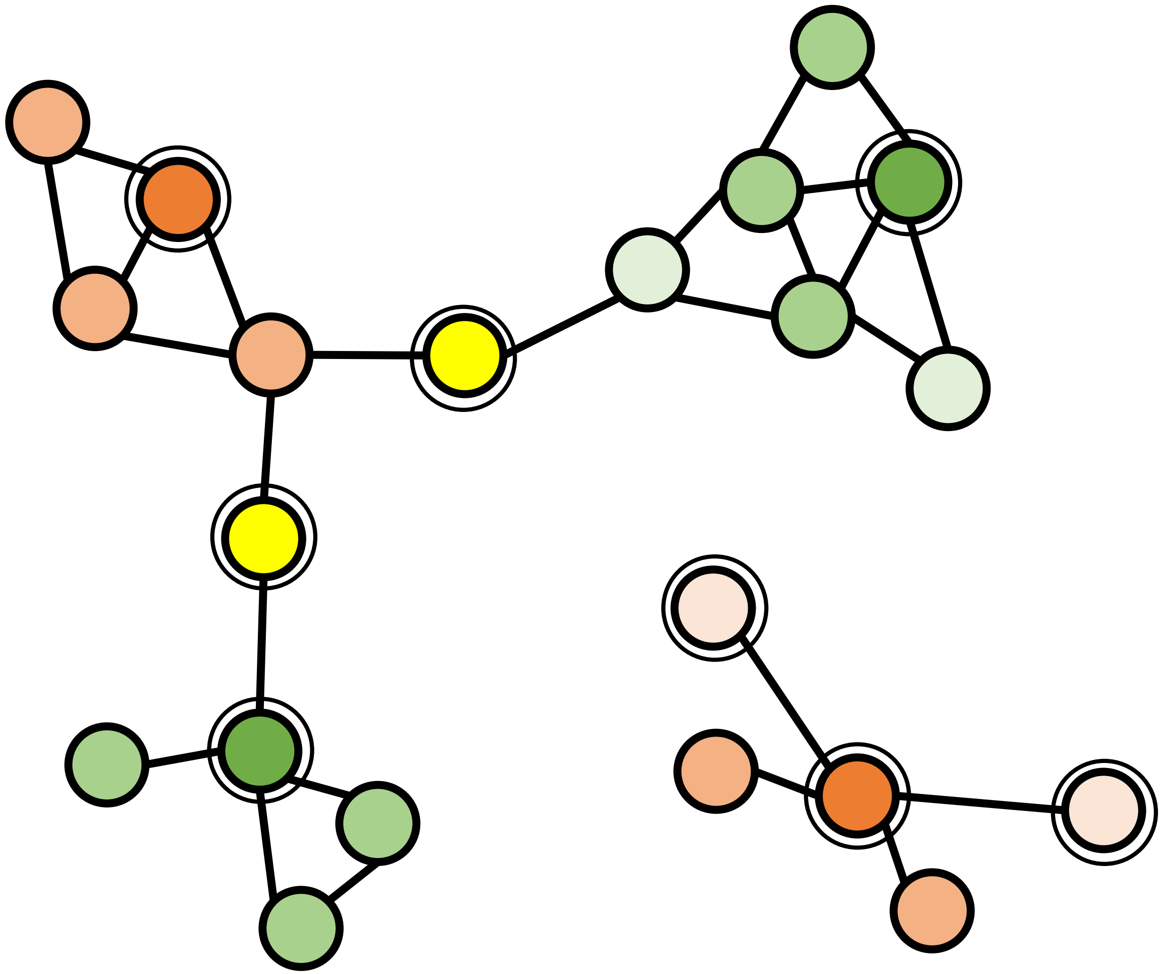
- Cluster and get labels for cluster representatives
- Create a “similarity graph” over data points
  - May use a Mercer kernel for edge weights
  - Don’t include edges with very small weight – sparse graph!
- “Propagate” labels from labelled to unlabelled points

Active Learning Attempt I



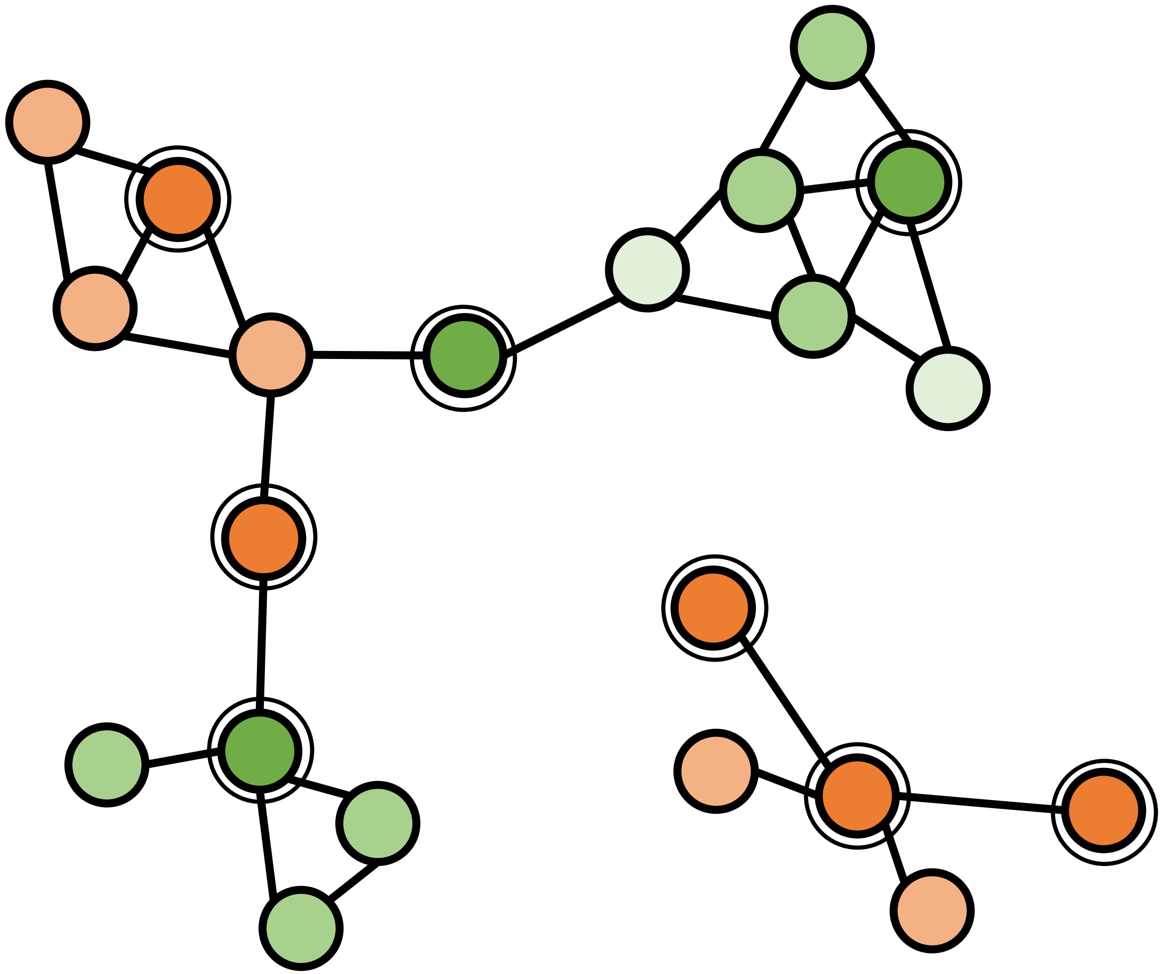
- Cluster and get labels for cluster representatives
- Create a "similarity graph" over data points
  - May use a Mercer kernel for edge weights
  - Don't include edges with very small weight – sparse graph!
- "Propagate" labels from labelled to unlabelled points
- Query unlabelled data points that are "confused"

Active Learning Attempt I



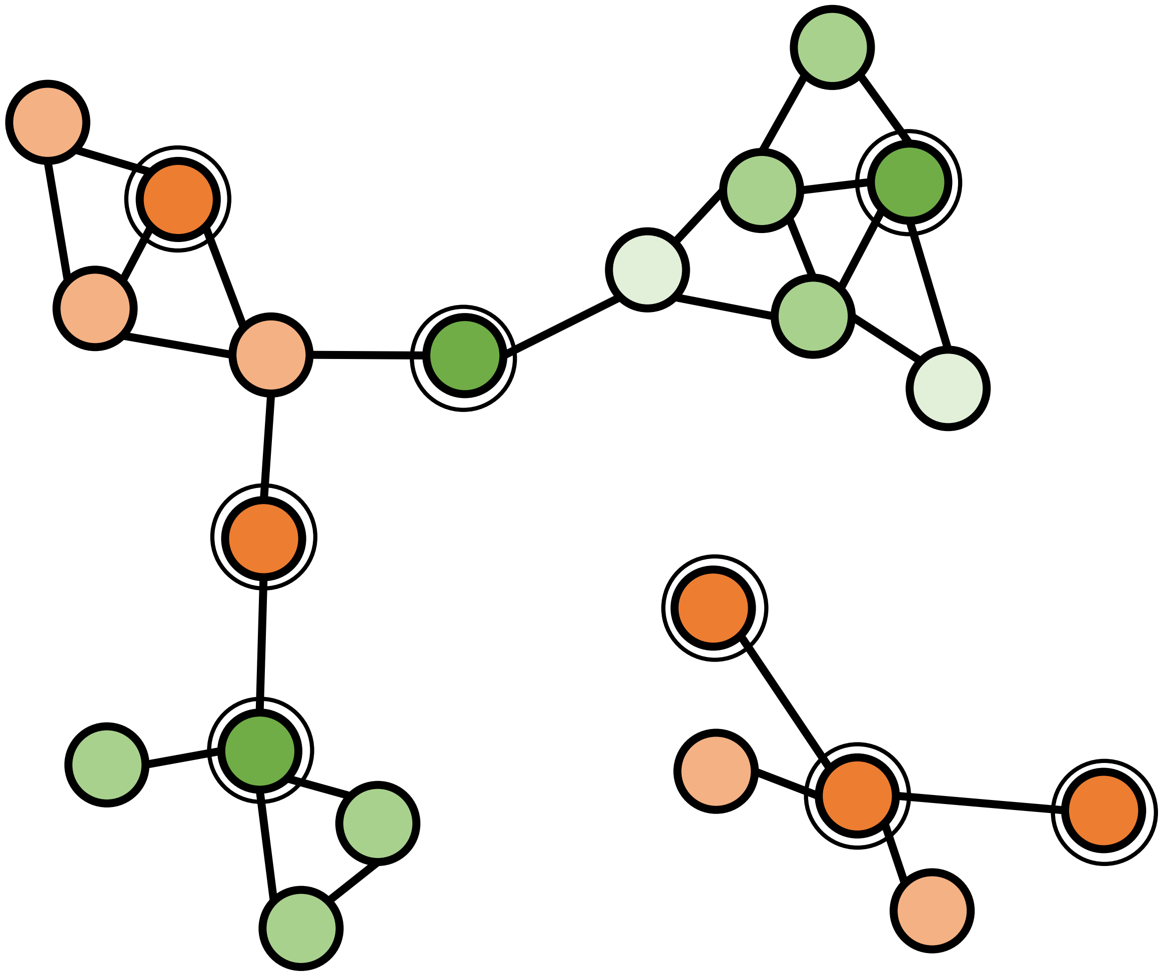
- Cluster and get labels for cluster representatives
- Create a “similarity graph” over data points
  - May use a Mercer kernel for edge weights
  - Don’t include edges with very small weight – sparse graph!
- “Propagate” labels from labelled to unlabelled points
- Query unlabelled data points that are “confused”

Active Learning Attempt I



- Cluster and get labels for cluster representatives
- Create a "similarity graph" over data points
  - May use a Mercer kernel for edge weights
  - Don't include edges with very small weight – sparse graph!
- "Propagate" labels from labelled to unlabelled points
- Query unlabelled data points that are "confused"

Active Learning Attempt I



- Cluster and get labels for cluster representatives
- Create a “similarity graph” over data points
  - May use a Mercer kernel for edge weights
  - Don’t include edges with very small weight – sparse graph!
- “Propagate” labels from labelled to unlabelled points
- Query unlabelled data points that are “confused”
- Repeat!

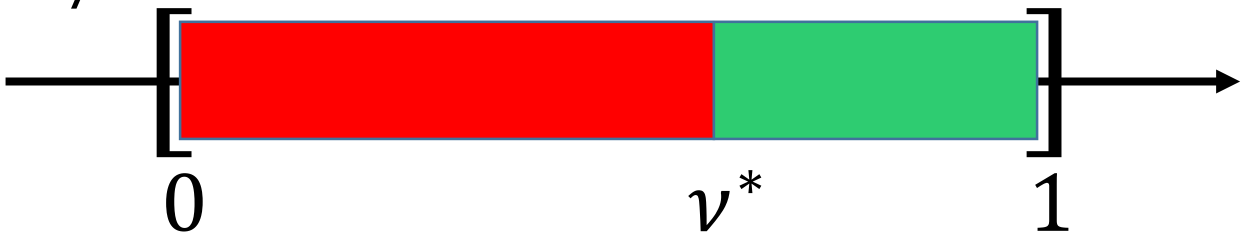
Active Learning Attempt II

Learning theoretic result

- To learn classifier with ϵ error, SVM/NN need $\mathcal{O}\left(\frac{1}{\epsilon}\right)$ labeled data pts
- Suppose data $S \subset [0,1]$ and true classifier is a threshold

$$y = 2 \cdot \mathbb{I}\{x \geq v^*\} - 1$$

- Notice that we can perform binary search to estimate v^*



- Let $a = 0, b = 1$
- For $t = 1,2, \ldots$
  - Query the label of the data point $(a+b)/2$
  - If label is -1 let $a \leftarrow (a+b)/2$ else let $b \leftarrow (a+b)/2$
  - If $b - a < \epsilon$, stop and output $\hat{v} = (a+b)/2$
- We queried only $\log(1/\epsilon)$ labels and yet ensured $|\hat{v} - v^*| < \epsilon$
- Exponential improvement in number of labelled samples needed!

Active Learning Attempt II

- This is the most simple form of *version space* search
- Search the space of all classifiers
  - In the previous example, classifiers corresponded to thresholds
- Can get tricky to do this efficiently even over linear classifiers
- Also tricky to handle cases that are not linearly separable
- Can't we do boosting or bagging to solve this problem?
- Yes, search for ActBoost (the name is lame but the algo isn't)

Active Learning

- Many other techniques to query labels
- If working with a Bayesian model, query data point whose predictive posterior variance is largest
- Query-by-committee – maintain a committee of several models and query data point on which committee disagrees the most
- Related concept of disagreement coefficient
- Expected model drift – maintain a single model but query the data point whose label, if known, will change the model the most

Semi-supervised Learning

Semi-supervision

- Labeled data is expensive since manual effort often involved
- Unlabelled data often free to obtain by crawling repo/www
- Why is unlabelled data of any use?
- Since it gives us valuable access to $\mathbb{P}[x]$
- If we have a super awesome model of $\mathbb{P}[x]$ then under some assumptions, a few labelled data points enough to learn $\mathbb{P}[y,x]$
- Assumptions?? Yes, SSL usually works by imposing a heavy inductive bias
  - Points deemed similar by blah notion of similarity have similar labels
  - Points in the same cluster given by blah clustering algo have same labels
  - Some assumption necessary to relate labelled and unlabelled data
- If assumption is inappropriate then SSL method suffers

SSL Attempt I – Generative Learning

- Labelled $L = \{x^i, y^i\}_{i=1,\ldots,n}$ and unlabeled data $U = \{x^j\}_{j=n+1,\ldots,n+m}$
- Learn a generative model Θ that models $\mathbb{P}[x]$ and $\mathbb{P}[y \mid x]$

$$\log \mathbb{P}[L \cup U \mid \Theta] = \log \mathbb{P}[L \mid \Theta] + \log \mathbb{P}[U \mid \Theta]$$

$$= \sum_{i=1}^{n} \log \mathbb{P}\left[y^i \mid x^i, \Theta\right] + \log \mathbb{P}\left[x^i \mid \Theta\right] + \sum_{j=n+1}^{n+m} \log \mathbb{P}\left[x^j \mid \Theta\right]$$

- Introduce latent variables $z^j \in [C]$ to model labels of points in U

$$= \sum_{i=1}^{n} \log \mathbb{P}\left[y^i \mid x^i, \Theta\right] + \log \mathbb{P}\left[x^i \mid \Theta\right] + \sum_{j=n+1}^{n+m} \log\left(\sum_{z^j=1}^{C} \mathbb{P}\left[z^j \mid x^j, \Theta\right] \cdot \mathbb{P}\left[x^j \mid \Theta\right]\right)$$

- Have fun executing the EM algorithm on this ☺

SSL Attempt II - Bootstrapping

- Basically the hard-assignment version for the previous slide
- Makes sense even in non-PML settings though
- Use any nice algo on L to learn a classifier f^0
- For $t = 1,2, \ldots$
  - Use f^{t-1} to label data points in U i.e. $z^{t,j} = f^{t-1}(x^j), j = n+1, \ldots,$
  - Use the nice algo on the dataset $L \cup \{x^j, z^{t,j}\}$ to learn f^t
  - Stop when tired
- Very simple to try out first before attempting more fancy methods
- Be careful – may reinforce wrong predictions
- No way to even detect if f^t getting better on U or not

SSL Attempt III - Regularization

- Incorporate a prior to force "similar" points to have similar labels
- Lets switch to a regression problem for simplicity
- Assume a notion of similarity between points $a_{ij} = K(x^i, x^j)$
  - Should not depend on the label i.e. can be computed over U as well
- Want $f(x^i)$ to be close to $f(x^j)$ if a_{ij} is large
- Can enforce this by asking $a_{ij} \cdot \left(f(x^i) - f(x^j)\right)^2$ to be small

$$\arg\min_f \sum_{i=1}^{n} \ell(f(x^i), y^i) + \sum_{i,j=1}^{m+n} a_{ij} \cdot \left(f(x^i) - f(x^j)\right)^2$$

- Can also incorporate a usual regularizer $r(f)$

SSL Attempt III - Regularization

- Incorporate a prior to force
- Lets switch to a regression
- Assume a notion of similarit
  - Should not depend on the lab
- Want $f(x^i)$ to be close to $f(x^j)$ i $_{ij}$ is large
- Can enforce this by asking $a_{ij} \cdot \left(f(x^i) - f(x^j)\right)^2$ to be small

$$\arg\min_{f} \sum_{i=1}^{n} \ell\left(f(x^i), y^i\right) + \sum_{i,j=1}^{m+n} a_{ij} \cdot \left(f(x^i) - f(x^j)\right)^2$$

- Can also incorporate a usual regularizer $r(f)$

If $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle$ and $r(\mathbf{w}) = \lambda \cdot \|\mathbf{w}\|_2^2$, then

$$\arg\min_{w} \sum_{i=1}^{n} \ell\left(\langle \mathbf{w}, \mathbf{x}^i \rangle, y^i\right) + \lambda \cdot \|\mathbf{w}\|_2^2 + \mathbf{w}^\top X L X^\top \mathbf{w}$$

where $X = [\mathbf{x}^1, \dots, \mathbf{x}^{m+n}] \in \mathbb{R}^{d \times (m+n)}$. Find L?

Additional regularization $\mathbf{w}^\top XLX^\top\mathbf{w}$. Overall regularization $\mathbf{w}^\top R\mathbf{w}$ where $R = XLX^\top + \lambda \cdot I$

ation

If $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle$ and $r(\mathbf{w}) = \lambda \cdot \|\mathbf{w}\|_2^2$, then

$$\arg\min_{w} \sum_{i=1}^{n} \ell(\langle \mathbf{w}, \mathbf{x}^i \rangle, y^i) + \lambda \cdot \|\mathbf{w}\|_2^2 + \mathbf{w}^\top XLX^\top \mathbf{w}$$

where $X = [\mathbf{x}^1, \ldots, \mathbf{x}^{m+n}] \in \mathbb{R}^{d\times(m+n)}$. Find L?

- Incorporate a prior to force
- Lets switch to a regression p
- Assume a notion of similarit
  - Should not depend on the lab
- Want $f(x^i)$ to be close to $f(x^j)$ i $_{ij}$ is large
- Can enforce this by asking $a_{ij} \cdot \left(f(x^i) - f(x^j)\right)^2$ to be small

$$\arg\min_{f} \sum_{i=1}^{n} \ell(f(x^i), y^i) + \sum_{i,j=1}^{m+n} a_{ij} \cdot \left(f(x^i) - f(x^j)\right)^2$$

- Can also incorporate a usual regularizer $r(f)$

SSL Attempt IV – Transductive SVM

- Transduction is a special form of learning where test features are used to train the model – note, test labels still taboo to look at
- However, the "transductive" SVM works in SSL settings too
- Minimizes hinge loss on labelled points and keeps unlabelled points away from the hyperplane on at least one side

$$\min_{\mathbf{w},\{\xi_i\}} \frac{1}{2}\|\mathbf{w}\|_2^2 + \sum_{i=1}^{m+n} \xi_i$$

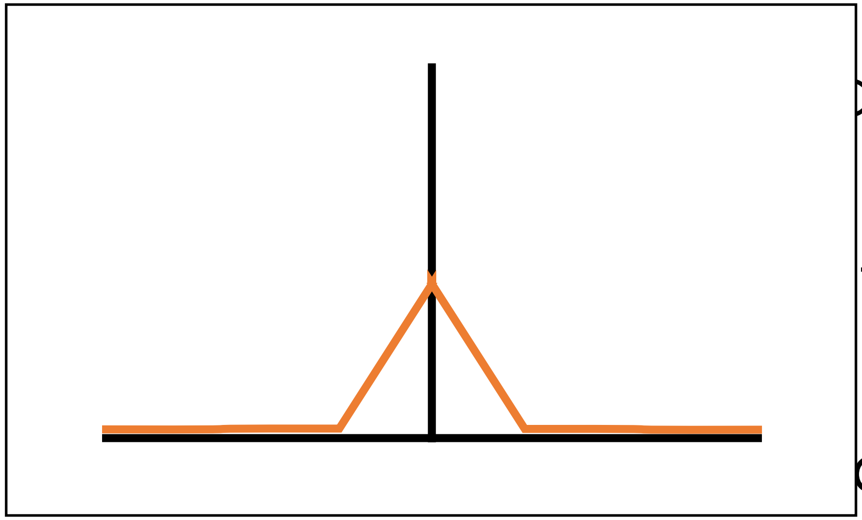
$$y^i \cdot \langle \mathbf{w}, \mathbf{x}^i \rangle \geq 1 - \xi_i, i = 1, \ldots, n$$

$$|\langle \mathbf{w}, \mathbf{x}^j \rangle| \geq 1 - \xi_j, j = n+1, \ldots, m+n$$

$$\xi_i \geq 0, i = 1, \ldots, n, n+1, \ldots, m+n$$

- Results in a non-convex optimization problem – difficult to solve

SSL Attempt IV – Transductive SVM



- al for
  – no
- tive" SVM
- label points an
  ane on at least on

Can show that for unlabelled points, TSVM uses the symmetric hinge loss function $\ell(\mathbf{w}, \mathbf{x}^j) = \min\{[1 - \langle \mathbf{w}, \mathbf{x}^j \rangle]_+, [1 + \langle \mathbf{w}, \mathbf{x}^j \rangle]_+\}$

Note that if $|\langle \mathbf{w}, \mathbf{x}^j \rangle| \geq 1 - \xi_j$ then $z^j \cdot \langle \mathbf{w}, \mathbf{x}^j \rangle \geq 1 - \xi_j$ for either $z^j = 1$ or $z^j = -1$

$$\min_{\mathbf{w}, \{\xi_i\}} \frac{1}{2}\|\mathbf{w}\|_2^2 + \sum_{i=1}^{m+n} \xi_i$$

$$y^i \cdot \langle \mathbf{w}, \mathbf{x}^i \rangle \geq 1 - \xi_i, i = 1, \ldots, n$$

$$|\langle \mathbf{w}, \mathbf{x}^j \rangle| \geq 1 - \xi_j, j = n+1, \ldots, m+n$$

$$\xi_i \geq 0, i = 1, \ldots, n, n+1, \ldots, m+n$$

- Results in a non-convex optimization problem – difficult to solve

SSL Attempt IV – Transductive SVM

Exercise: show TSVM actually solves

$$\min_{\mathbf{w},\{\xi_i\},\{z^j\}} \frac{1}{2}\|\mathbf{w}\|_2^2 + \sum_{i=1}^{m+n} \xi_i$$

$$y^i \cdot \langle \mathbf{w}, \mathbf{x}^i \rangle \geq 1 - \xi_i, i = 1, \ldots, n$$

$$z^j \cdot \langle \mathbf{w}, \mathbf{x}^j \rangle \geq 1 - \xi_j, j = n+1, \ldots, m+n$$

$$\xi_i \geq 0, i = 1, \ldots, m+n$$

$$z^j \in \{-1, +1\}, j = n+1, \ldots, m+n$$

i.e. jointly optimizes latent vars and model

points, TSVM
ss function
$+ \langle \mathbf{w}, \mathbf{x}^j \rangle]_+\}$

$, \mathbf{x}^j \rangle| \geq 1 - \xi_j$
$\rangle \geq 1 - \xi_j$ for
or $z^j = -1$

$$|\langle \mathbf{w}, \mathbf{x}^j \rangle| \geq 1 - \xi_j, j = n + 1, \ldots, m + n$$

$$\xi_i \geq 0, i = 1, \ldots, n, n + 1,$$

- Results in a non-convex optimization p

Bootstrapping using an SVM would have solved the same problem using hard AltMin

Time to Wrap Up!

A whirlwind tour of topics we did not discuss

Many topics we did not cover

- Bayesian learning (applicable to several problem areas – CS775)
- Graphical models (model interactions within data feat. - CS772)
- Transfer learning (when exam syllabus may have a "drift")
- Reinforcement learning (learn from an evolving teacher – CS773)
- Multi-task learning (solve several problems on the same data)
- Multi-view learning (different feature rep. of the same data)
- Sparse recovery (learn space-efficient models – CS772/CS777)
- Time series modelling (HMMs, AR, ARMA, RNNs)
- Learning theory (provable goodness of learnt models – CS777)
- Learning on data streams (experts, bandits – CS773)
- Advanced Optimization Techniques (CS774/CS777)

Please give your Feedback

http://tinyurl.com/ml17-18afb