Introduction to Machine Learning

CSE474/574: Lecture 3

Varun Chandola <chandola@buffalo.edu>

30 Jan 2015

Outline

Outline

A Simple Algorithm (Find-S [1, Ch. 2])

1. Start with $h = \emptyset$
2. Use next input $\{x, c(x)\}$
3. *If* $c(x) = 0$, *goto* step 2
4. $h \leftarrow h \wedge x$ (*pairwise-and*)
5. *If* more examples: *Goto* step 2
6. *Stop*

Pairwise-and rules:

$$a_h \wedge a_x = \begin{cases} a_x & : \quad if \ a_h = \emptyset \\ a_x & : \quad if \ a_h = a_x \\ ? & : \quad if \ a_h \neq a_x \\ ? & : \quad if \ a_h = ? \end{cases}$$

Simple Example

Target concept

{?,large,?,?,thick}

- How many positive examples can there be?
- What is the minimum number of examples need to be seen to learn the concept?
  1. {circular,**large**,light,smooth,**thick**}, malignant
  2. {oval,**large**,dark,irregular,**thick**}, malignant
- Maximum?

Simple Example

Target concept

{?,large,?,?,thick}

- How many positive examples can there be?
- What is the minimum number of examples need to be seen to learn the concept?
  1. {circular,**large**,light,smooth,**thick**}, malignant
  2. {oval,**large**,dark,irregular,**thick**}, malignant
- Maximum?

Simple Example

Target concept

{?,large,?,?,thick}

- How many positive examples can there be?
- What is the minimum number of examples need to be seen to learn the concept?
  1. {circular,**large**,light,smooth,**thick**}, malignant
  2. {oval,**large**,dark,irregular,**thick**}, malignant
- Maximum?

Partial Training Data

1. {circular,large,light,smooth,thick}, malignant
2. {circular,large,light,irregular,thick}, malignant
3. {oval,large,dark,smooth,thin}, benign
4. {oval,large,light,irregular,thick}, malignant
5. {circular,small,light,smooth,thick}, benign

- Concept learnt:
  - {?,large,light,?,thick}
  - What mistake can this "concept" make?

Partial Training Data

1. {circular,large,light,smooth,thick}, malignant
2. {circular,large,light,irregular,thick}, malignant
3. {oval,large,dark,smooth,thin}, benign
4. {oval,large,light,irregular,thick}, malignant
5. {circular,small,light,smooth,thick}, benign

- Concept learnt:
  - {?,large,light,?,thick}
  - What mistake can this "concept" make?

Partial Training Data

1. {circular,large,light,smooth,thick}, malignant
2. {circular,large,light,irregular,thick}, malignant
3. {oval,large,dark,smooth,thin}, benign
4. {oval,large,light,irregular,thick}, malignant
5. {circular,small,light,smooth,thick}, benign

- Concept learnt:
  - {?,large,light,?,thick}
  - What mistake can this "concept" make?

Recap of Find-S

- Objective: Find *maximally specific* hypothesis
- Admit all positive examples and nothing more
- Hypothesis never becomes any more specific

Questions

- Does it converge to the target concept?
- Is the most specific hypothesis the best?
- Robustness to errors
- **Choosing best among potentially many maximally specific hypotheses**

Recap of Find-S

- Objective: Find *maximally specific* hypothesis
- Admit all positive examples and nothing more
- Hypothesis never becomes any more specific

Questions

- Does it converge to the target concept?
- Is the most specific hypothesis the best?
- Robustness to errors
- **Choosing best among potentially many maximally specific hypotheses**

Version Spaces

1. {circular,large,light,smooth,thick}, malignant
2. {circular,large,light,irregular,thick}, malignant
3. {oval,large,dark,smooth,thin}, benign
4. {oval,large,light,irregular,thick}, malignant
5. {circular,small,light,smooth,thin}, benign

- Hypothesis chosen by **Find-S**:
  - {?,large,light,?,thick}
- Other possibilities that are **consistent** with the training data?
- What is **consistency**?
- **Version space**: Set of *all* consistent hypotheses.

Version Spaces

1. {circular,large,light,smooth,thick}, malignant
2. {circular,large,light,irregular,thick}, malignant
3. {oval,large,dark,smooth,thin}, benign
4. {oval,large,light,irregular,thick}, malignant
5. {circular,small,light,smooth,thin}, benign

- Hypothesis chosen by **Find-S**:
  - {?,large,light,?,thick}
- Other possibilities that are **consistent** with the training data?
- What is **consistency**?
- **Version space**: Set of *all* consistent hypotheses.

List Then Eliminate

1. $VS \leftarrow \mathcal{H}$
2. *For Each* $\langle x, c(x) \rangle \in D$:

   Remove every hypothesis h from VS such that $h(x) \neq c(x)$
3. *Return* VS

- Issues?
- How many hypotheses are removed at every instance?

List Then Eliminate

1. $VS \leftarrow \mathcal{H}$
2. *For Each* $\langle x, c(x) \rangle \in D$:

   Remove every hypothesis h from VS such that $h(x) \neq c(x)$
3. *Return* VS

- Issues?
- How many hypotheses are removed at every instance?

Compressing Version Space

More_General_Than Relationship

$$h_j \geq_g h_k \quad if \quad h_k(x) = 1 \Rightarrow h_j(x) = 1$$

$$h_j >_g h_k \quad if \quad (h_j \geq_g h_k) \wedge (h_k \not\geq_g h_j)$$

- In a version space, there are:
  1. Maximally general hypotheses
  2. Maximally specific hypotheses
- Boundaries of the version space

References


T. M. Mitchell.
Machine Learning.
McGraw-Hill, Inc., New York, NY, USA, 1 edition, 1997.