An algebraic approach to machine learning and knowledge discovery

Zdravko Markov
Department of Computer Science, Central Connecticut State University
1615 Stanley Street, New Britain, CT 06050
E-mail: markovz@ccsu.edu

March 31, 2000

Talk outline

- Introduction to ML by example
- Lgg (infimum) based learning algorithm
- Semi-distance on semi-lattices
- Semantic semi-distance
- Conceptual clustering of Horn clauses

Papers

- Markov, Z. An algebraic approach to inductive learning, FLAIRS-2000, AAAI Press 2000, to appear.
- Markov, Z. and Marinchev, I. Metric-based inductive learning using semantic height functions, ECML-2000, LNCS, Springer, 2000, to appear.
- Monjardet, B. Metrics on Partially Ordered Sets – a Survey, *Discrete Mathematics*, Vol. 35 (1981), 173-184.

Play tennis example

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
	x_1	x_2	x_3	x_4	y
1	sunny	hot	high	weak	no
2	sunny	hot	high	strong	no
3	overcast	hot	high	weak	yes
4	rain	mild	high	weak	yes
5	rain	cool	normal	weak	yes
6	rain	cool	normal	strong	no
7	overcast	cool	normal	strong	yes
8	sunny	mild	high	weak	no
9	sunny	cool	normal	weak	yes
10	rain	mild	normal	weak	yes
11	sunny	mild	normal	strong	yes
12	overcast	mild	high	strong	yes
13	overcast	hot	normal	weak	yes
14	rain	mild	high	strong	no
15	sunny	mild	normal	weak	?

$X = (x_1, x_2, x_3, x_4)$

Instance space $= |\{X\}| = 36$

Concept learning/regression: f?, $y = f(X)$

$f(sunny, \_\,, normal, \_\,) = yes$
$f(rain, \_\,, \_\,, strong) = no$
$f(rain, \_\,, \_\,, weak) = yes$
$f(overcast, \_\,, \_\,, \_\,) = yes$
$f(sunny, \_\,, high, \_\,) = no$

$Error(f) = \frac{|\{X|f(X) \neq y\}|}{36}$

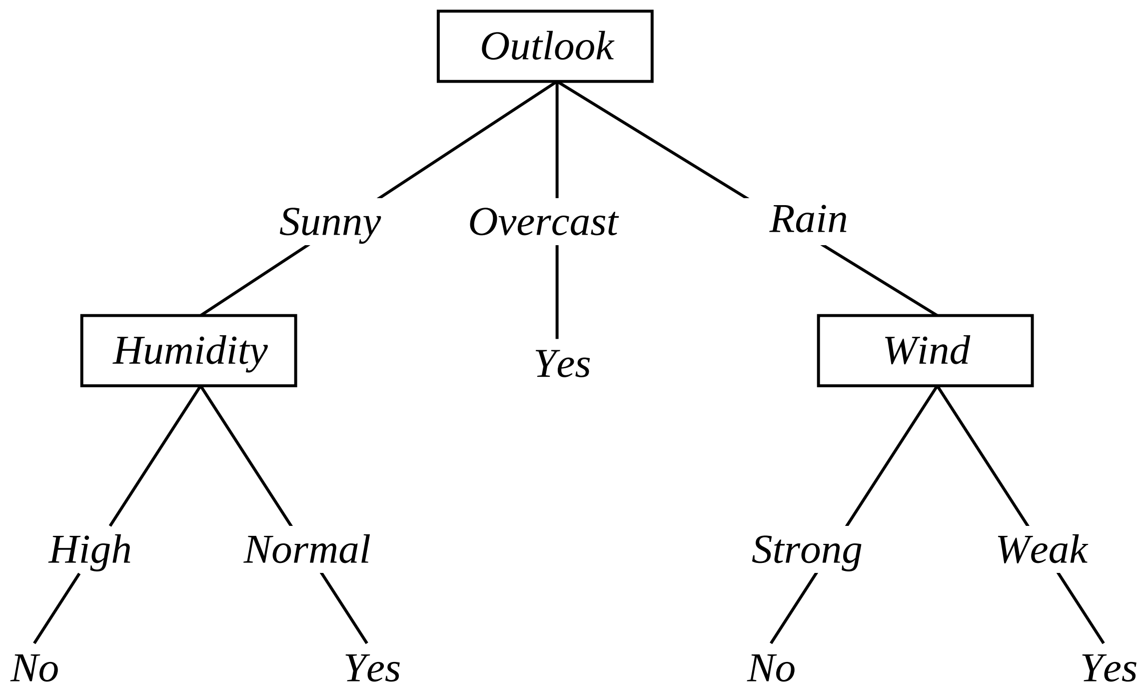
Prediction: $f(sunny, mild, normal, weak) = ?$

K-Nearest Neighbor (k-NN)

X	8	9	10	11	1	...	14
d(15,X)	1	1	1	1	2	...	3
PlayTennis	no	yes	yes	yes	no	...	no

$\Rightarrow f(sunny, mild, normal, weak) = yes$

Decision Tree for *PlayTennis*



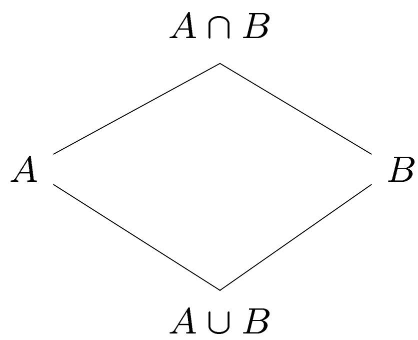
3

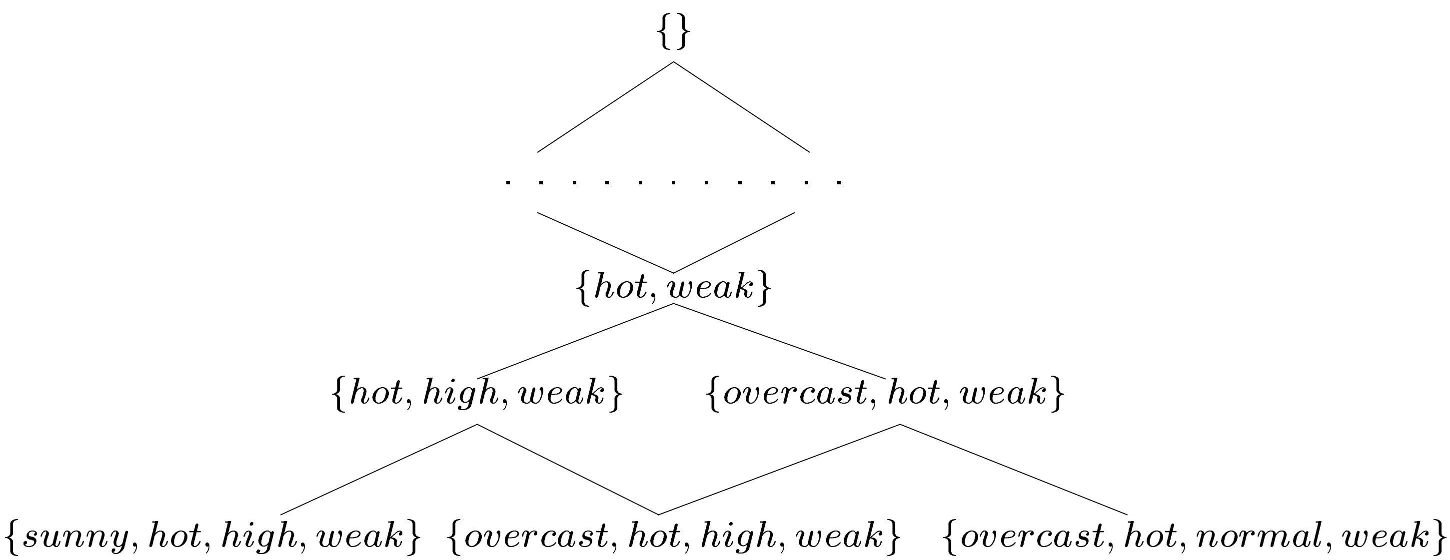
Conceptual Clustering

No.	covering	milk	homeothermic	habitat	eggs	gills
1	hair	t	t	land	f	f
2	none	t	t	sea	f	f
3	hair	t	t	sea	t	f
4	hair	t	t	air	f	f
5	scales	f	f	sea	t	t
6	scales	f	f	land	t	f
7	scales	f	f	sea	t	f
8	feathers	f	t	air	t	f
9	feathers	f	t	land	t	f
10	none	f	f	land	t	f

```
[A,B,C,D,E,F]
   [A,f,B,C,t,D]
      [A,f,B,C,t,f]
         [A,f,f,land,t,f]
            [scales,f,f,land,t,f]
            [none,f,f,land,t,f]
         [feathers,f,t,A,t,f]
            [feathers,f,t,air,t,f]
            [feathers,f,t,land,t,f]
      [scales,f,f,sea,t,A]
         [scales,f,f,sea,t,t]
         [scales,f,f,sea,t,f]
   [A,t,t,B,C,f]
      [A,t,t,sea,B,f]
         [none,t,t,sea,f,f]
         [hair,t,t,sea,t,f]
      [hair,t,t,A,f,f]
         [hair,t,t,land,f,f]
         [hair,t,t,air,f,f]
```

Least General Generalization (lgg, infimum)





Lgg-based clustering/concept learning

Given: set of examples E
Find: semi-lattice G (E is the set of all maximal elements of G)

1. Initialization: $G = E$, $M = E$;

2. If $|M| = 1$ then exit;

3. $T = \{h | h = lgg(a, b), a, b \in M\}$

4. $T = T \setminus \{h | \exists a \in C_1, \exists b \in C_2, C_1 \neq C_2, h \preceq a, h \preceq b\}$ (for concept learning only);

5. Let $h_{min} \in T$ satisfy some minimality condition min(T);

6. $M = M \setminus \{a | a \in M, h_{min} \preceq a\}$;

7. $G = G \cup \{h_{min}\}$, $M = M \cup \{h_{min}\}$;

8. go to step 2.

Example: Attribute-value language

- $a = \{outlook{=}sunny, temperature{=}hot, humidity{=}high, wind{=}weak\}$

- $a \preceq b \Leftrightarrow a \subseteq b$

- $lgg(a, b) = a \cap b$

- min(T) $= \{h | h \in T, h = lgg(a, b), a, b \in M, d(a, b) = \min_{x,y \in M} d(x, y)\}$.

- $d(a, b) = |a| - |a \cap b|$

- min(T): The *closest* elements produce the minimal lgg.

Semi-distance on join semi-lattices (size functions)

Semi-distance (Quasi-metric) $d : O \times O \rightarrow \Re$

1. $d(a, a) = 0$ and $d(a, b) \geq 0$

2. $d(a, b) = d(b, a)$

3. $d(a, b) \leq d(a, c) + d(c, b)$

Join/Meet semi-lattice. A partially ordered set $(A, \preceq)$ in which every two elements $a, b \in A$ have an infimum/supremum.

Size function $s : A \times A \rightarrow \Re$

1. $s(a, b) \geq 0, \forall a, b \in A$ and $a \preceq b$.

2. $s(a, a) = 0, \forall a \in A$.

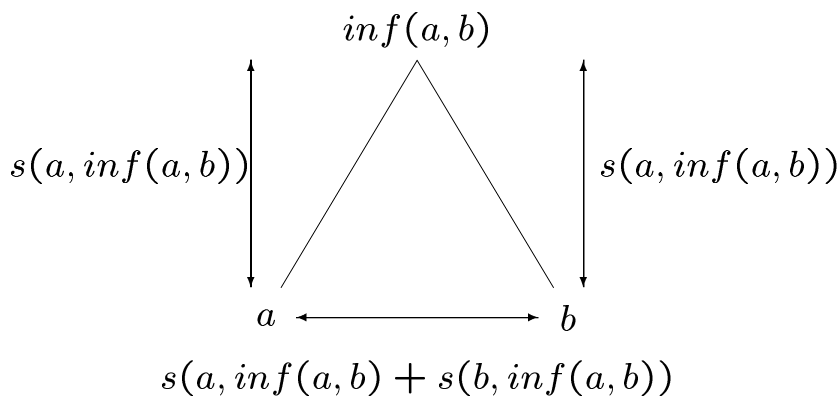
3. $\forall a, b, c \in A$, $a \preceq c$, $c \preceq b \Rightarrow s(a, b) \leq s(a, c) + s(c, b)$ and $s(c, b) \leq s(a, b)$.

4. $c = inf\{a, b\}$, $a, b \in A$, $d \in A$, $a \preceq d$ and $b \preceq d \Rightarrow s(c, a) + s(c, b) \leq s(a, d) + s(b, d)$.

Example 1 (attribute-value): $a \preceq b$ $(a \subseteq b)$, $s(a, b) = |b| - |a|$

Example 2 (first order terms): $t(X, Y) \preceq_\theta t(a, b)$, $t(X, Y)\theta = t(a, b)$, $\theta = \{X/a, Y/b\}$, $s(t(X, Y), t(a, b)) = |\theta| = 2$

Semi-distance on join semi-lattices



Theorem 1. Let $(A, \preceq)$ be a join semi-lattice and s − a size function. Then $d(a, b) = s(inf\{a, b\}, a) + s(inf\{a, b\}, b)$ is a *semi-distance* on $(A, \preceq)$.

Height function. A function h is called *height* of the elements of a partially ordered set $(A, \preceq)$ if it satisfies the following two properties:

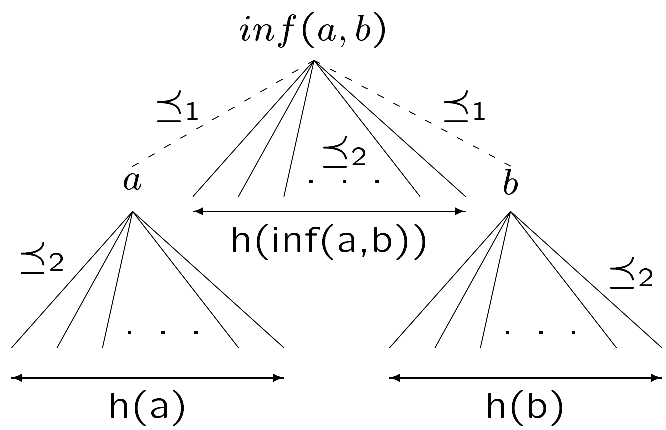
1. $\forall a, b \in A,\ a \preceq b \Rightarrow h(a) \leq h(b)$ (isotone).

2. $\forall a, b \in A,\ c = inf\{a, b\},\ d \in A,\ a \preceq d,\ b \preceq d \Rightarrow h(a) + h(b) \leq h(c) + h(d)$.

Theorem 2. Let $(A, \preceq)$ be a join semi-lattice and h be a height function. Let $s(a, b) = h(b) - h(a), \forall a \preceq b \in A$. Then s is a *size function* on $(A, \preceq)$.

Corollary 1. Let $(A, \preceq)$ be a join semi-lattice and h be a height function. Then the function $d(a, b) = h(a) + h(b) - 2h(inf\{a, b\}), \forall a, b \in A$ is a *semi-distance* on $(A, \preceq)$.

Semantic semi-distance



Set of objects A and two binary relations:

- $\preceq_1$ - *syntactic* relation, $(A, \preceq_1)$ - join semi-lattice
- $\preceq_2$ - *semantic* relation (hypothesis evaluation)

Ground elements. $GA = \{a | a \in A,\ \neg\exists b \in A : a \preceq_1 b\}$.

Ground coverage (semantics of a hypothesis).
$S_h = \{a | a \in GA,\ h \preceq_2 a\}$.

Theorem 3. If:

1. $\forall a, b \in A,\ a \preceq_1 b \Rightarrow |S_a| \geq |S_b|$
2. $\forall a, b \in A,\ c = inf\{a, b\},\ \exists d = sup\{a, b\}$ *one* of the following is true:
   - $|S_d| < |S_a|$ and $|S_d| < |S_b|$
   - $|S_d| = |S_a|$ and $|S_c| = |S_b|$
   - $|S_d| = |S_b|$ and $|S_c| = |S_a|$

Then $h(a) = x^{-|S_a|}$ is a *height function* on A, $x \in \Re$ and $x \geq 2$.

Clustering Horn clauses

A - the set of Horn clauses with same predicate at their heads:

- $\preceq_1$ = θ-subsumption, $C \preceq_\theta D \Leftrightarrow C\theta \subseteq D$
- $\preceq_2$ = *semantic entailment* (logical implication), $C \models D \Leftrightarrow$ every model if C is a model of D.