Concept Learning

Outline:

- Learning boolean functions
- Most general and most specific consistent hypothesis.
- Mitchell's version space algorithm
- Probably approximately correct (PAC) learning.
- Sample complexity for PAC.
- Vapnik-Chervonenkis (VC) dimension.
- Improved sample complexity bounds.

Learning concepts

Assume objects (examples) described in terms of attributes:

Sky	Air-Temp	Humidity	Wind	Water	Forecast	EnjoySport
Sunny	Warm	Normal	Strong	Warm	Same	yes
Rainy	Cold	Normal	Strong	Warm	Change	no

Concept = a set of objects

- Concept learning:
  Given a sample of labeled objects we want to learn a boolean mapping from objects to T/F identifying an underlying concept
  - E.g. EnjoySport concept
- Concept (hypothesis) space H
  - Restriction on the boolean description of concepts

Learning concepts

- Object (instance) space X
- Concept (hypothesis) spaces H

$$H \neq X \quad !!!!$$

Assume n binary attributes (e.g. true/false, warm/cold)

- Instance space X:

  2^n different objects

- Concept space H:

  2^{2^n} possible concepts

  = all possible subsets of objects

Learning concepts

- Problem: Concept space too large
- Solution: restricted hypothesis space H
- Example: conjunctive concepts

  $(Sky = Sunny) \wedge (Weather = Cold)$

  3^n possible concepts   Why?

- Other restricted spaces:

  3-CNF (or k-CNF) $(a_1 \vee a_3 \vee a_7) \wedge (...)$

  3-DNF (or k-DNF) $(a_1 \wedge a_5 \wedge a_9) \vee (...)$

Learning concepts

- After seeing k examples the hypothesis space (even if restricted) can have many consistent concept hypotheses
- Consistent hypothesis: a concept *c* that evaluates to T on all positive examples and to F on all negatives.

- What to learn?
  - General to specific learning. Start from all true and refine with the maximal (consistent) generalization.
  - Specific to general learning. Start from all false and refine with the most restrictive specialization.
  - Version space learning. Keep all consistent hypothesis around – the combination of the above two cases.

Specific to general learning
(for conjunctive concepts)

Assume two hypotheses:

$h1 = (Sunny, ?, ? \; Strong, ?, ?)$

$h2 = (Sunny, ?, ?, ?, ?, ?)$

Then we say that:

h2 is more general than *h1,*
h1 is a special case (specialization of) h2

Specific to general learning:

- start from the all-false hypothesis $h0 = (-,-,-,-,-,-)$
- by scanning samples, gradually refine the hypothesis (make it more general) whenever it does not satisfy the new sample seen (keep the most restrictive specialization of positives)

Specific to general learning. Example

Conjunctive concepts, target is a conjunctive concept

$h = (-,-,-,-,-,-)$ All false

(Sunny,Warm, Normal, Strong, Warm, Same) T

$h = (Sunny, Warm, Normal, Strong, Warm, Same)$

(Rainy, Cold, Normal, Strong, Warm, Change) F

$h = (Sunny, Warm, Normal, Strong, Warm, Same$

(Sunny,Warm, High, Strong, Warm, Same) T

$h = (Sunny, Warm, ?, Strong, Warm, Same)$

(Sunny,Warm, High, Strong, Cool, Same) T

$h = (Sunny, Warm, ?, Strong, ?, Same)$

General to specific learning

- Dual problem to the specific to general learning
- Start from the all true hypothesis $h0 = (?,?,?,?,?,?)$
- Refine the concept description such that all samples are consistent (keep maximal possible generalization)

$h = (?,?,?,?,?,?)$

(Sunny,Warm, Normal, Strong, Warm, Same) T

$h = (?,?,?,?,?,?)$

(Sunny,Warm, High, Strong, Warm, Same) T

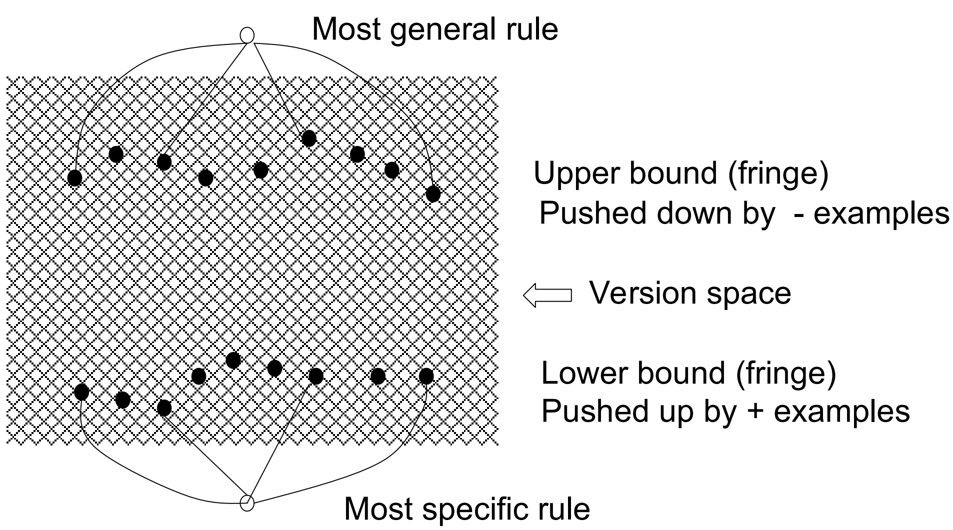
$h = (?,?,?,?,?,?)$

(Rainy, Cold, Normal, Strong, Warm, Change) F

$h = (Sunny,?,?,?,?,?),(?,Warm\,?,?,?,?),$
$(?,?,?,?,?,Same)$

Mitchell's version space algorithm

- Keeps the space of consistent hypotheses



Mitchell's version space algorithm

- Keeps and refines the fringes of the version space
- Converges to the target concept whenever the target is a member of the hypotheses space H
- Assumption:
  – No noise in the data samples (the same example has always the same label)

- The hope is that the fringe is always small

  Is this correct ?

Exponential fringe set – example

Conjunctive concepts, upper fringe (general to specific)

Samples: $(true, true, true, true, ..., true)$ T

$$\frac{1}{2}n \left\{ \begin{array}{ll} (false, false, true, true, ..., true) & F \\ (true, true, false, false, ..., true) & F \\ ... & \\ (true, true, true, ..., false, false) & F \end{array} \right.$$

Maximal generalizations – different hypotheses we need to remember

$$2^{\frac{n}{2}} \left\{ \begin{array}{l} (true, ?, true, ?, ..., true, ?) \\ (?, true, true, ?, ..., true, ?) \\ (true, ?, ?, true, ..., true, ?) \\ ... \\ (?, true, ?, true, ..., ?, true) \end{array} \right.$$

Learning concepts

- Version space algorithm may require large number of samples to converge to the target concept
  - In the worst case we must see all concepts before converging to it.
  - The samples may come from different distributions – it may take a very long time to see all examples
- The fringe can go exponential in the number of attributes
- Alternative solution: Select a hypothesis that is consistent after some number of (+, -) samples is seen by our algorithm