

$$E_j = \frac{S_j}{S_T} \sum -p_j \log(p_j)$$

$$\text{Entropy}(H) = - \sum P(H) \cdot \log P(H)$$

① معیار انتخاب شده: آنتروپی

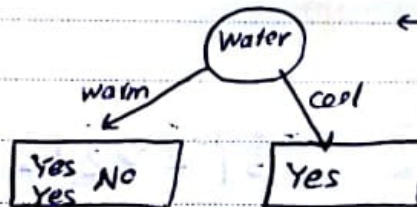
②

برای تمامی متغیرها این معیار را محاسبه کنیم. اما هر بزرگدلی بر حسب دای تکلیف است!

* با بررسی داده‌ها متوجه می‌شویم که فیلدهای sky و Air Temp به طور مستقیم در حال دیکراسازی داده‌ها هستند و نیازی نیست برای آن‌ها آنتروپی محاسبه کنیم.

بعد جستجوی هر کدام از داده‌ها نیاز به انتخاب بهترین ویژگی و محاسبه آنتروپی نداریم.

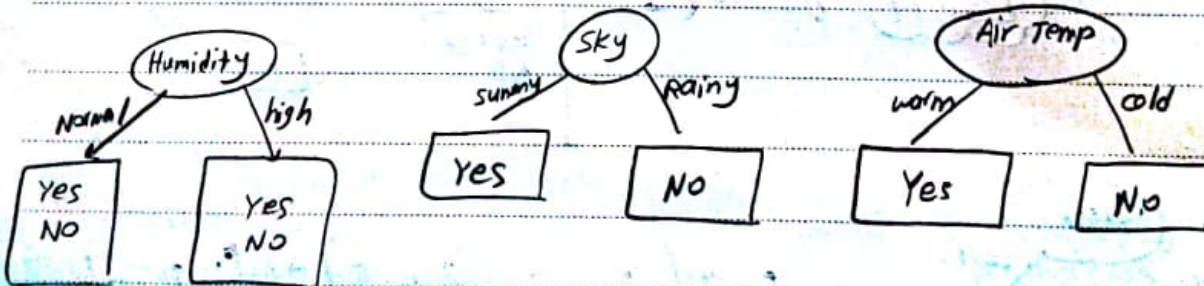
با بقیدی متغیرها به تخطی سطر classification انجام داد. $\left\{ \begin{array}{l} \text{sky} \\ \text{Air Temp} \end{array} \right.$ فقط



$$\text{Entropy} \rightarrow -\frac{3}{4} \log \frac{2}{3} \times \frac{2}{3} + \frac{1}{3} \log \frac{1}{3} + \frac{1}{4} [-\log 1 + 0]$$



$$\approx 0.44 \rightarrow \text{پس ویژگی دمای هوا را می‌خواهیم}$$

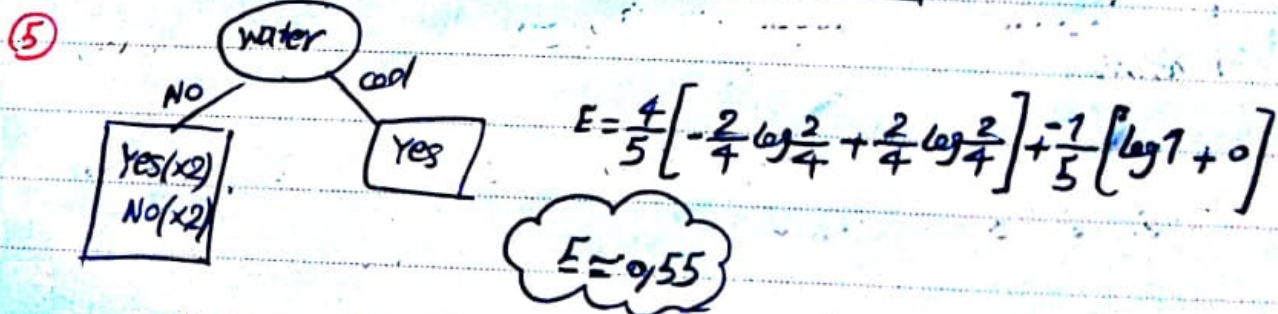
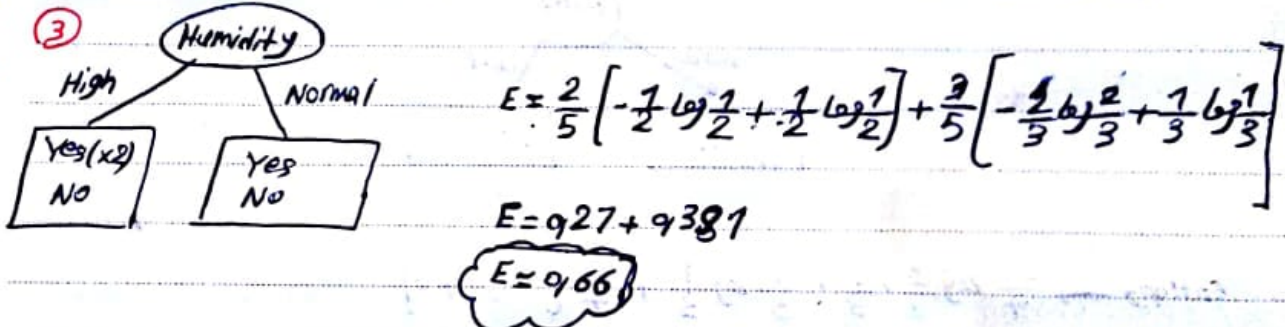
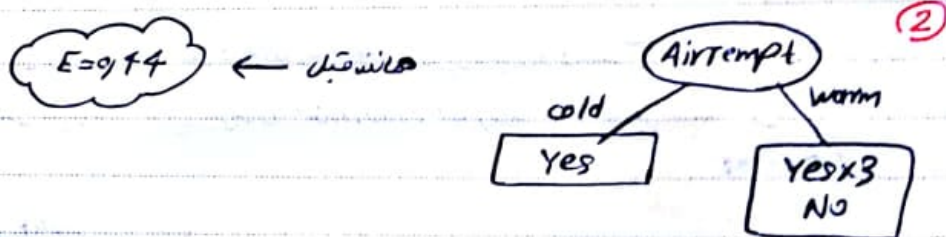
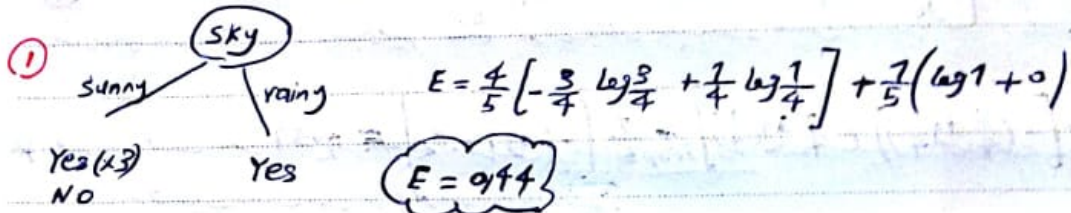


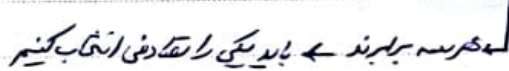
* Wind حذف می‌شود. جهت همش داریم. Result می‌دهد و اسدرا "مباسازی ای انجام نمی‌دهد.

نکته: دقت تقسیم بهتر است با دو فیلد $\left\{ \begin{array}{l} \text{sky} \\ \text{Air Temp} \end{array} \right.$ (یکبار کلا) اضافه شود. دیگر نیازی به سطح دیگر نیست.

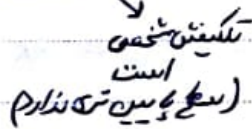
Subject: _____
Date: _____

(b) ملاک مقیاس جدید اضافه شده، (سپید جدید طلسم)، جدید دقت تقسیم را از نو بکشیم.





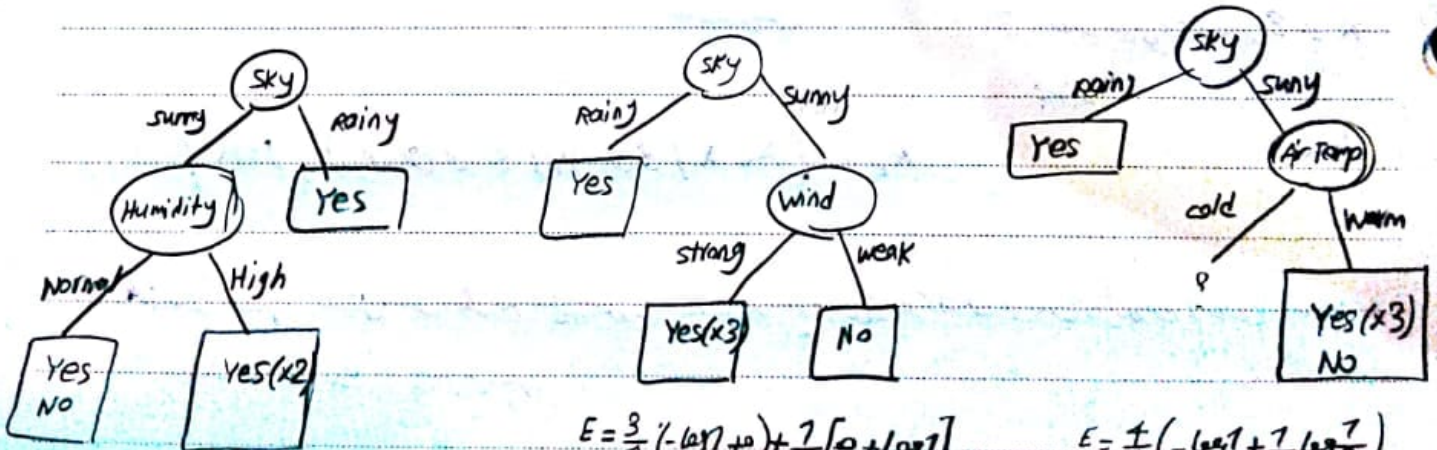
۵۷ باید بدخل دوم را انتخاب کنیم.



من فکری که سر در این بود که گفتم

از بین سه فیچر که کمترین استفاده را داشتند اکثر این جا به استفاده بیشتر است. بین wind و Air Attempt. کاسه ای آتروپه انجام شود و wind انتخاب شد.

راه دیگر این است که محل فیچرها را جابجایی کنیم و یه یه هم آتروپه را صلب کنیم.



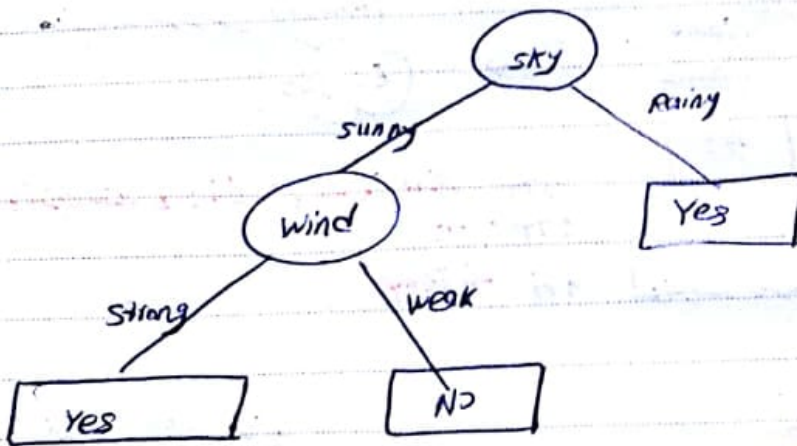
$$E = \frac{1}{2} \left[-\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2} \right] + \frac{1}{2} [-\log 1 + 0]$$

$$E = \frac{1}{4} \left(-\log 1 + \frac{1}{3} \log \frac{7}{3} \right)$$

↑ 0

$E = 0,5$

← چوں که ی بر لایک کا میٹر یک کر شوند ، کار سائن (حقیقت) تمام بندہ است .



(2)

تفسیر سؤال: منطقه‌ی R_K برای هر نقطه‌ی P_K مجموعه‌ی نقاطی x که فاصله‌ی آن‌ها با P_K نزدیکتر از فاصله‌ی آن‌ها با P_j است.

منیت



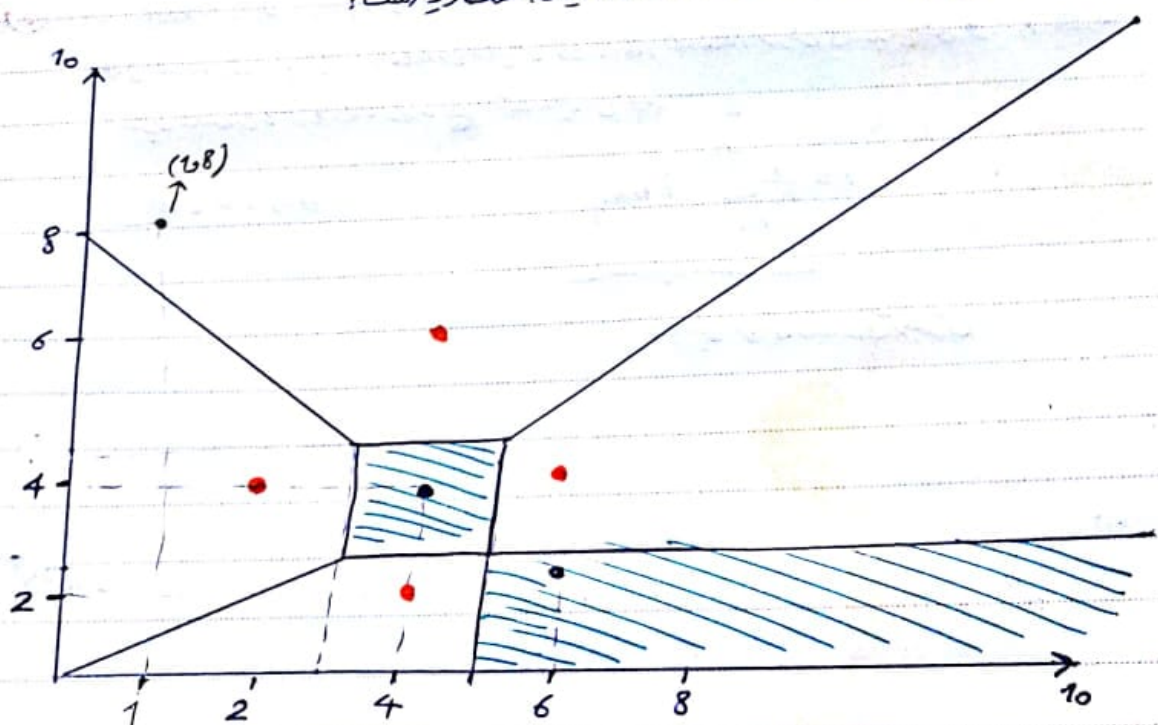
در نقطه‌ی x داخل مجموعه‌ی خود، باید یک مرز تعیین کنی → فاصله‌ی x با نقطه‌ی K دارد، کوچکتر یا برابر از فاصله‌ی آن با P_j باشد.

$$R_K = \{x \in X \mid d(x, P_K) \leq d(x, P_j) ; j \neq K\}$$

مرزهای تعیین در شکل مورد نظر به این صورت است که بین نقاط هر کلاس و بین هر کلاس (داخل هر کلاس) خط‌های جداکننده را رسم می‌کنیم.

برای دو نقطه‌ی هر مرز یک مرز تعیین کنی مثل شد چون نقطه‌ی x که اطراف آن‌ها هستند، به نقطه‌ی یکی از آن‌ها نزدیک‌ترند. پس می‌توان ناحیه‌ی درونی را این ناحیه رسم کرد.

در صورت تقسیم بیرون گداس که می استفاده از حاصل من اقلیدس به صورت زیر است:



(b) نقطه (6, 8) در گداس قرمز (نقطه قرمز) قرار می گیرد. * رنگه من با شکل منق دایه و آبی و قرمز
اصلی
خام برعکس: دی

(c) اثر k را فیل بزرگ نایم، یعنی تعداد فیل زیاد می هم می برای یک نقطه انتخاب می کنیم.
الگو توزیع به صورتی باشد که داده ها ضعیف نباشند (معمولا" اینطور است)، انظر دلای جالبی
می کنیم و عملا" کار انجام می دهیم.
چون در واقع دلای هم می های زیادی را که از گداس های دیگری هستند در یک می لای و دسته بندی فیل
می کنیم.

Subject :
Date :

نکته : kNN برای رگرسیون هم قابل استفاده است.

وقتی یک داده ی جدید می آید ، فاصله ی آن با تمام داده های آموزش می گسسته می شود و K نقطه که کمترین فاصله را با آن داده دارند انتخاب می شوند.

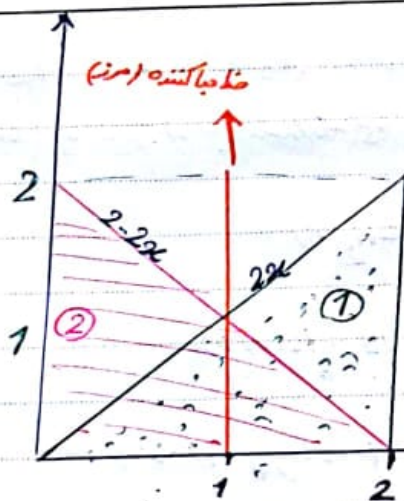
$$Y = \frac{1}{K} \sum_{i=1}^K Y_{train_i}$$

میانگین برابر می شود :

میانگین خروجی داده های آموزش

Subject:
Date:

(a)



(3)

$$P(x/w_1) = \begin{cases} 2x & 0 \leq x < 1 \\ 0 & 0.5 \leq x \leq 1 \end{cases}$$

$$P(x/w_2) = \begin{cases} 2-2x & 0 \leq x < 1 \\ 0 & 0.5 \leq x \leq 1 \end{cases}$$

(b)

مقیاسی مناسبی دو کلاس ← باید error سبب مقدار ممکن شود.

$$P(\text{error}/x) \rightarrow \text{error} : \begin{cases} P(w_1/x) \rightarrow w_2 \\ P(w_2/x) \rightarrow w_1 \end{cases}$$

$$\text{کفایت} \rightarrow P(\text{error}/x) = \int_{-\infty}^{\infty} P(\text{error}/x) \cdot P(x) dx$$

چگونه از رابطه ی بالا استفاده کنیم؟

← فرض کنیم توزیع احتمالات - صوتی باشد احتمال وجود نمونه در یک کلاس بیشتر از دیگران شود.

$$\text{مثلا: } P(w_2/x) > P(w_1/x) \rightarrow P(w_2) \cdot P(x/w_2) > P(w_1) \cdot P(x/w_1)$$

$$\text{دری} \rightarrow \frac{P(x/w_1)}{P(x/w_2)} > \frac{P_1}{P_2}$$

خطی کل را اگر $E(x)$ نشان دهیم:
 $E(x) = P_1 E_1(x) + P_2 E_2(x)$

خطی مناسبی هر دو کلاس می باشد.

Subject: _____
Date: _____

امتیازی

برای کسبی مرزها کننده (دو کلاس)، باید از کلاسifikاسیون Bayesian استفاده کنیم.

برای اینکه حد جدا کننده را پیدا کنیم، از همین رابطه ای که پیدا کردیم استفاده میکنیم.

$$Err(x) = P_1 \cdot (x) + P_2 \cdot Err_2(x)$$

$$Err(x) = P_1 \int_0^1 P_1(x) dx + P_2 \int_0^1 P_2(x) dx$$

$$Err(x) = 0.5 \int_0^1 2x dx + 0.5 \int_0^1 (2-2x) dx = 0.5 \times 2 \times \frac{x^2}{2} \Big|_0^1 + 0.5 \times \left(2x - \frac{2x^2}{2} \Big|_0^1 \right)$$

$$Err(x) = \frac{1}{2} [1+1] = 1$$

مقدار نفعی جدا کننده (مرز)، در نقطه ① است. (در شکل مشخص کرده ام - معنی قبل)

Subject: _____
Date: _____

④ استناد قرارگیری یک داده در یک کلاس:

$$p(w_1/x) = \frac{p(x/w_1) p(w_1)}{p(x)}$$

$$p(w_2/x) = \frac{p(x/w_2) p(w_2)}{p(x)}$$

صورت عبارت $p(x)$ در هر دو ثابت است، پس باید صورت کسرها را مقایسه کنیم.

$$\max \{ p(x/w_1) \cdot p(w_1) \text{ و } p(x/w_2) \cdot p(w_2) \}$$

Yes کلاس $\rightarrow p(\text{Yes} | \text{Income} = \text{low}, \text{education} = \text{MS}, \text{Debt} = \text{high})$

$$p(\text{Yes}) \cdot p(\text{Income} = \text{low} | \text{Yes}) \cdot p(\text{education} = \text{MS} | \text{Yes}) \cdot p(\text{Debt} = \text{High} | \text{Yes})$$

$$\frac{1}{2} \times \frac{2}{3} \times \frac{1}{4} \times \frac{2}{3} = \frac{4}{48} = \frac{1}{12}$$

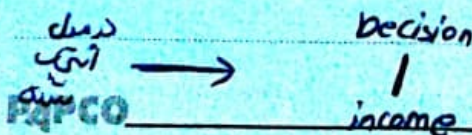
No کلاس $\rightarrow p(\text{No} | \text{Income} = \text{low}, \text{education} = \text{MS}, \text{Debt} = \text{High})$

$$p(\text{No}) \cdot p(\text{Income} = \text{low} | \text{No}) \cdot p(\text{education} = \text{MS} | \text{No}) \cdot p(\text{Debt} = \text{High} | \text{No})$$

$$\frac{1}{2} \times \frac{1}{3} \times \frac{3}{4} \times \frac{1}{3} = \frac{1}{16}$$

سپید فقط کلاس Yes انتخاب می شود

⑥ جمع در کلاسهای Bayesian و شبکه های از هم مستقل:



Subject: _____
Date: _____

©

	Yes	No
decision	$\frac{1}{2}$	$\frac{1}{2}$

احتمال اول

Income

	Yes	No
low	$\frac{2}{3}$	$\frac{1}{3}$
Medium	$\frac{1}{2}$	$\frac{1}{2}$
High	$\frac{1}{3}$	$\frac{2}{3}$

Education

	Yes	No
HS	1	0
BS	1	0
MS	$\frac{2}{4}$	$\frac{3}{4}$
phd	0	1

dept

	Yes	No
low	$\frac{1}{3}$	$\frac{2}{3}$
Medium	$\frac{1}{2}$	$\frac{1}{2}$
High	$\frac{2}{3}$	$\frac{1}{3}$

Subject: _____
Date: _____

(a) فرض استناد \rightarrow $H_0: \mu = 0 \rightarrow x_i \sim N(0, \Delta)$
 $H_1: \mu = m \rightarrow x_i \sim N(m, \Delta)$

(5)

تست درست-نادر: $\log \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - m)^2}{2}}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i)^2}{2}}}$

ساده سازی $\rightarrow \sum_{i=1}^n \frac{-(x_i - m)}{2} + \frac{x_i^2}{2} = m \sum_{i=1}^n x_i - \frac{nm^2}{2}$

$\rightarrow m \sum_{i=1}^n x_i \underset{H_0}{\overset{H_1}{>}} T_1$ then $\sum_{i=1}^n x_i \underset{H_0}{\overset{H_1}{>}} \left(\frac{T_1}{m}\right) T_2$ یک مرز جدید

(b) $PFA = P(H_0 > T | \mu = 0) = \int_{T_2}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x)^2}{2}} dx$

PM (تست خلاف فرض) های PFA است.

$PM = P(H_0 > T | \mu = m) = \int_{-\infty}^{T_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2}} dx$

$\rightarrow P_D = 1 - PM$
 $P_D = 1 - \int_{-\infty}^{T_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2}} dx = \int_{T_2}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2}} dx$

اینجا نتیجه نهایی!

Subject: _____
Date: _____

$$\textcircled{a} \quad \Pr(\neg P_3) = \Pr(\neg P_3|P_2) \Pr(P_2) + \Pr(\neg P_3|\neg P_2) \Pr(\neg P_2) \quad \textcircled{6}$$

$$\Pr(\neg P_3) = 0.18 \underbrace{\Pr(P_2)} + 0.17 \underbrace{\Pr(\neg P_2)} \quad \textcircled{I}$$

$$\rightarrow \Pr(P_2) = \Pr(P_2|P_1) \cdot \Pr(P_1) + \Pr(P_2|\neg P_1) \Pr(\neg P_1)$$

$$\Pr(P_2) = 0.8 \times 0.4 + 0.5 \times 0.6$$

$$\Pr(P_2) = 0.32 + 0.3$$

$$\Pr(P_2) = 0.62 \quad \textcircled{II}$$

$$1 - \Pr(P_2|P_1) \Pr(P_1) + 1 - \Pr(P_2|\neg P_1) \Pr(\neg P_1)$$

$$\rightarrow \Pr(\neg P_2) = \Pr(\neg P_2|P_1) \Pr(P_1) + \Pr(\neg P_2|\neg P_1) \Pr(\neg P_1)$$

$$= 0.2 \times 0.4 + 0.5 \times 0.6$$

$$\Pr(\neg P_2) = 0.2 \times 0.4 + 0.5 \times 0.6$$

$$\Pr(\neg P_2) = 0.08 + 0.3$$

$$\Pr(\neg P_2) = 0.38 \quad \textcircled{III}$$

II skip

III

I

$$\rightarrow 0.8 (0.62) + 0.17 (0.38)$$

$$\rightarrow \Pr(\neg P_3) = 0.76$$

Subject,
Date

(b)

اگر کارهای سه نفر کنیم می بینیم که متغیر P_2 از P_3 مستقل است.

$$Pr(P_2 | \neg P_3) = Pr(P_2) = 0.62$$

یعنی وابستگی P_2 که در استقلال قبیل ۳ دست آورده
وجود دارد، باز نویسی می کنیم.

یعنی P_2 مستقل از P_3 است،
مقدار این عبارت برابر با
 $Pr(P_2)$
می شود.

باسم سعادت میرمحمد

۹۳۳/۵۲۲