



یادگیری ماشین (Machine Learning)

یادگیری مفهوم
ترتیب کل به خاص (بخش سوم)
Concept Learning
and the General-To-Specific Ordering

دانشگاه فردوسی مشهد
دانشکده مهندسی
رضا منصفی

عناوین مورد بحث (یادگیری مفهوم)

یادگیری:

استقراء توابع کلی (*Inducing General Functions*)

از مثال‌های آموزشی خاص (*Specific Training Examples*)،

مسأله‌ی محوری یادگیری است.

بحث این درس در مورد یادگیری مفهوم

(۱) یادگیری مفهوم عبارت است از:-

اکتساب تعریف رسته‌ای عمومی از روی مثال‌های آموزشی

مثبت (عضو) و منفی (غیرعضو) داده شده در آن رسته

(*acquiring the definition of a general category given a sample of positive and negative training examples of the category*)

(۲) یادگیری مفهوم به‌عنوان مسأله‌ی جستجو

بیان فضای فرضیه‌های محتمل از پیش تعریف‌شده، جهت یافتن بهترین فرضیه منطبق (برازش بیش‌تر) بر مثال‌های آموزش

(*Concept Learning can be formulated as a problem of searching through a predefined space of potential Hypothesis for the hypothesis that best fits the training examples*)

(۳) انجام این جستجو با استفاده از ساختاری که به‌طور طبیعی در فضای فرضیه‌ها رخ می‌دهد، یعنی؛

ترتیب عام (کل) به خاص (جزء) از فرضیه‌ها (*General-to-specific ordering of hypothesis*)

(۴) ارائه الگوریتم یادگیری (*FIND-S*) و بررسی وضعیت‌های هم‌گرایی آن به فرضیه درست

(۵) بررسی ماهیت یادگیری استقرایی، و توجیه تعمیم‌پذیری (*Generalization*) موفقیت‌آمیز هر برنامه و رای داده‌های آموزشی مشاهده شده



فضای مدل (Version Space) و الگوریتم حذف کاندید (Candidate-Elimination)

نکته:

اگرچه الگوریتم **Find-S** فرضیه‌ای را از میان H برمی‌گزیند و به‌عنوان **خروجی** برمی‌گرداند

(که با مثال‌های آموزشی سازگار است)، اما این فرضیه تنها یکی از فرضیه‌هایی است که برازش مناسب با داده‌های آموزشی دارد و البته خاص‌ترین است.

معرفی:

در این گفتار ارائه دومین الگوریتم با نام **حذف کاندید (Candidate-Elimination)** و سعی در مرتفع نمودن

تعدادی از محدودیت‌های الگوریتم **Find-S** توسط این الگوریتم (**حذف کاندید**)

الگوریتم حذف کاندید: توصیفی از مجموعه‌ی تمامی فرضیه‌های سازگار با مثال‌های آموزشی به‌عنوان **خروجی**

توجه مهم: ارائه توصیفی از چنین مجموعه‌ای بدون شمارش تمامی اعضای مجموعه به‌طور صریح (**Explicitly**)

چگونه:

به کمک ترتیب جزئی (**partial ordering**) «کلی‌تر بودن» و فراهم نمودن نمایشی خلاصه‌شده

(توصیفی)، از مجموعه فرضیه‌های سازگار و به تدریج پایش نمایش آن با بررسی مثال‌های

آموزشی جدید

محدودیت: در کاربردهای واقعی الگوریتم‌های **حذف کاندید** و **Find-S**،

خوب عمل ننمودن ۲ الگوریتم فوق با داده‌های آموزشی نویزی (ضعف بسیار مهم)

نکته: البته برای اهداف این درس (یعنی **یادگیری مفهوم**) و مهم‌تر از همه، الگوریتم **حذف کاندید** **چارچوب مفهومی مفیدی** برای معرفی تعدادی از مسائل **اساسی** در **یادگیری ماشین** فراهم می‌آورد در مابقی این درس الگوریتم و مسائل مرتبط با آن مورد بررسی قرار می‌گیرد.

نکته: در ادامه، بررسی الگوریتم‌های **یادگیری (درخت‌های تصمیم)** که با داده‌های آموزشی **نویزی** بهتر عمل نموده و **بیش‌تر** به کار می‌رود، ارائه می‌شود.

تعاریف و نحوه نمایش

تعریف:

الگوریتم حذف کاندید کلیدی فرضیه‌های قابل توصیف و سازگار با داده‌های آموزشی مشاهده‌شده را پیدا می‌کند.

تعریف:

فرضیه‌ای با داده‌های آموزشی سازگار است اگر فرضیه به درستی آن مثال‌ها را دسته‌بندی کند.

سازگار (*Consistent*): گوئیم فرضیه h با مثال‌های آموزشی D سازگار است، اگر و فقط اگر به ازای کلیدی مثال‌های D ، $h(x) = c(x)$ باشد.

$$\text{Consistent}(h, D) \equiv (\forall \langle x, c(x) \rangle \in D) h(x) = c(x)$$

توجه:

به تفاوت کلیدی بین این دو که پیش‌تر گفته شد، دقت کنید.

تعریف:

برآورده‌نمودن/اقناع‌شدن (*Satisfy*)

گفته می‌شود که مثالی مانند x فرضیه h را برآورده (*Satisfy*) می‌کند \square

اگر $h(x) = 1$ باشد، بدون در نظر گرفتن این که آیا x مثالی مثبت یا منفی از مفهوم هدف است (مستقل از مفهوم هدف).

سازگاربودن/نامتناقض (*Consistent*)

در حالی که سازگاربودن چنین مثالی با h ، به مفهوم هدف وابسته است، در واقع \square

$h(x) = c(x)$ است.

الگوریتم حذف کاندید:

ارائه کلیه فرضیه‌های سازگار با مثال‌های آموزشی

فضای مدل (*Version Space*): به زیرمجموعه از کل فرضیه‌ها، فضای مدل نسبت به

□ فضای فرضیه H و

□ مثال‌های آموزشی D (*data set*)

گفته می‌شود،

زیرا شامل تمامی مدل‌های (نسخه‌های) محتمل (ممکن) از

مفهوم هدف (*target concept*) است.

تعریف فضای مدل (*Version Space*):

فضای مدل نسبت به فضای فرضیه H و مثال‌های آموزشی D ، که با

$VS_{H,D}$ نشان داده می‌شود، عبارت است از زیرمجموعه‌ای از فرضیه‌های H که با مثال‌های آموزشی موجود در D سازگار است.

$$VS_{H,D} \equiv \{h \in H \mid consistent(h, D)\}$$

توجه: راهی ساده برای نمایش فضای مدل (*Version Space*)، لیست نمودن تمامی اعضای آن است. این کار به الگوریتم یادگیری ساده با نام الگوریتم **LIST-THEN-ELIMINATE** منتهی می‌شود.

The LIST-THEN-ELIMINATE Algorithm

1. $VersionSpace \leftarrow$ a list containing every hypothesis in H
2. For each training example, $\langle x, c(x) \rangle$
remove from $VersionSpace$ any hypothesis h for which $h(x) \neq c(x)$
3. Output the list of hypotheses in $VersionSpace$

الگوریتم

❖ ابتدا تمامی فرضیه‌های H را در فضای مدل قرار بده

❖ هر فرضیه‌ای که با مثال‌های آموزشی سازگار نیست، حذف کن

❖ تا زمانی که (در حالت ایده‌آل) تک فرضیه‌ای (احتمالاً مفهوم هدف مطلوب) باقی بماند

□ اگر داده کافی موجود نباشد تا فضای مدل را به تک فرضیه‌ی (مفهوم هدف) کاهش دهد،

✓ آن وقت الگوریتم می‌تواند تمامی مجموعه فرضیه‌های سازگار با داده‌های مشاهده‌شده را به‌عنوان خروجی برگرداند

□ به‌طور کلی الگوریتم **LIST-THEN-ELIMINATE** هر جا که فضای فرضیه H محدود باشد، قابل استفاده است

□ این الگوریتم تضمین می‌کند که تمام فرضیه‌های سازگار با داده‌های آموزشی را به‌عنوان خروجی برگرداند

الگوریتم فوق نیاز دارد که تمام فرضیه‌های موجود در H را برشمرد، این امر متأسفانه برای بسیاری از

فضاهای فرضیه‌ها، شدنی نیست!!!

الگوریتم حذف کاندید نمایش فشرده تری از فضای مدل

توجه: الگوریتم حذف کاندید مشابه اصول و قواعد الگوریتم قبلی (*LIST-THEN-ELIMINATE*) کار می کند.

توجه: اما الگوریتم فوق بازنمونی بسیار ساده تری (فشرده تری) برای فضای مدل به کار می گیرد. در واقع فضای مدل توسط خاص ترین (اختصاصی ترین) و عام ترین (کلی ترین) عضوهایش نمایش داده می شود.

نکته: این اعضاء مجموعه های کرانی خاص و عام محدوده فضای مدل را در فضای فرضیه (مرتب شده ی جزئی) معین می کند.

Version Space: فضای مدل نسبت به فضای فرضیه H و مثال های آموزشی D عبارت است از زیرمجموعه ای از فرضیه های H که با مثال های آموزشی موجود در D سازگار است (شامل تمامی مدل های ممکن از مفهوم هدف).

$$VS_{H,D} \equiv \{h \in H \mid \text{Consistent}(h, D)\}$$

حد کلی (کران عام) G : نسبت به فضای فرضیه H و داده های آموزشی D

عبارت است از مجموعه ای از اعضاء حداکثر کلی (عام) از H که با D سازگارند.

حد اختصاصی (کران خاص) S : نسبت به فضای فرضیه H و داده های آموزشی D

برابر است با مجموعه اعضاء حداقل کلی (حداکثر خاص) از H که با D سازگارند.

$$G \equiv \{g \in H \mid \text{Consistent}(g, D) \wedge (\neg \exists g' \in H)[(g' >_g g) \wedge \text{Consistent}(g', D)]\}$$

$$S \equiv \{s \in H \mid \text{Consistent}(s, D) \wedge (\neg \exists s' \in H)[(s >_g s') \wedge \text{Consistent}(s', D)]\}$$

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

فضای مدل
(حذف کاندید)

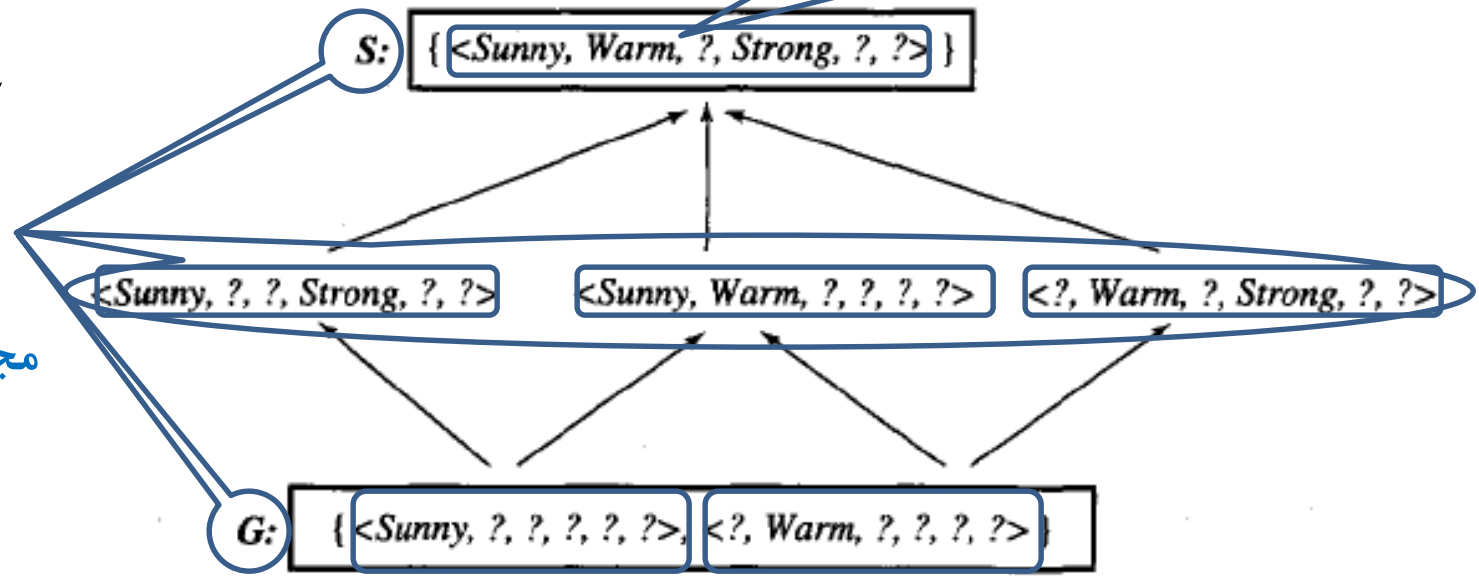
Find-s

$x_1 = \langle \text{Sunny Warm Normal Strong Warm Same} \rangle, +$
 $x_2 = \langle \text{Sunny Warm High Strong Warm Same} \rangle, +$
 $x_3 = \langle \text{Rainy Cold High Strong Warm Change} \rangle, -$
 $x_4 = \langle \text{Sunny Warm High Strong Cool Change} \rangle, +$

$h_0 = \langle \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset \rangle$
 $h_1 = \langle \text{Sunny Warm Normal Strong Warm Same} \rangle$
 $h_2 = \langle \text{Sunny Warm ? Strong Warm Same} \rangle$
 $h_3 = \langle \text{Sunny Warm ? Strong Warm Same} \rangle$
 $h_4 = \langle \text{Sunny Warm ? Strong ? ?} \rangle$



6 فرضیه
که تشکیل دهنده
فضای مدل
(Version Space)
نسبت به این
مجموعه داده‌ها و
این نوع نمایش
فرضیه‌ها



More general than

الگوریتم یادگیری حذف کاندید و ویژگی‌ها آن

ویژگی: ارائه توصیفی در مورد مجموعه‌ی همه فرضیه‌های سازگار با مثال‌های آموزشی، بدون مشخص نمودن تک‌تک اعضای آن

توجه: الگوریتم حذف کاندید فضای مدل را پیدا می‌کند که شامل تمامی فرضیه‌هایی از H است که با دنباله‌ی مشاهده شده از مثال‌های آموزشی سازگار است.

الگوریتم: با مقداردهی اولیه‌ی فضای مدل با مجموعه‌ی همه فرضیه‌های H آغاز می‌شود. یعنی با مقداردهی کران عام G_0 با عام‌ترین فرضیه در H (کف مطلق) و $G_0 \leftarrow \{ \{?, ?, ?, ?, ?, ?\} \}$ مقداردهی کران خاص S_0 با خاص‌ترین فرضیه در H (سقف مطلق) $S_0 \leftarrow \{ \{\emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset\} \}$ این ۲ مجموعه‌ی کرانی، کل فضای فرضیه را محدود و معین می‌کند، زیرا مابقی فرضیه‌های موجود در H ، کلی‌تر از S_0 و خاص‌تر از G_0 است.

توجه: با واریسی هر مثال آموزشی، مجموعه‌های کرانی S_0 و G_0 به‌مرور تعمیم‌یافته‌تر و اختصاصی‌تر می‌شود، متعاقباً فرضیه‌هایی که با مثال‌های جدید مشاهده‌شده، ناسازگار است از فضای مدل حذف می‌شود.

توجه: بعد از بررسی تمامی مثال‌های آموزشی، فضای مدل حاصل، حاوی تمامی فرضیه‌های سازگار با این مثال‌ها و فقط همین فرضیه‌ها خواهد بود.

الگوریتم حذف کاندید

Initialize G to the set of maximally general hypotheses in H

Initialize S to the set of maximally specific hypotheses in H

For each training example d , do

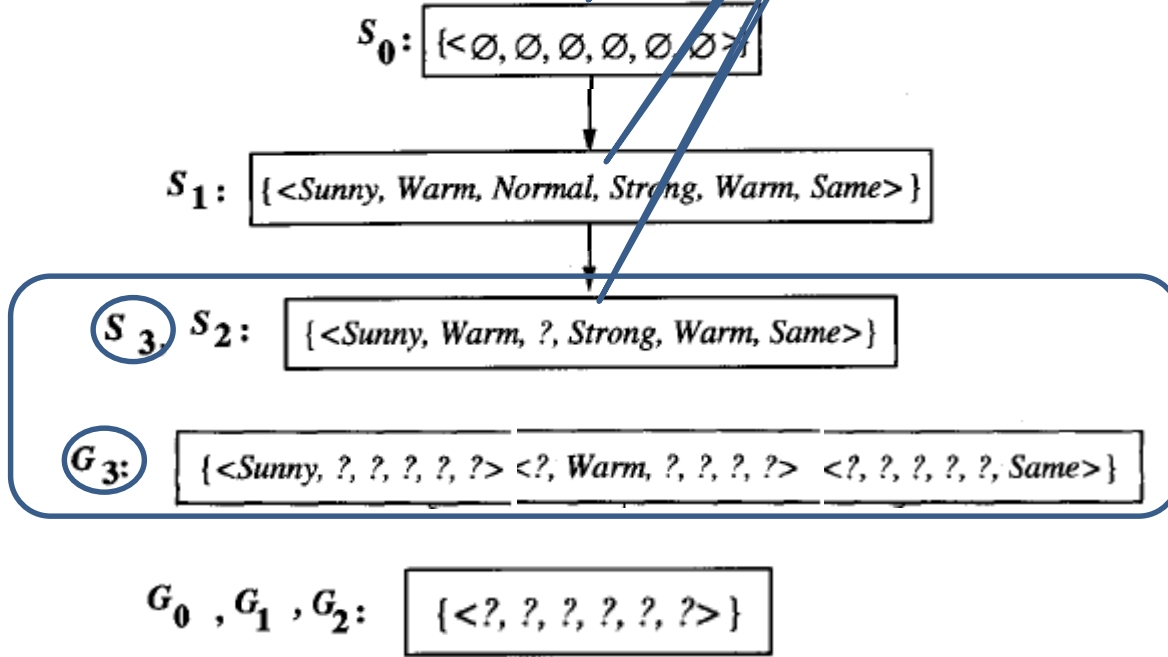
- If d is a positive example
 - Remove from G any hypothesis inconsistent with d
 - For each hypothesis s in S that is not consistent with d
 - Remove s from S
 - Add to S all minimal generalizations h of s such that
 - h is consistent with d , and some member of G is more general than h
 - Remove from S any hypothesis that is more general than another hypothesis in S
- If d is a negative example
 - Remove from S any hypothesis inconsistent with d
 - For each hypothesis g in G that is not consistent with d
 - Remove g from G
 - Add to G all minimal specializations h of g such that
 - h is consistent with d , and some member of S is more specific than h
 - Remove from G any hypothesis that is less general than another hypothesis in G

تکلیف: کد این الگوریتم

در متلب نوشته شود و با داده‌های آموزشی موجود، آموزش داده شود.

$x_1 = \langle \text{Sunny Warm Normal Strong Warm Same} \rangle, +$
 $x_2 = \langle \text{Sunny Warm High Strong Warm Same} \rangle, +$
 $x_3 = \langle \text{Rainy Cold High Strong Warm Change} \rangle, -$
 $x_4 = \langle \text{Sunny Warm High Strong Cool Change} \rangle, +$

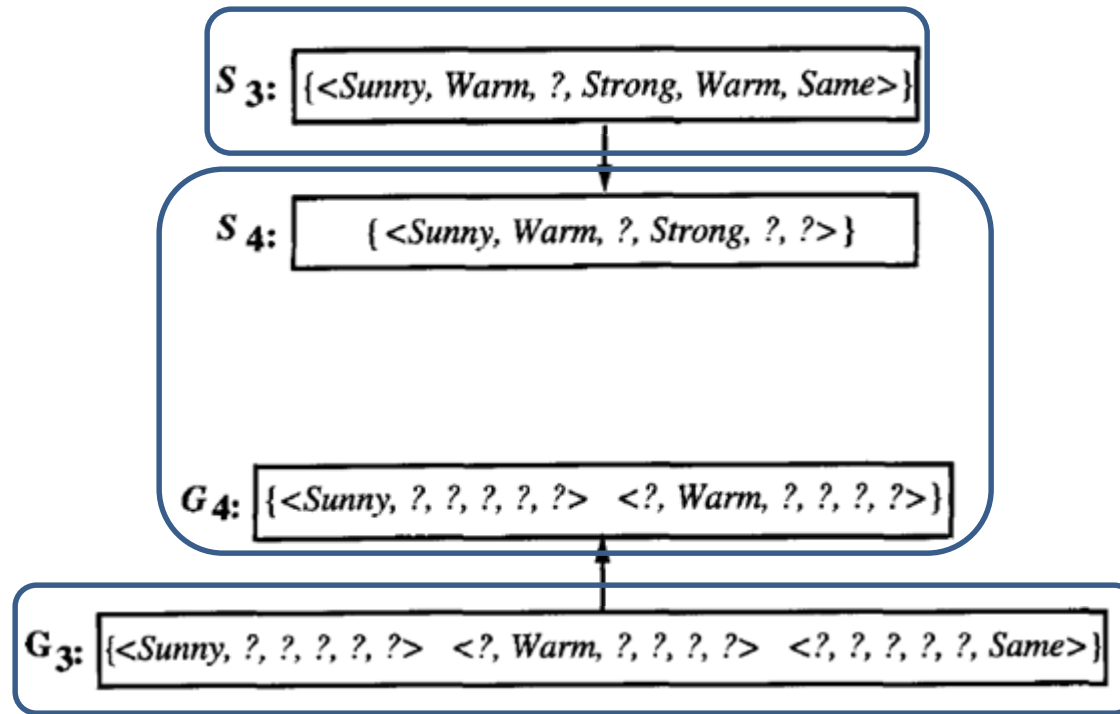
$h_0 = \langle \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset \rangle$
 $h_1 = \langle \text{Sunny Warm Normal Strong Warm Same} \rangle$
 $h_2 = \langle \text{Sunny Warm ? Strong Warm Same} \rangle$
 $h_3 = \langle \text{Sunny Warm ? Strong Warm Same} \rangle$
 $h_4 = \langle \text{Sunny Warm ? Strong ? ?} \rangle$



Training examples:

1. $\langle \text{Sunny, Warm, Normal, Strong, Warm, Same} \rangle, \text{Enjoy Sport} = \text{Yes} \quad +$
2. $\langle \text{Sunny, Warm, High, Strong, Warm, Same} \rangle, \text{Enjoy Sport} = \text{Yes} \quad +$
3. $\langle \text{Rainy, Cold, High, Strong, Warm, Change} \rangle, \text{Enjoy Sport} = \text{No} \quad -$

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes



Training examples:

1. $\langle \text{Sunny, Warm, Normal, Strong, Warm, Same} \rangle, \text{Enjoy Sport} = \text{Yes}$ +
2. $\langle \text{Sunny, Warm, High, Strong, Warm, Same} \rangle, \text{Enjoy Sport} = \text{Yes}$ +
3. $\langle \text{Rainy, Cold, High, Strong, Warm, Change} \rangle, \text{EnjoySport} = \text{No}$ -
4. $\langle \text{Sunny, Warm, High, Strong, Cool, Change} \rangle, \text{EnjoySport} = \text{Yes}$ +

مثال (جواب نهایی)

$S_4: \{ \langle \text{Sunny, Warm, ?, Strong, ?, ?} \rangle \}$

$\langle \text{Sunny, ?, ?, Strong, ?, ?} \rangle$

$\langle \text{Sunny, Warm, ?, ?, ?, ?} \rangle$

$\langle \text{?, Warm, ?, Strong, ?, ?} \rangle$

$G_4: \{ \langle \text{Sunny, ?, ?, ?, ?, ?} \rangle$

$\langle \text{?, Warm, ?, ?, ?, ?} \rangle \}$

Training examples:

1. $\langle \text{Sunny, Warm, Normal, Strong, Warm, Same} \rangle, \text{Enjoy Sport} = \text{Yes} \quad +$
2. $\langle \text{Sunny, Warm, High, Strong, Warm, Same} \rangle, \text{Enjoy Sport} = \text{Yes} \quad +$
3. $\langle \text{Rainy, Cold, High, Strong, Warm, Change} \rangle, \text{EnjoySport} = \text{No} \quad -$
4. $\langle \text{Sunny, Warm, High, Strong, Cool, Change} \rangle, \text{EnjoySport} = \text{Yes} \quad +$

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

$S_0 = \langle \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset \rangle$
 $S_1 = \langle \text{sunny, warm, normal, strong, warm, same} \rangle$
 $S_2 = \langle \text{sunny, warm, ?, strong, warm, same} \rangle$
 $S_3 = S_2$

$S_4 = \langle \text{sunny, warm, ?, strong, ?, ?} \rangle$

$\langle \text{sunny, ?, ?, strong, ?, ?} \rangle$

$\langle \text{sunny, warm, ?, ?, ?, ?} \rangle$

$\langle \text{?, warm, ?, strong, ?, ?} \rangle$

$G_4 = \langle \text{sunny, ?, ?, ?, ?, ?} \rangle \langle \text{?, warm, ?, ?, ?, ?} \rangle$

$G_3 = \langle \text{?, ?, ?, ?, ?, same} \rangle \langle \text{sunny, ?, ?, ?, ?, ?} \rangle \langle \text{?, warm, ?, ?, ?, ?} \rangle$
 $G_2 = G_1$
 $G_1 = G_0$
 $G_0 = \langle \text{?, ?, ?, ?, ?, ?} \rangle$

خاتمه ای این بخش بوده

