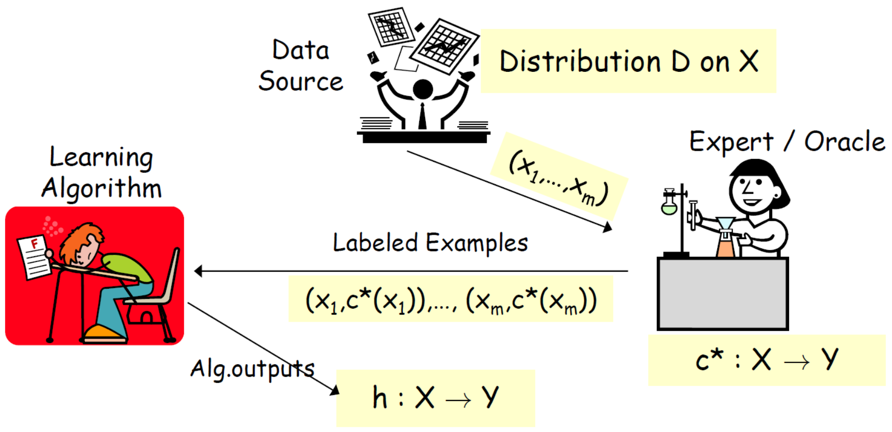
Discussion 09

2022.5.12

Superviesd VS semi-supervised

Fully Supervised Learning



S_l={$(x_1, y_1), \dots,(x_{m_l}, y_{m_l})$}

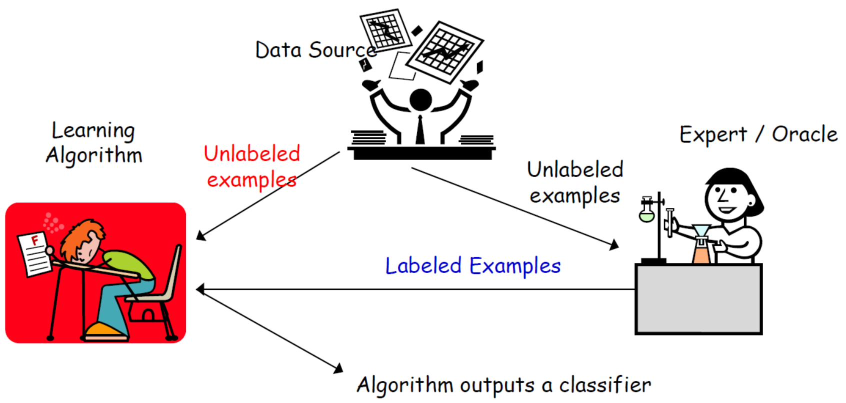
x_i drawn i.i.d from D, $y_i = c^*(x_i)$

Goal: h has small error over D.

$err_D(h) = \Pr_{x\sim D}(h(x) \neq c^*(x))$

VS

Semi-Supervised Learning



S_l={$(x_1, y_1), \dots,(x_{m_l}, y_{m_l})$}

x_i drawn i.i.d from D, $y_i = c^*(x_i)$

S_u={$x_1, \dots,x_{m_u}$} drawn i.i.d from D

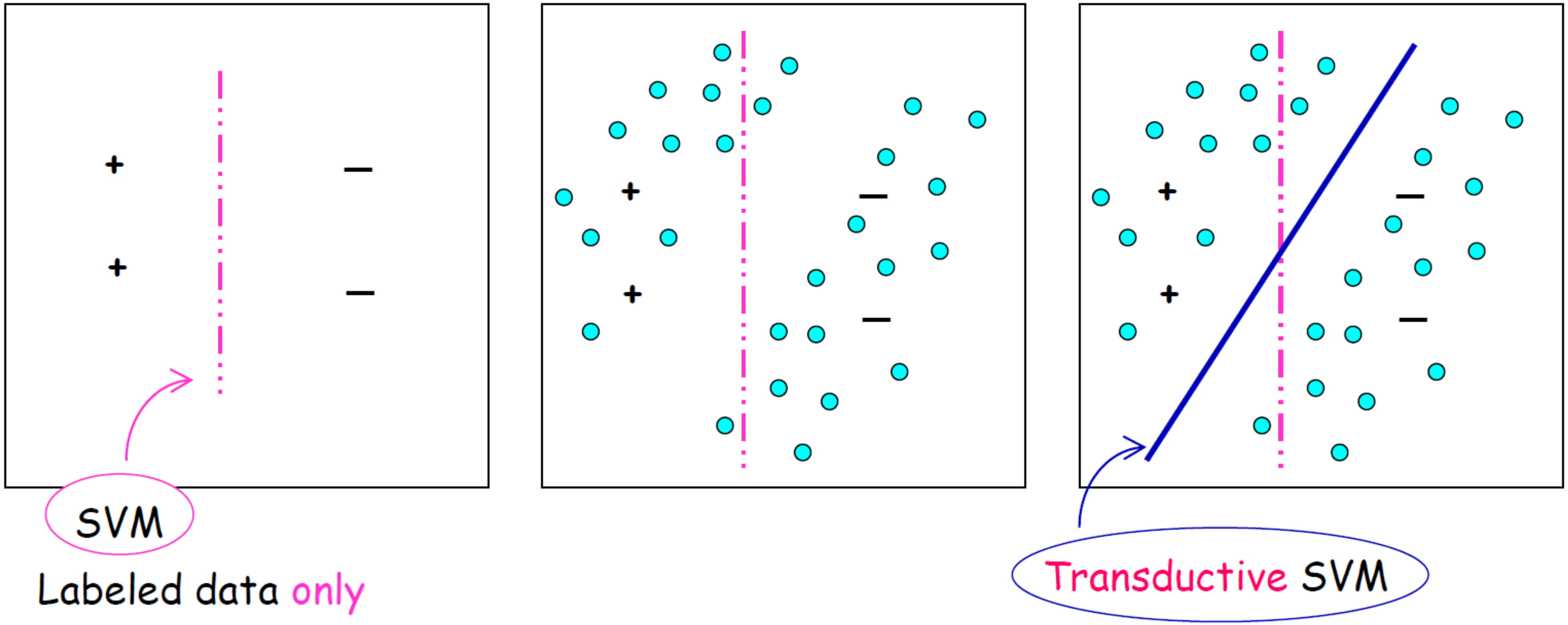
Goal: h has small error over D.

$err_D(h) = \Pr_{x\sim D}(h(x) \neq c^*(x))$

Semi-supervised Learning

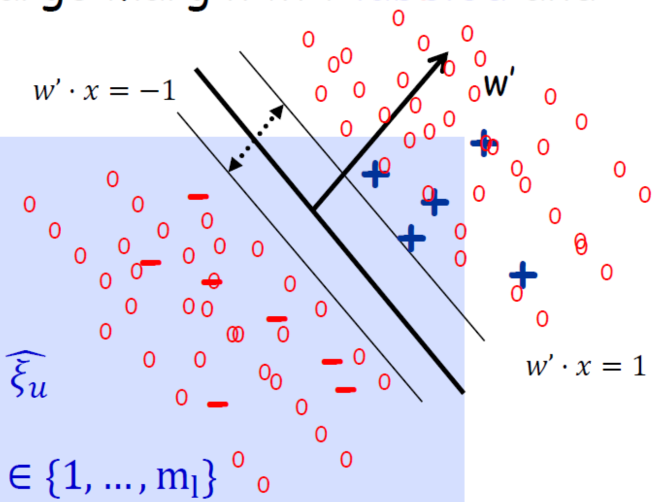
- Transductive SVM
- Co-training
- Graph-based methods

Transductive SVM



Transductive SVM

Optimize for the separator with large margin wrt labeled and unlabeled data. [Joachims '99]



Input: $S_l=\{(x_1,y_1), \ldots,(x_{m_l},y_{m_l})\}$

$S_u=\{x_1, \ldots,x_{m_u}\}$

$\mathrm{argmin}_w \, ||w||^2 + C \sum_i \xi_i + C \sum_u \widehat{\xi_u}$

- $y_i \, w \cdot x_i \geq 1-\xi_i$, for all $i \in \{1, \ldots, m_l\}$
- $\widehat{y_u} w \cdot x_u \geq 1 - \widehat{\xi_u}$, for all $u \in \{1, \ldots, m_u\}$
- $\widehat{y_u} \in \{-1, 1\}$ for all $u \in \{1, \ldots, m_u\}$

Find a labeling of the unlabeled sample and w s.t. w separates both labeled and unlabeled data with maximum margin.

NP-hard

Convex after know the labels

Too many guess

Transductive SVM

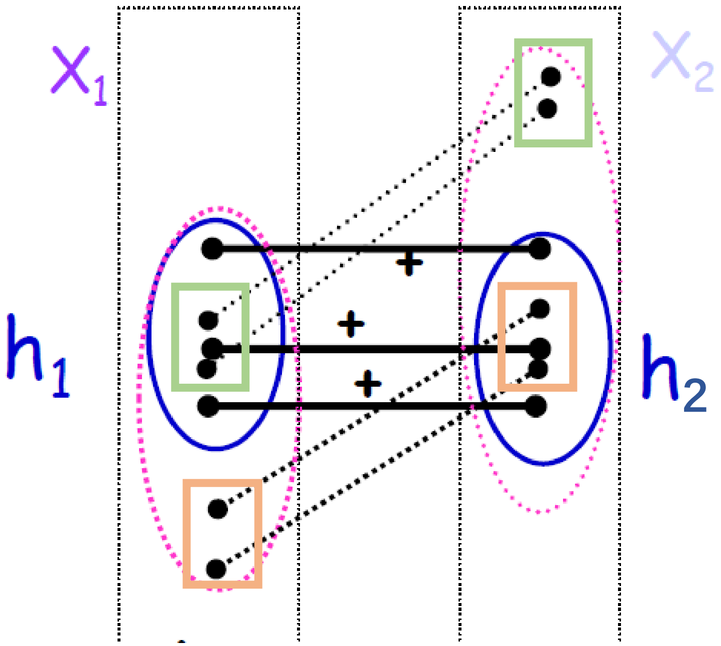
<u>Heuristic (Joachims) high level idea:</u>

- First maximize margin over the labeled points
- Use this to give initial labels to unlabeled points based on this separator.
- Try flipping labels of unlabeled points to see if doing so can increase margin

Local optimal

Co training

$x = \langle x_1, x_2 \rangle$

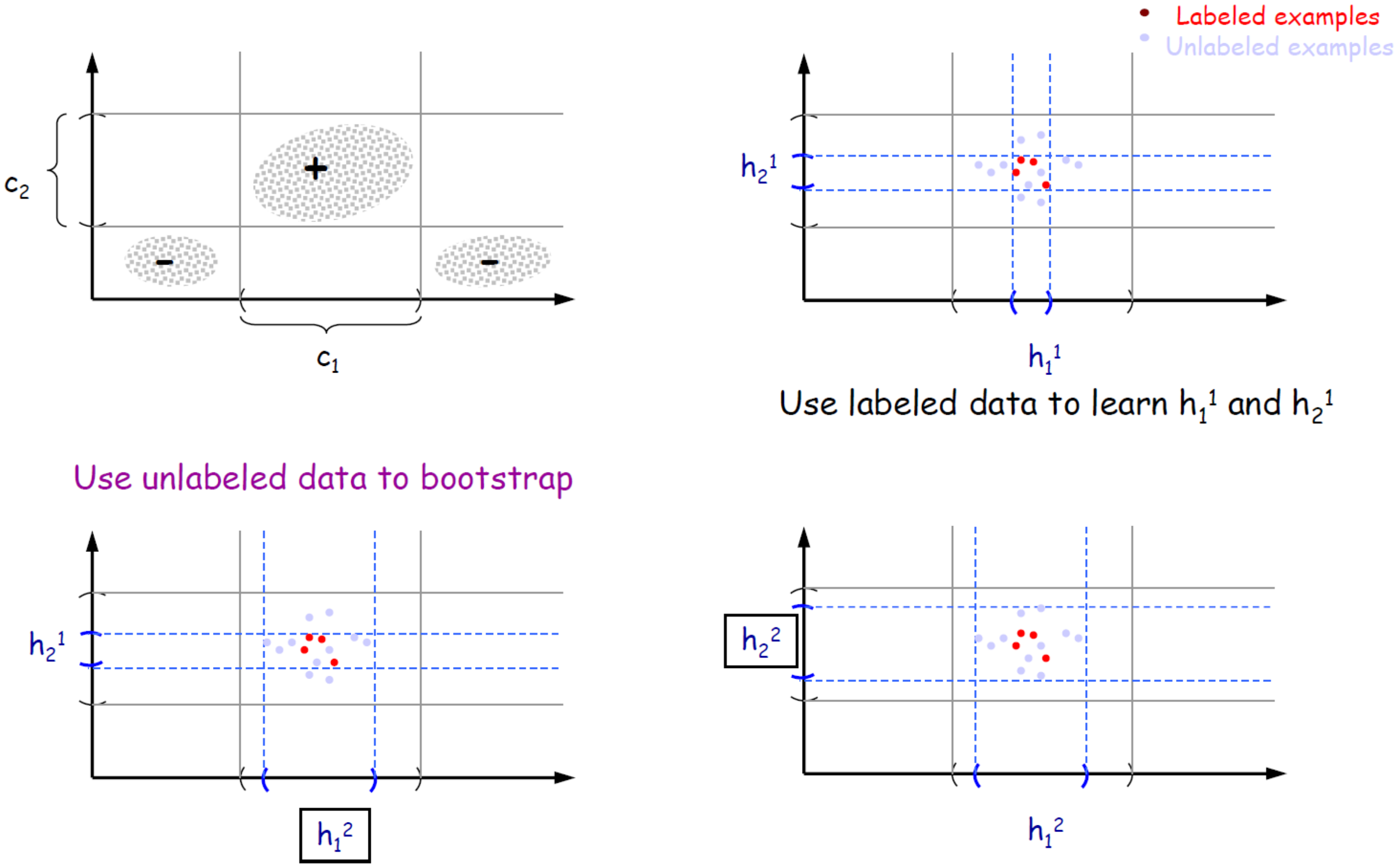


Add h1’s confident predictions to L2

Add h2’s confident predictions to L1

Remove these point from unlabeled data

Co-training

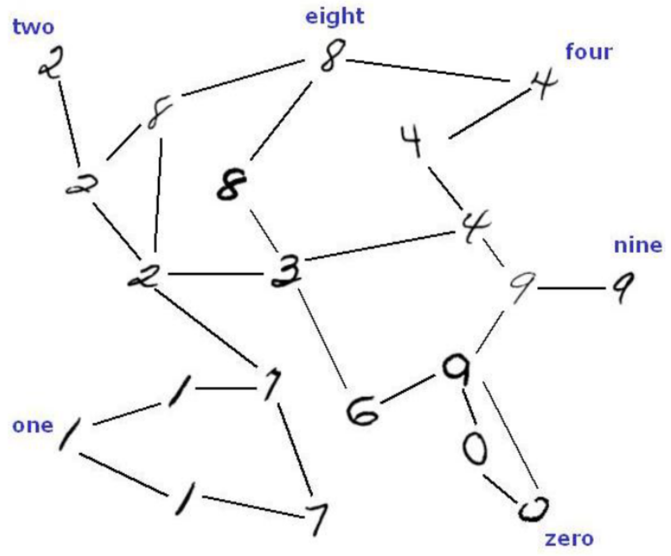


Labeled examples
Unlabeled examples
c_2
c_1
h_2^1
h_1^1
Use labeled data to learn h_1^1 and h_2^1
Use unlabeled data to bootstrap
h_2^1
h_1^2
h_2^2
h_1^2

Graph based methods

Idea: construct a graph with edges between very similar examples.

Unlabeled data can help “glue” the objects of the same class together.



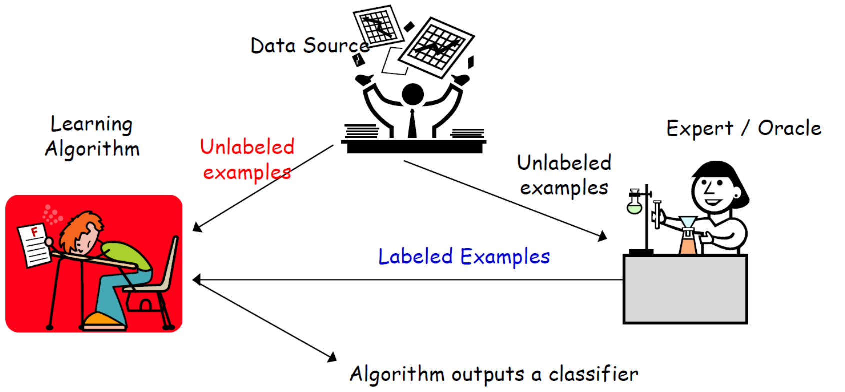
Run a graph partitioning alg to separate the graph into pieces

$$\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

Minimize $\sum_{e=(i,j)} w_e \|f_i - f_j\|^2$

Semi-supervised vs Active

Semi-Supervised Learning



$S_l=\{(x_1, y_1), \ldots, (x_{m_l}, y_{m_l})\}$

x_i drawn i.i.d from D, $y_i = c^*(x_i)$

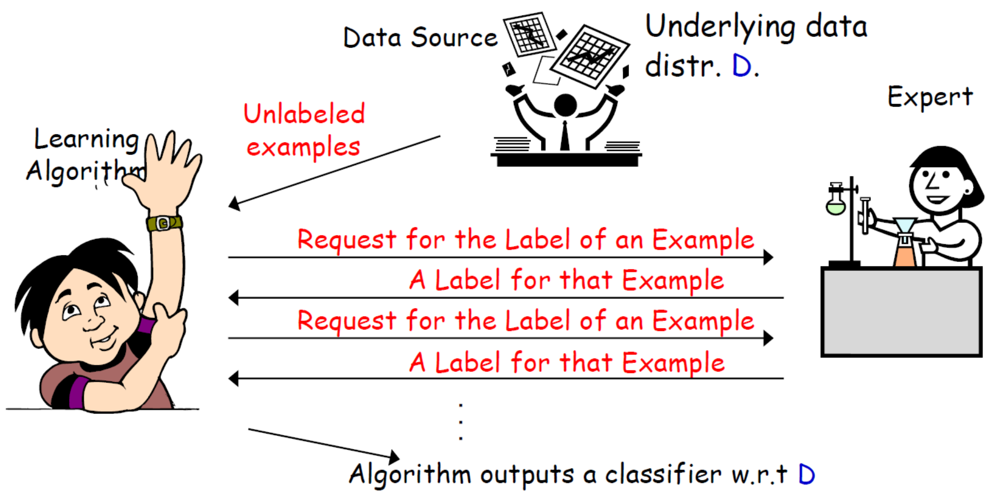
$S_u=\{x_1, \ldots, x_{m_u}\}$ drawn i.i.d from D

Goal: h has small error over D.

$$err_D(h) = \Pr_{x\sim D}(h(x) \neq c^*(x))$$

VS

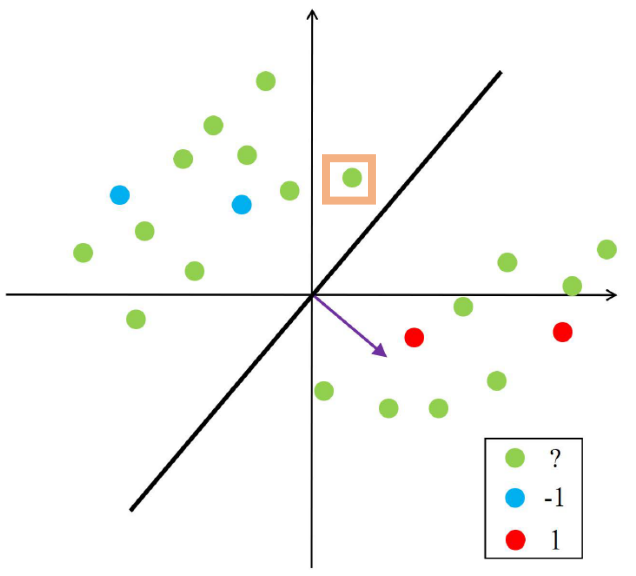
Batch Active Learning



fewer labels request

More informative labels

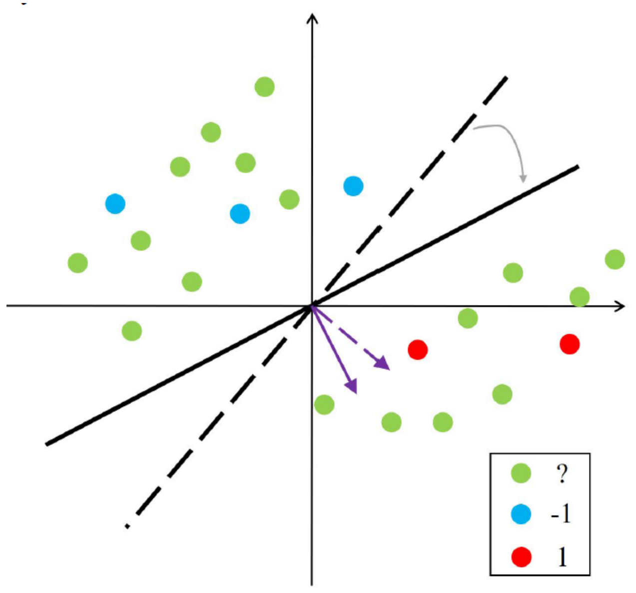
Active learning



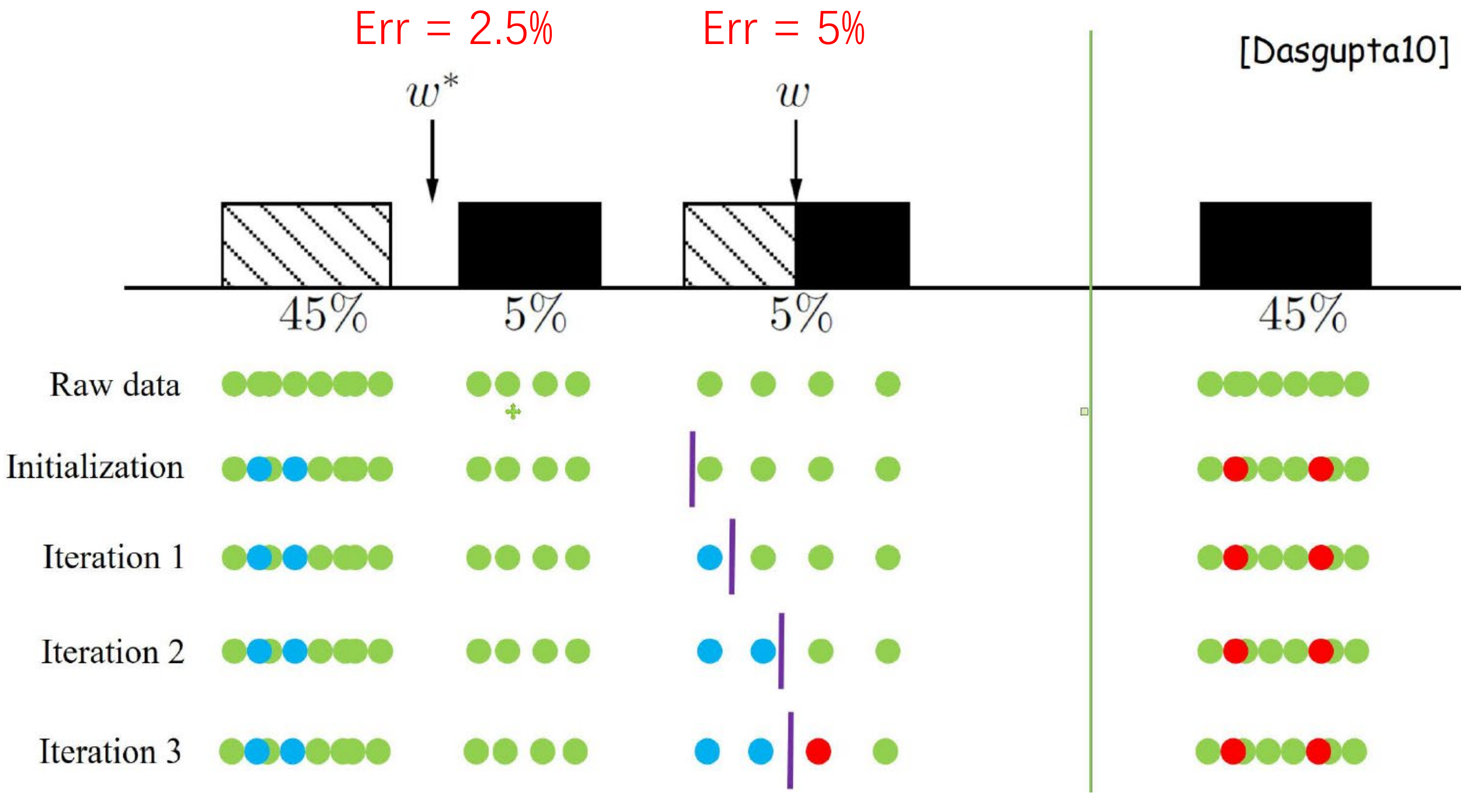
For $t = 1, \ldots,$

- Find w_t the max-margin separator of all labeled points so far.
- Request the label of the example closest to the current separator: minimizing $|x_i \cdot w_t|$.

  (highest uncertainty)



Active learning



sampling bias

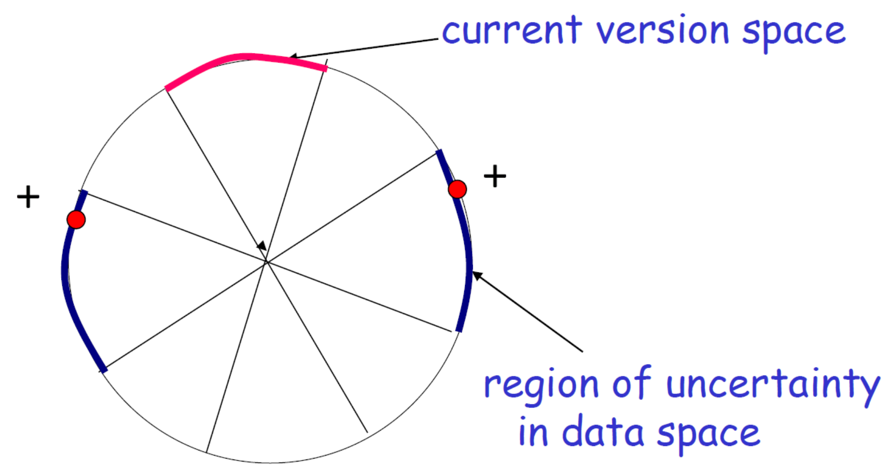
Active learning

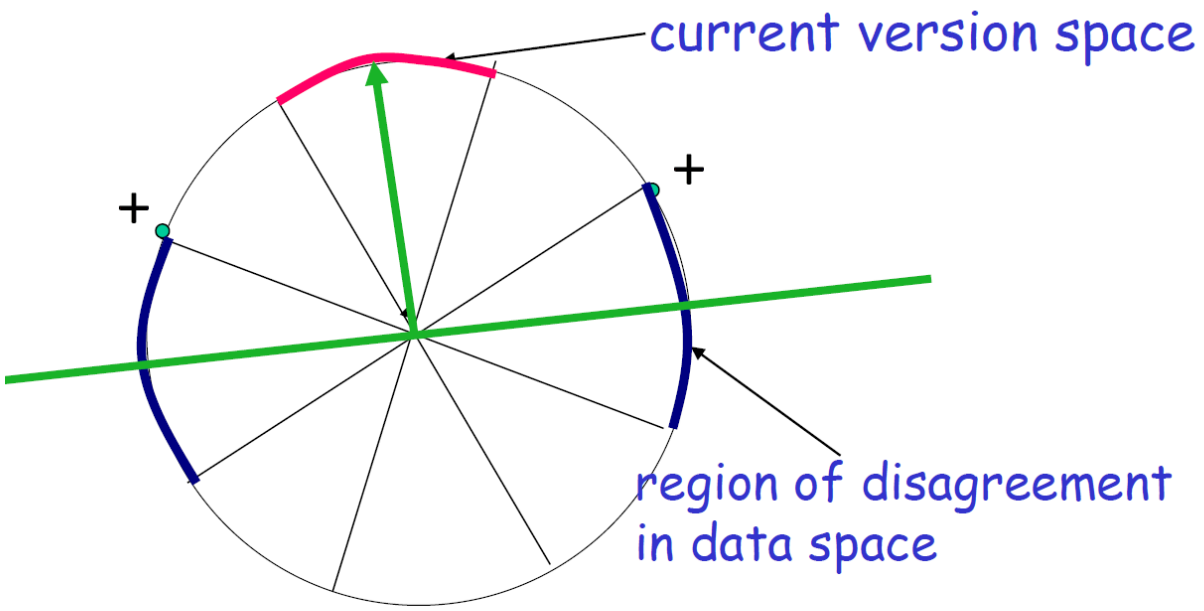
Definition (Mitchell'82) Assume realizable case: $c^* \in H$.

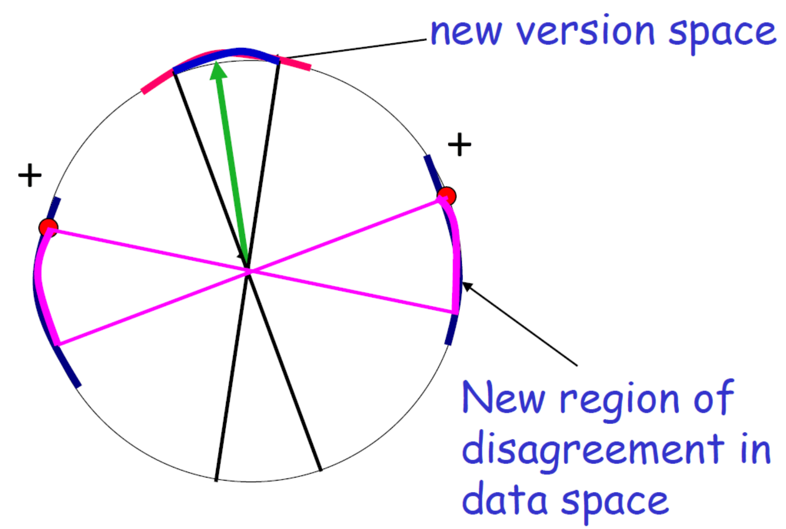
Given a set of labeled examples $(x_1, y_1), \ldots, (x_{m_l}, y_{m_l})$, $y_i = c^*(x_i)$

Version space of H: part of H consistent with labels so far.

I.e., $h \in VS(H)$ iff $h(x_i) = c^*(x_i)\ \forall i \in \{1, \ldots, m_l\}$.





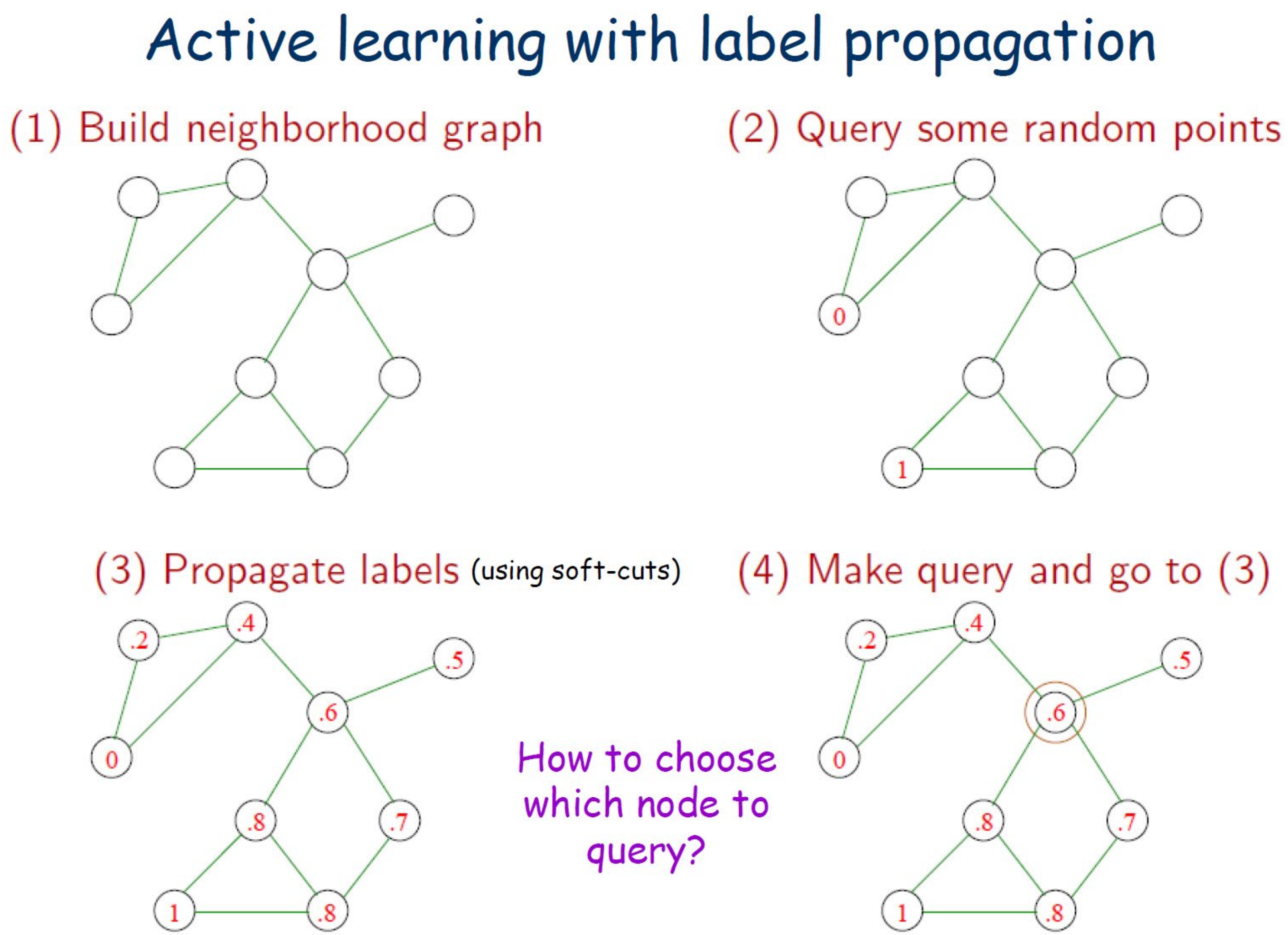


Active learning

The DHN Agnostic Active Learner [DHN'07]

$S = \emptyset$ (points with inferred labels)
$T = \emptyset$ (points with queried labels)
For $t = 1, 2, \ldots$:
    Receive x_t
    If ($h_{+1} = \texttt{learn}(S \cup \{(x_t, +1)\}, T)$) fails: Add $(x_t, -1)$ to S and break
    If ($h_{-1} = \texttt{learn}(S \cup \{(x_t, -1)\}, T)$) fails: Add $(x_t, +1)$ to S and break
    If $\mathrm{err}(h_{-1}, S \cup T) - \mathrm{err}(h_{+1}, S \cup T) > \Delta_t$: Add $(x_t, +1)$ to S and break
    If $\mathrm{err}(h_{+1}, S \cup T) - \mathrm{err}(h_{-1}, S \cup T) > \Delta_t$: Add $(x_t, -1)$ to S and break
    Request y_t and add (x_t, y_t) to T

Active learning



THANKS