

مفاهیم دسته بندی

تعاریف

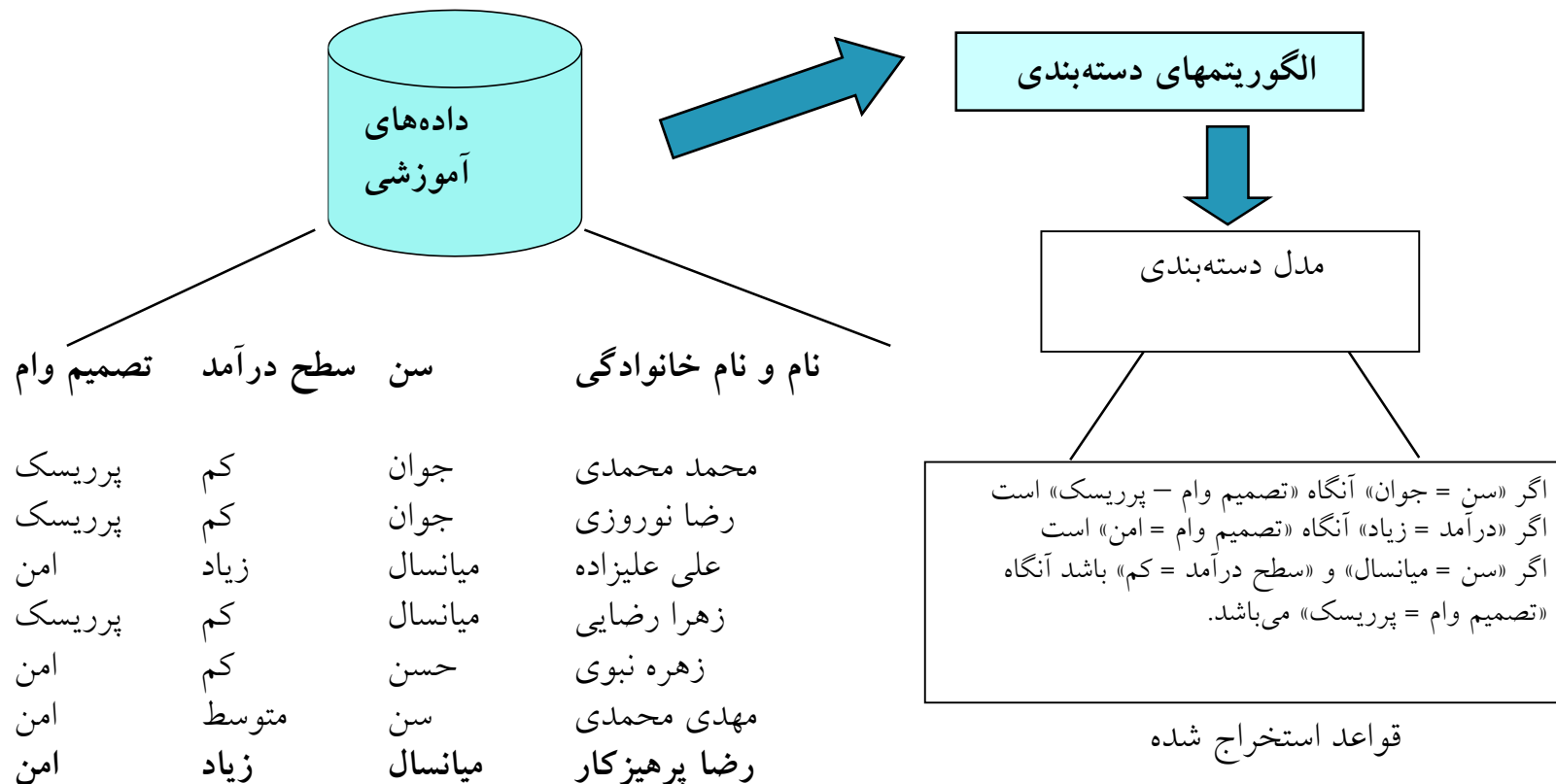
- دسته بندی یعنی تخصیص یک برچسب دسته به مجموعه های ازداده ها که هنوز دسته بندی نشده اند.
- داده ها دارای k ویژگی هستند که به صورت A_1, \dots, A_k نشان داده می شود.
- هر مورد یا مثال بوسیله مقادیر ویژگیها و یک برچسب دسته توصیف می شود.
- اغلب یکی از ویژگی های داده دسته آن را نشان می دهد.

• تفاوت دسته بندی و خوشه بندی



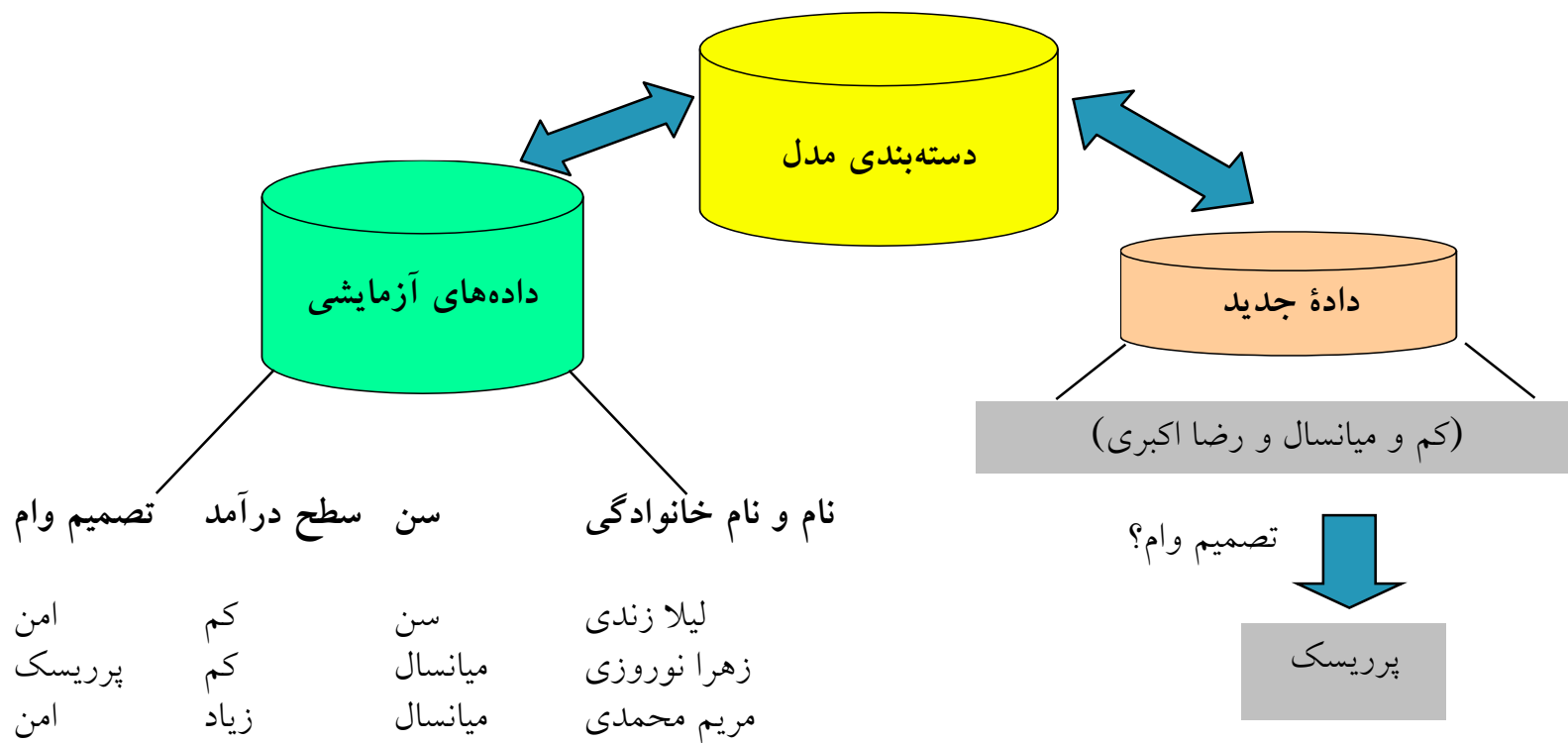
فرایند دو مرحله ای در دسته بندی

• ساخت مدل



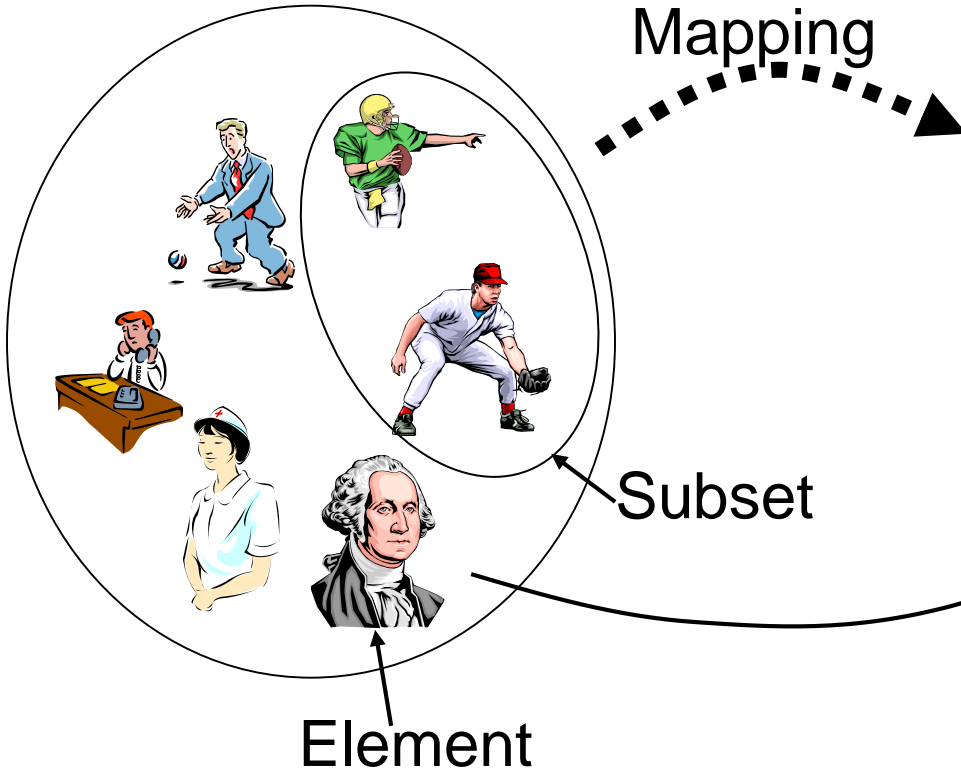
فرایند دو مرحله ای در دسته بندی

• استفاده از مدل

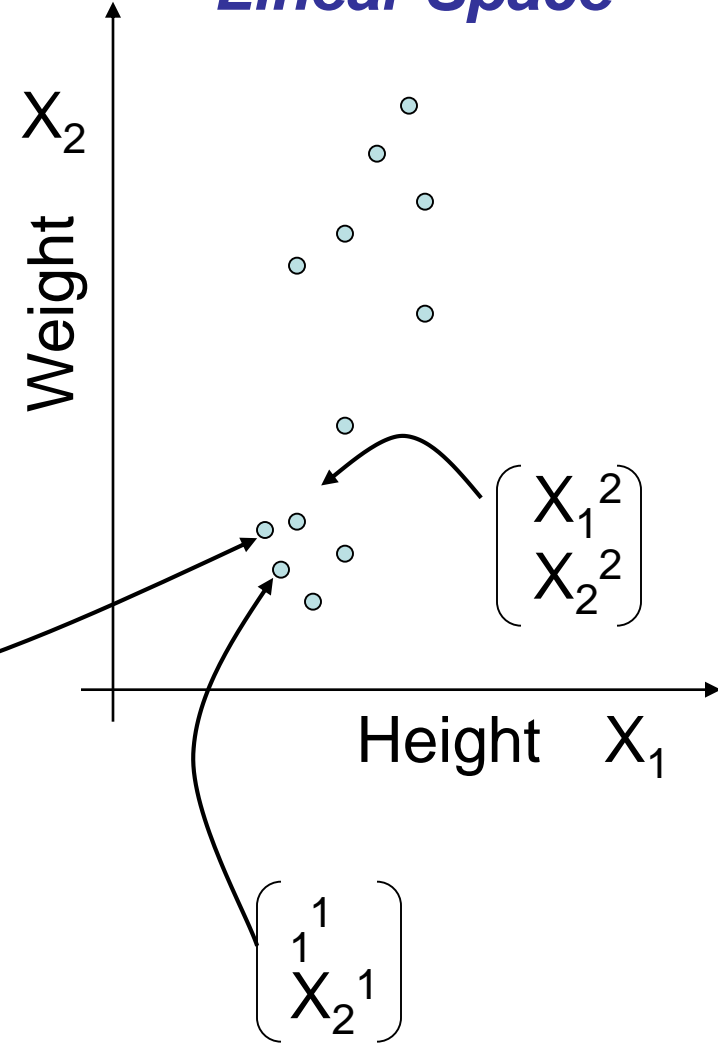


Mapping

Set



Linear Space



$$Y = F(X)$$

روشهای ارزیابی دسته بندی ها

- صحت پیش بینی (Accuracy) و بقیه پارامترها
- زمان لازم برای ساخت مدل و زمان لازم برای استفاده از مدل
- پایداری (Robustness) :
- توانایی برخورد با داده های غیر معمول (noise) و یا مقادیر جا افتاده (missing values)
- قابلیت تفسیر (Interpretability) :
- قابل فهم بودن مدل و فراهم سازی بینش لازم توسط آن
- جمع و جور بودن (Compactness) مدل :
- اندازه درخت یا تعداد قواعد و ...



شاخص دقت در دسته بندی

Classes	<i>buys_computer = yes</i>	<i>buys_computer = no</i>	Total	Recognition (%)
<i>buys_computer = yes</i>	6,954	46	7,000	99.34
<i>buys_computer = no</i>	412	2,588	3,000	86.27
Total	7,366	2,634	10,000	95.52

	C_1	C_2
C_1	true positives	false negatives
C_2	false positives	true negatives

$$sensitivity = \frac{t_pos}{pos}$$

$$specificity = \frac{t_neg}{neg}$$

$$precision = \frac{t_pos}{(t_pos + f_pos)}$$

$$accuracy = sensitivity \frac{pos}{(pos + neg)} + specificity \frac{neg}{(pos + neg)}$$

مساله تعمیم

مساله تعمیم

- کلاس بندی الگو فقط می تواند برای نمونه های موجود مدل کلاس بندی را مشخص نماید. اما در هر حال می دانیم که معمولا همه داده ها در دسترس و موجود نیستند و مشکل از اینجا ناشی می شود

