

# Machine Learning

**By: Dr. Omid Mahdi Ebadati**

Ph.D. (Computer Science), Delhi



Kharazmi University

Kharazmi University

# **Machine Learning Applications**

- 1. Association**
- 2. Supervised Learning**
  - 2.1 Classification
  - 2.2 Regression
- 3. Unsupervised Learning**
- 4. Reinforcement Learning**

# دسته بندی یادگیری

## ○ یادگیری با ناظر:

یک مجموعه از مثالهای یادگیری وجود دارد بازای هر ورودی، مقدار خروجی و یا تابع مربوطه نیز مشخص است. هدف سیستم یادگیر بدست آوردن فرضیه ای است که تابع و یا رابطه بین ورودی و یا خروجی را حدس بزند.

## ○ یادگیری بدون ناظر:

یک مجموعه از مثالهای یادگیری وجود دارد که در آن فقط مقدار ورودی ها مشخص است و اطلاعاتی در مورد خروجی صحیح در دست نیست. یادگیری بدون ناظر برای دسته بندی ورودیها و یا پیش بینی مقدار بعدی بر اساس موقعیت فعلی بکار میرود.

# دسته بندی یادگیری (ادامه...)

## ○ یادگیری تقویتی:

مثالها بصورت ورودی/خروجی نیستند بلکه بصورت وضعیت/پاداش هستند که یادگیر در وضعیت های مختلف عملیات مختلفی را انجام داده و پاداشهای متفاوتی دریافت و بر اساس مجموع پاداش های دریافتی عمل متناسب با هر وضعیت را یاد میگیرد.

## ○ یادگیری نیمه نظارتی:

مثالها طوری هستند که برای تعداد کمی از آنها مقدار خروجی موجود است اما برای مثالهای زیادی مقدار خروجی مشخص نیست.

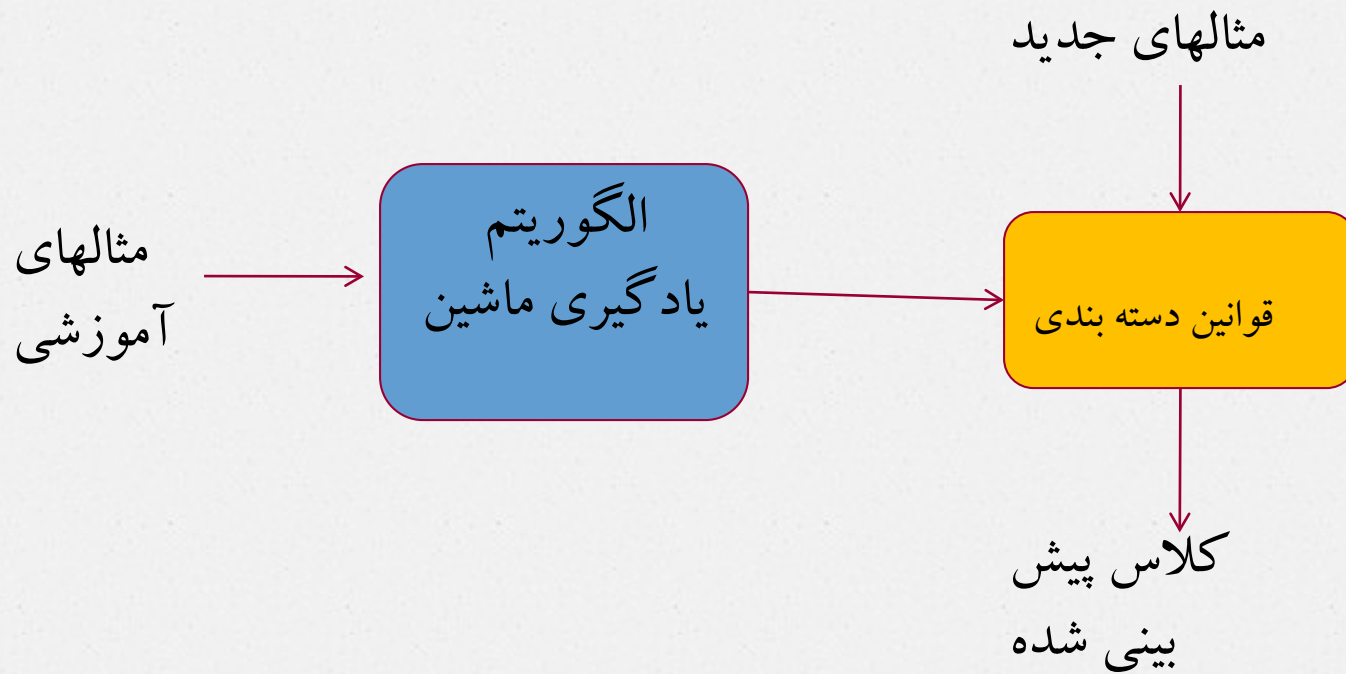
# 1. Learning Associations

- o **Basket analysis:**

$P(Y|X)$  probability that somebody who buys  $X$  also buys  $Y$  where  $X$  and  $Y$  are products/services.

Example:  $P(\text{chips} | \text{beer}) = 0.7$

# یادگیری با ناظر



# مثالی از یادگیری با ناظر

یک دیتا بیس شامل قیمت ۵۰ خانه و مساحت خانه ها وجود دارد چگونه میتوان نرخ خانه ها را بر اساس تابعی از اندازه آنها یاد گرفت؟

قیمت	مساحت
۳۷	۵۶۰
۷۹	۱۰۱۲
۷۶	۸۹۳
۱۳۰	۲۱۹۶
۸۲	۹۳۶

# مثالی از یادگیری با ناظر

- Input feature:  $x(i)$  = مساحت خانه
- Output feature:  $y(i)$  = قیمت خانه

$\langle x(i), y(i) \rangle$  مثال یادگیری

$\{ x(i), y(i), i=1:m \}$  مجموعه یادگیری

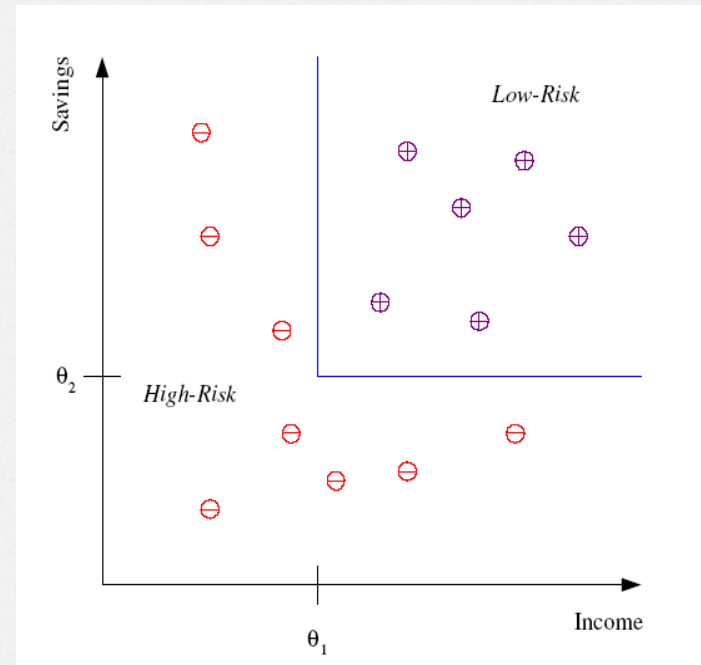
عمل یادگیری:

با داشتن یک مجموعه یادگیری می‌خواهیم تابعی بصورت  $x \rightarrow y$  یاد بگیریم که  $h(x)$  بتواند مقدار  $y$  را بخوبی حدس بزند. این تابع فرضیه و یا **hypothesis** نامیده میشود.



## 2.1. Classification

- Example: Credit scoring
- Differentiating between **low-risk** and **high-risk** customers from their *income* and *savings*



**Discriminant:** IF  $income > \theta_1$  AND  $savings > \theta_2$   
THEN **low-risk** ELSE **high-risk**

# دسته بندی

دسته بندی (Classification):

وقتی که  $y$  بتواند تعداد محدودی مقدار گسسته  
**Discrete Value** بگیرد مسئله یادگیری یک مسئله دسته  
بندی خواهد بود.

مثلا: آیا خانه مورد نظر یک آپارتمان است؟

# 2.1.1. Classification: Applications

- o Aka Pattern recognition
- o **Face recognition:** Pose, lighting, occlusion (glasses, beard), make-up, hair style
- o **Character recognition:** Different handwriting styles.
- o **Speech recognition:** Temporal dependency.
  - o Use of a dictionary or the syntax of the language.
  - o Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- o **Medical diagnosis:** From symptoms to illnesses
- o ...

# Face Recognition

Training examples of a person

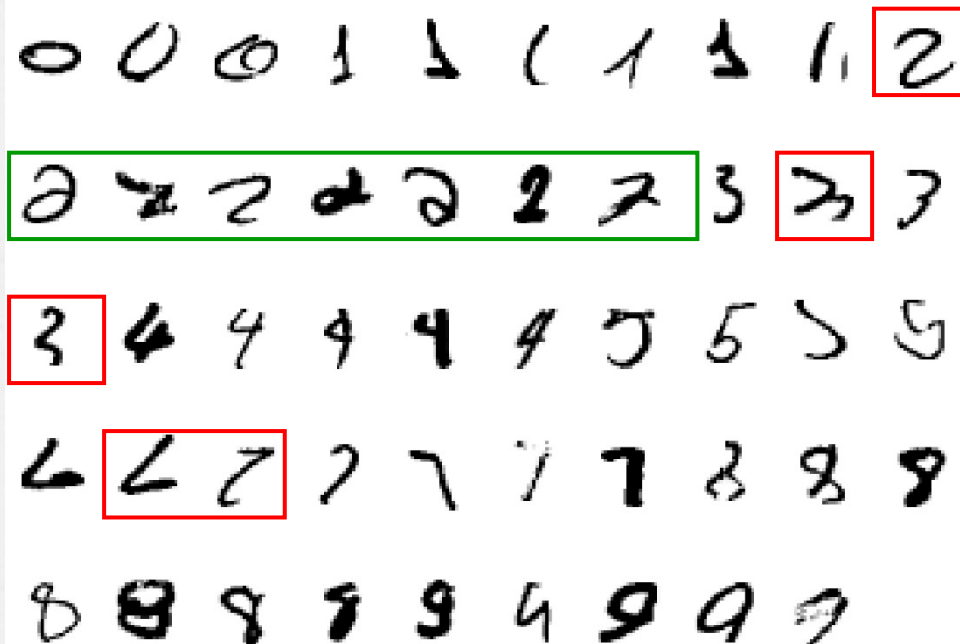


Test images



# Character recognition

- o A classic example of a task that requires machine learning: **It is very hard to say what makes a 2**



## 2.2 Regression / بازگشتی

o Example: Price of a used car

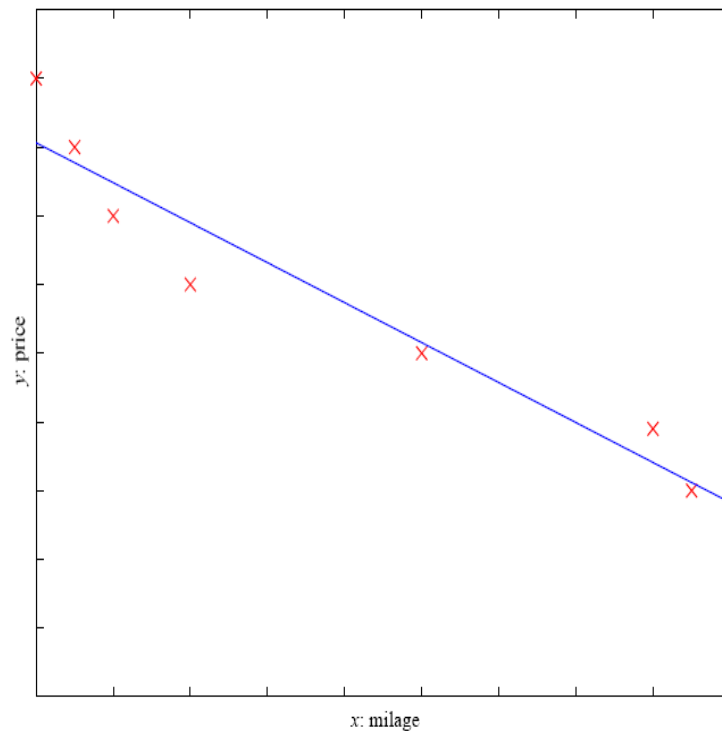
o  $x$  : car attributes

$y$  : price

$$y = g(x | \theta)$$

$g()$  model,

$\theta$  parameters



# رگرسیون

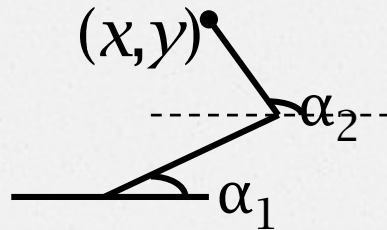
○ رگرسیون:

وقتی که تابع هدف بصورت پیوسته باشد مسئله یادگیری یک مسئله رگرسیون خواهد بود.

مثلا: یادگیری رابطه قیمت و مساحت خانه ها

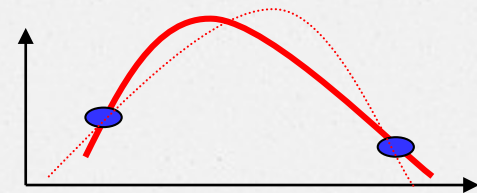
# Regression Applications

- Navigating a car: Angle of the steering wheel (CMU NavLab)
- Kinematics of a robot arm



$$\alpha_1 = g_1(x, y)$$

$$\alpha_2 = g_2(x, y)$$



- Response surface design



# Supervised Learning: Uses

- o **Prediction of future cases:** Use the rule to predict the output for future inputs
- o **Knowledge extraction:** The rule is easy to understand
- o **Compression:** The rule is simpler than the data it explains
- o **Outlier detection:** Exceptions that are not covered by the rule, e.g., fraud

# 3. Unsupervised Learning

- o Learning “what normally happens”
- o No output
- o **Clustering: Grouping similar instances**
- o Example applications
  - o Customer segmentation in CRM
  - o Image compression: Color quantization
  - o Bioinformatics: Learning motifs

# 4. Reinforcement Learning

- o Learning a policy: A **sequence of outputs**
- o No supervised output but delayed reward
- o Credit assignment problem
- o Game playing
- o Robot in a maze
- o Multiple agents, partial observability, ...

# برخی کاربردهای موفق یادگیری ماشین

- شناسائی الگو
- شناسائی چهره و حالات آن
- شناسائی حروف دست نویس
- شناسائی گفتار
- تولید الگو
- تولید الگو عکس یا حرکتهای متوالی
- شناسائی رفتارهای نادرست
- تشخیص خرابی سیگنالهای سنسورها
- تشخیص سوء استفاده از کارت های اعتباری
- پیش بینی
- قیمت سهام
- پیش بینی قیمت ارز

# Some web-based examples of machine learning

- The web contains a lot of data. **Tasks with very big datasets often use machine learning**
  - **especially if** the data is noisy or non-stationary.
- **Spam filtering, fraud detection:**
  - **The enemy adapts, so we must adapt** too.
- Recommendation systems:
  - Lots of noisy data. Million dollar prize!
- Information retrieval:
  - **Find documents or images with similar** content.
- Data Visualization:
  - **Display a huge database in a revealing** way

# تکنیک‌های مختلف یادگیری

## ○ یادگیری استنتاجی Inductive

یادگیری بر مبنای مثالهای متعدد انجام میشود. مثل درخت های تصمیم

## ○ یادگیری Connectionist

یادگیری بر مبنای مدل مغز بشر صورت میپذیرد. مثل شبکه های عصبی مصنوعی

## ○ یادگیری Bayesian

فرضیه های مختلفی در مورد داده ارائه میشود.

## ○ یادگیری Reinforcement

از سنسورها و تجربه در محیط استفاده میشود.

## ○ یادگیری Evolutionary

مثل الگوریتم ژنتیک



**Concept  
Learning**

# مفهوم چیست؟

○ دسته بندی مغز برای اشیاء، حوادث، یا ایده ها که دارای مجموعه مشترکی از ویژگیها هستند.

○ مفاهیم ما را در دسته بندی اشیاء و رویدادها کمک می کنند

○ مثال: مفهوم کتاب چیست؟

○ مسلماً معادل دانستن کتاب با مجموعه ای از کاغذ گویای مفهوم کتاب نیست. مثلاً یک دسته کاغذ باید دارای جلد هم باشند. اما این نیز کافی نیست.

○ بسیاری از اسامی که ما بکار میبریم نظیر خانه، کشور، کودک و غیره برای بیان مفاهیم است.

○ برای یادگیری یک مفهوم باید بر ویژگیهای مشترک تاکید کرده و ویژگیهای نامربوط حذف شود



# یادگیری مفهوم در انسان

○ مکتب فکری (Clark Hall) :

تئوری نسبت دادن محرک - پاسخ (Stimulus-response association):  
برطبق این تئوری ما یک مفهوم را به مجموعه ای از محرکها که معرف آن هستند نسبت می دهیم

سگ : پوزه - چهارپا - دم - صدای خاص - ....

○ مکتب فکری (Rosh (1978) :

مفاهیم طبیعی روزمره از طریق مثالهای متعدد یاد گرفته می شوند و نه از طریق قوانین

سگ : با دیدن تعداد زیادی سگ

در طول سالیان متمادی ما محرک های مختلفی دریافت میکنیم که نحوه پاسخ ما به آنها بستگی به علائق، باورها، ارزش ها و تجربیات ما دارد. ایجاد مفهوم نوعی از تفکر است که ما را در شناخت بهتر دنیای اطرافمان کمک میکند.

# یادگیری مفهوم

○ عبارت است از بدست آوردن یک مفهوم کلی از روی مثالهای مثبت و منفی.

○ این کار از طریق جستجو در فضای فرضیه های ممکن از قبل دانسته برای پیدا کردن بهترین فرضیه منطبق بر داده های آموزشی انجام می شود.

در عمل با موارد زیادی از یادگیری مواجه هستیم که در آن یک مفهوم کلی از روی مثالهای آموزشی یاد گرفته میشود. **چنین مفاهیمی را میتوان توسط یک تابع بولی نشان داد.**

## مثال: مفهوم هدف

○ یادگیری روزهایی از هفته که شخص از ورزش آبی لذت برده است

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

○ ترکیب تمام ویژگیهای موجود در مثالها باشد

<sky,temp,humidity,wind,water,forecast>

# نحوه نمایش فرضیه

○ برای هر ویژگی :

○ آنرا با "؟" نشان می دهیم اگر هر مقداری برای آن **قابل قبول** باشد (don't care)

○ آنرا با "∅" نشان می دهیم اگر هیچ مقداری برای آن **قابل قبول نباشد**

○ آنرا با یک مقدار مشخص نشان می دهیم (مثلا گرم) اگر این مقدار حتمی باشد

اگر مثالی مثل  $x$  تمامی قیود فرضیه  $h$  را برآورده کند آنگاه  $h$  آنرا یک مثال مثبت تعیین میکند یعنی:  $h(x) = 1$

مثلا فرضیه  $\langle ?, cold, high, ?, ?, ? \rangle$  برای مجموعه مثال فوق یک فرضیه غلط است  
 $h(x) = 0$

**عمومی ترین فرضیه :**  $\langle ?, ?, ?, ?, ?, ? \rangle$  تمام روزها یک مثال مثبت هستند

**اختصاصی ترین فرضیه :**  $(\emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset)$  هیچ روزی مثال مثبت نخواهد بود

# تعاریف کلی

- نمونه (Instance): مجموعه ای از ویژگیها که مفهوم را تعریف می کنند.
- مفهوم هدف (Target Concept): تابعی که باید یاد گرفته شود (در این فصل با توابع بولی کار داریم)

$$C : x \in \{0,1\}$$

در جدول مثال فوق داریم:

$C(x)=1$  if  $EnjoySport=Yes$        $C(x)=0$  if  $EnjoySport=No$   
مثبت یا منفی

- خروجی یادگیر: فرضیه ای مثل  $h$  بنحویکه  $h(x)=c(x)$  باشد (برای تمامی  $x$ ها)
- ورودی یادگیر: مجموعه ای از مثالهای آموزشی  $x$  به همراه مقدار تابع هدف برای آنها  $\langle x, c(x) \rangle$

**مجموعه تمام فرضیه ها را با  $H$  نشان می دهیم**

# فرض اساسی یادگیری استنتاجی

## Inductive

○ هر تابعی که بتواند تقریب خوبی از تابع هدف برای یک مجموعه آموزشی به اندازه کافی بزرگ باشد قادر خواهد بود که تابع هدف را در مورد مثالهای مشاهده نشده هم تقریب بزند.

○ از آنجائیکه در این نوع از یادگیری تنها اطلاعات موجود مجموعه مثالهای آموزشی است لذا در بهترین حالت یک الگوریتم یادگیری میتواند فرضیه ای را ارائه دهد که تابع هدف را بر روی مثالهای آموزشی تقریب بزند.

# یادگیری مفهوم بعنوان جستجو

- میتوان به مسئله یادگیری مفهوم بعنوان جستجو در فضای بزرگی از فرضیه ها نگاه کرد. این فضا بصورت غیر صریح با نحوه نمایش فرضیه ها تعریف میشود.
- **انتخاب نحوه نمایش فرضیه**، فضای فرضیه های قابل نمایش و قابل یادگیری را مشخص خواهد کرد.
- در مثال قبل:

○ تعداد کل فرضیه ها برابر است با:  $5120 = 5,4,4,4,4,4$

○ تعداد کل نمونه های ممکن برای مثال فوق:  $3*2*2*2*2*2=96$

○ با توجه به اینکه تهی برای تمامی مثالها مقدار منفی خواهد داشت تعداد فرضیه های با معنا عبارتند از:  $1+(4.3.3.3.3.3)=973$

- *Sky* (with possible values *Sunny*, *Cloudy*, and *Rainy*),
- *AirTemp* (with values *Warm* and *Cold*),
- *Humidity* (with values *Normal* and *High*),
- *Wind* (with values *Strong* and *Weak*),
- *Water* (with values *Warm* and *Cool*), and
- *Forecast* (with values *Same* and *Change*).

# ترتیب کلی به جزئی فرضیه ها

• الگوریتمهای یادگیری زیادی مبنای جستجوی خود را بر پایه ساختار مرتب سازی، کلی به جزئی فرضیه ها بنا نهاده اند

• مثال:

$$h_1 = \langle \text{Sunny}, ?, ?, \text{Strong}, ?, ? \rangle$$

$$h_2 = \langle \text{Sunny}, ?, ?, ?, ?, ? \rangle$$

چون  $h_2$  قید کمتری دارد تعداد بیشتری از نمونه ها را مثبت ارزیابی می کند.

هر نمونه ای که  $h_1$  آنرا مثبت ارزیابی کند  $h_2$  نیز مثبت خواهد دانست، بنابراین  $h_2$  از  $h_1$  کلی تر است.



# Deductive Learning (Ex)

- o **top-down logic** contrasts with **inductive (bottom-up logic)**
- o In deductive, a **conclusion is reached from general statements**, but in **inductive the conclusion is reached from specific examples**.
- o An **example** of a deductive argument:
  - o All men are mortal.
  - o Aristotle is a man.
  - o Therefore, Aristotle is mortal.

# Deductive Learning (Cont.)

- o The first premise states **that all objects classified as "men" have the attribute "mortal"**. The **second premise** states that **"Aristotle" is classified as a "man" – a member of the set "men"**.
- o The conclusion then states that **"Aristotle" must be "mortal" because he inherits this attribute** from his classification as a "man".

# Law of Detachment

- o A single conditional statement is made, and a **hypothesis (P) is stated. The conclusion (Q) is then deduced** from the statement and the hypothesis. The most basic form is listed below:
  - o  $P \rightarrow Q$  (conditional statement)
  - o P (hypothesis stated)
  - o Q (conclusion deduced)

# Deductive Learning (Cont.)

- o In deductive, we can conclude Q from P by using the law of detachment. However, if the conclusion (Q) is given instead of the hypothesis (P) then there is no valid conclusion.
- o The following is an example of an argument using the law of detachment in the form of an if-then statement:
  - o If an angle satisfies  $90^\circ < A < 180^\circ$ , then A is an obtuse angle.
  - o  $A = 120^\circ$
  - o A is an obtuse angle, Since the measurement of angle A is greater than  $90^\circ$ , we can deduce that A is an obtuse angle.

# Law of Syllogism (قیاسی)

- o The law of syllogism **takes two conditional statements and forms a conclusion by combining the hypothesis of one statement with the conclusion of another.** Here is the general form, with the true premise P:
  - o  $P \rightarrow Q$
  - o  $Q \rightarrow R$
  - o Therefore,  $P \rightarrow R$ .

## (Cont.)

- o The following is an example:
  - o If Larry is sick, then he will be absent
  - o If Larry is absent, then he will miss his classwork.
  - o If Larry is sick, then he will miss his classwork.
- o We deduced the final statement by combining the hypothesis of the first statement with the conclusion of the second statement. **We also allow that this could be a false statement.** This is an example of the **Transitive Property** in mathematics. The Transitive Property is many times phrased in this form:
  - o  $A=B$
  - o  $B=C$
  - o Therefore  $A=C$

# Law of Contrapositive

- o The law of contrapositive states that, in a conditional, **if the conclusion is false, then the hypothesis must be false also**. The general form is the following:
  - o  $P \rightarrow Q$
  - o  $\sim Q$
  - o Therefore we can conclude  $\sim P$ .

# Law of Contrapositive (Cont.)

- o The following are examples:
  1. If it is raining, then there are clouds in the sky.
  2. There are no clouds in the sky.
  3. Thus, it is not raining.
  
- o **Deductive is generally thought of as a skill that develops without any formal teaching or training.**



# خاصیت اصلی یادگیری استقرائی

• یک سیستم یادگیر که هیچ پیش فرضی در مورد ماهیت تابع هدف نداشته باشد قادر به دسته بندی داده های مشاهده نشده نخواهد بود.

# Model Selection & Generalization

- Learning is an **ill-posed problem**; **data is not sufficient to find a unique solution**
- The need for **inductive bias**, assumptions about  $\mathcal{H}$
- **Generalization**: **How well a model performs on new data**
- Overfitting:  $\mathcal{H}$  more complex than  $C$  or  $f$
- Underfitting:  $\mathcal{H}$  less complex than  $C$  or  $f$

# Deductive vs Inductive

## **Deductive**

Generalization (or Rule) —————> Specific Examples or Activities

## **Inductive**

Specific Examples or Activities —————> Generalization (or Rule)

Figure 1, Deductive and Inductive Learning adapted from  
<http://www.sasked.gov.sk.ca/docs/policy/approach/instrapp05.html>

# Inductive Learning

- **The premises seek to supply strong evidence for (not absolute proof of) the truth of the conclusion.**
- **While the conclusion of a deductive argument is supposed to be certain, the truth of an inductive argument is supposed to be *probable*, based upon the evidence given.**
- **The premises of an inductive logical argument indicate some degree of support (inductive probability) for the conclusion but do not entail it; They suggest truth but do not ensure it.**

# Inductive Learning (Cont.)

- o Inductive is probabilistic; it only states that, given the premises, the **conclusion is *probable***.
- o A statistical syllogism is an example of inductive Learning:
  1. Almost all people are taller than 26 inches
  2. Gareth is a person
  3. Therefore, Gareth is almost certainly taller than 26 inches

# Inductive Learning (Cont.)

- As a stronger **example**:
  - 100% of biological life forms that we know of depend on liquid water to exist. Therefore, if we discover a new biological life form it will probably depend on liquid water to exist.
  - This **argument could** have been **made every time a new biological life form was found**, and would have been **correct** every time; this **does not mean it is impossible** that in the **future a biological life form** that does not require water could be discovered.

# Inductive Learning (Cont.)

- o As a result, the argument may be stated less formally as:
  - o All biological life forms that we know of depend on liquid water to exist.
  - o All biological life probably depends on liquid water to exist.

# Types of Inductive

- o **Generalization**
- o **Statistical syllogism**
- o **Simple induction**
- o **Argument from analogy**
- o **Causal inference**
- o **Prediction**



# Generalization

- o A generalization (more accurately, an *inductive generalization*) proceeds from **a premise about a sample to a conclusion about the population.**
- o **The proportion Q of the sample has attribute A. Therefore: The proportion Q of the population has attribute A.**

# Generalization (Ex.)

- There are **20 balls—either black or white**—in an urn. To estimate their respective numbers, you draw a **sample of four balls and find that three are black and one is white**. A good inductive generalization would be that there are **15 black, and five white**, balls in the urn.
- How much the premises support the conclusion depends upon (a) the **number in the sample group**, (b) the **number in the population**, and (c) the **degree to which the sample represents the population** (which may be achieved by taking a random sample).

# Statistical syllogism (قیاسی)

- A statistical syllogism **proceeds from a generalization to a conclusion** about an individual.
- A proportion Q of population P has attribute A. An individual X is a member of P.
- **Therefore:** There is a probability which corresponds to Q that X has A. The proportion in the first premise would be something like "3/5ths of", "all", "few", etc. Two dicto simpliciter fallacies can occur in statistical syllogisms: "accident" and "converse accident".

# Simple Induction

- Simple induction proceeds **from a premise about a sample group to a conclusion about another individual.**
- Proportion  $Q$  of the known instances of population  $P$  has attribute  $A$ . Individual  $I$  is another member of  $P$ .
- Therefore: There is a probability corresponding to  $Q$  that  $I$  has  $A$ . This is a **combination of a generalization and a statistical syllogism**, where the conclusion of the generalization is also the first premise of the statistical syllogism.

# Argument from analogy

(بحث از قیاس)

- o The process of analogical inference **involves noting the shared properties of two or more things**, and from this basis inferring that they also share some further property:
  - o P and Q are similar in respect to properties a, b, and c.
  - o Object P has been observed to have further property x. Therefore, Q probably has property x also.

# Causal inference

(براساس علت)

- A causal inference draws a conclusion about a causal connection **based on the conditions of the occurrence of an effect**. Premises about the **correlation of two things can indicate a causal relationship between them, but additional factors must be confirmed to establish the exact form of the causal relationship**.

# Prediction

- o A prediction draws a **conclusion about a future individual from a past sample.**
- o **Proportion Q of observed members of group G have had attribute A.**
- o **Therefore:** There is a **probability** corresponding to Q that other members of group G will have attribute A when next observed.

## (Ex.)

- Initialize  $h$  to the most specific hypothesis in  $H$ ;
- For each positive training instance  $x$ 
  - For each attribute  $a_i$  in  $h$ 
    - If the constraint  $a_i$  in  $h$  is not satisfied by  $x$  then replace  $a_i$  in  $h$  by the most general constraint that is satisfied by  $x$
- Output hypothesis  $h$ .
  
- Guaranteed to output the most specific hypothesis within  $H$  that is consistent with the positive training examples.
- Notice that negative examples are ignored.



موفق باشید