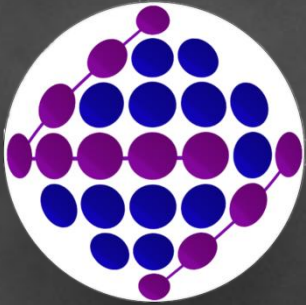


Decision Trees and Random Forests



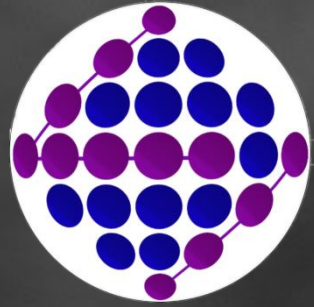
فصل دہم



شرکت فروش مصنوعی رسا

مدرس : خرید فاشمی نژاد

معرفی Decision Trees



جلسہ سے ام



شرکت فروش مصنوعي رسا

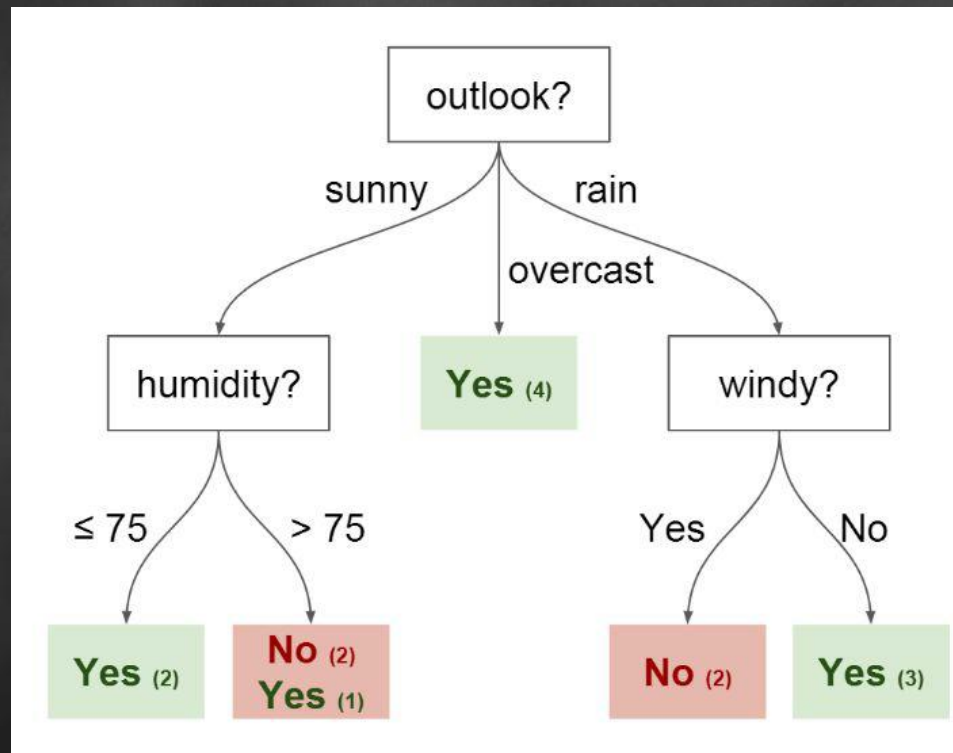
مدرس : خرید فاشمی نژاد

- تصور کنید که من هر جمعه تنیس بازی می کنم و همیشه از دوستم دعوت می کنم
- کافی دوست من می آید ، کافی نه .
- آمدن او به عوامل مختلفی بستگی دارد ، مانند: آب و هوا ،
- دما ، رطوبت ، باد و غیره
- میخوایم ببینیم که هر ویژگی در حاضر شدن اون تاثیر دارد یا نه

Temperature	Outlook	Humidity	Windy	Played?
Mild	Sunny	80	No	Yes
Hot	Sunny	75	Yes	No
Hot	Overcast	77	No	Yes
Cool	Rain	70	No	Yes
Cool	Overcast	72	Yes	Yes
Mild	Sunny	77	No	No
Cool	Sunny	70	No	Yes
Mild	Rain	69	No	Yes
Mild	Sunny	65	Yes	Yes
Mild	Overcast	77	Yes	Yes
Hot	Overcast	74	No	Yes
Mild	Rain	77	Yes	No
Cool	Rain	73	Yes	No
Mild	Rain	78	No	Yes

من می خوامم از این داده ها برای پیش بینی استفاده
کنم که آیا برای بازی حاضر می شود؟
درخت تصمیم گیری یک روش شهودی برای انجام این

کار است



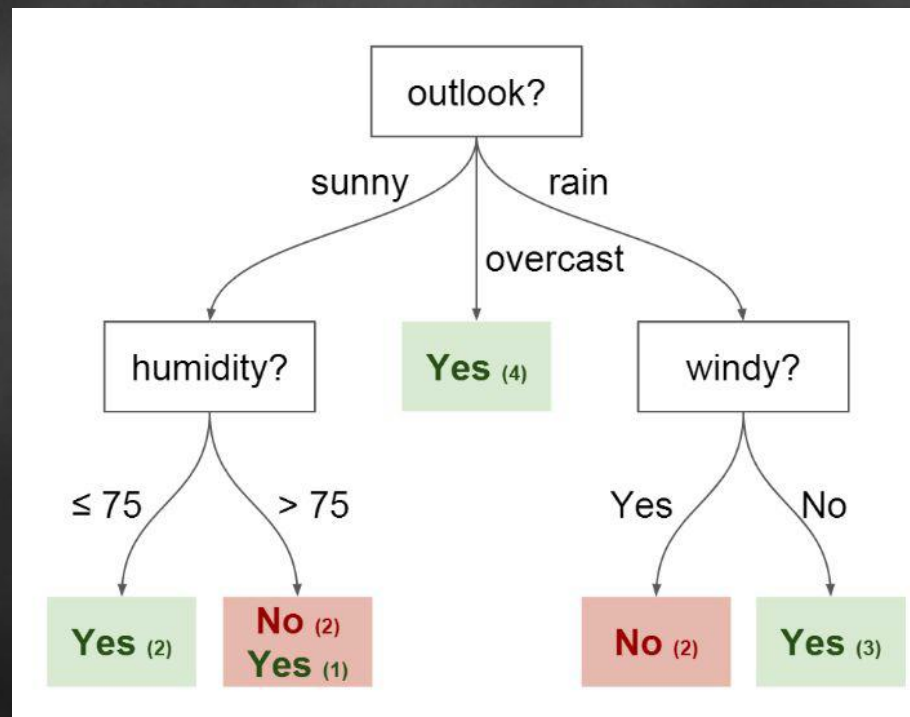
در این درخت ما داریم:

- گره ها

- دسته بندی ویژگی های خاص

- لبه ها

- نتیجه دسته بندی ها



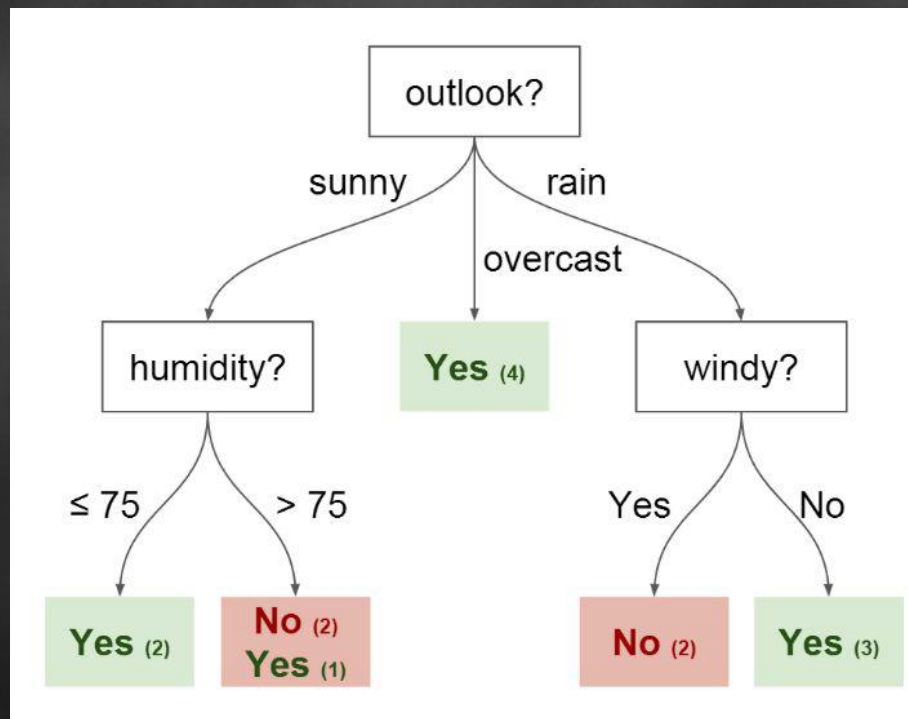
در این درخت ما داریم:

- ریشه

- گره های اولین تقسیم

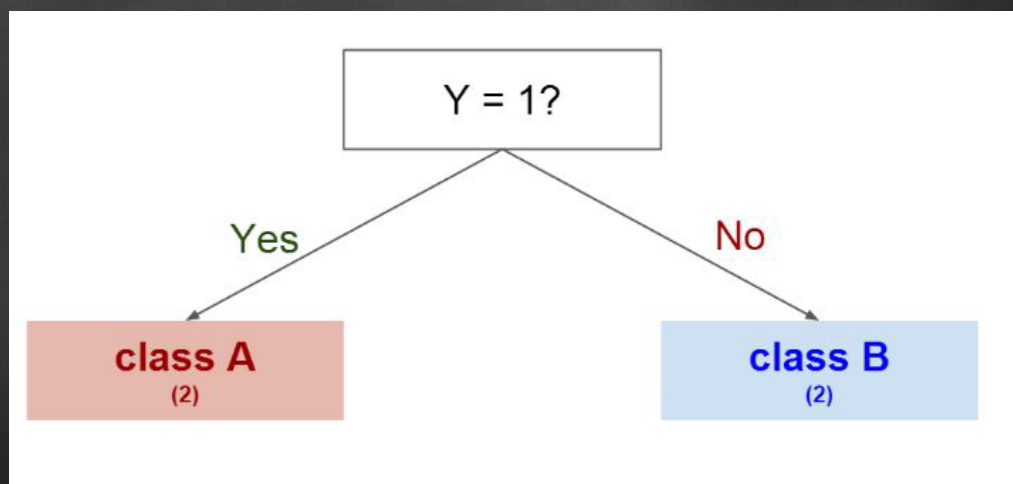
- برگ ها

- گره های پایانی که نتیجه را پیش بینی می کنند



X	Y	Z	Class
1	1	1	A
1	1	0	A
0	0	1	B
1	0	0	B

واضح است که:



ما همچنین می توانیم ابتدا ویژگی های دیگر را تقسیم کنیم:

first split on					
X		Y		Z	
X = 0	X = 1	Y = 0	Y = 1	Z = 0	Z = 1
B	A	B	A	A	A
	A		A		
	B				B

آنتروپی و بهره اطلاعات (Information Gain) روشهای ریاضی هستند برای
انتخاب بهترین تقسیم

Entropy:

$$H(S) = - \sum_i p_i(S) \log_2 p_i(S)$$

Information Gain:

$$IG(S, A) = H(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{S} H(S_v)$$

برای بهبود عملکرد ، می توانیم از درختان زیادی با
نمونه تصادفی از ویژگی های دسته انتخاب شده
استفاده کنیم

- یک نمونه تصادفی جدید از ویژگی های هر درخت در
هر دسته انتخاب شده

- برای طبقه بندی ، m معمولاً به عنوان ریشه مربع p
انتخاب می شود