سوال ۱:

الف)

$$H(Y) = -\frac{۹}{۱۴}\log_۲\frac{۹}{۱۴} - \frac{۵}{۱۴}\log_۲\frac{۵}{۱۴} = ۰٫۴۰۹۷۷۹۳ + ۰٫۵۳۰۵۱ = ۰٫۹۴$$

$$H(Y|credit\_rating) = -\frac{۸}{۱۴}\left(\frac{۶}{۸}\log_۲\frac{۶}{۸} + \frac{۱}{۴}\log_۲\frac{۱}{۴}\right) - \frac{۶}{۱۴}\left(\frac{۱}{۲}\log_۲\frac{۱}{۲} - \frac{۱}{۲}\log_۲\frac{۱}{۲}\right)$$

$$= ۰٫۴۶۳۵۸۷۵ + ۰٫۴۲۸۵ = ۰٫۸۹۲۱$$

$$\rightarrow (H(Y) - H(Y|credit\_rating)) = ۰٫۰۴۷۸$$

$$H(Y|student) = -\frac{۷}{۱۴}\left(\frac{۶}{۷}\log\frac{۶}{۷} + \frac{۱}{۷}\log\frac{۱}{۷}\right) - \frac{۷}{۱۴}\left(\frac{۴}{۷}\log\frac{۴}{۷} + \frac{۳}{۷}\log\frac{۳}{۷}\right)$$

$$= ۰٫۷۸۸۴۵$$

$$\rightarrow H(Y) - H(Y|student) = ۰٫۹۴ - ۰٫۷۸۸۴۵ = ۰٫۱۵۱۵۵$$

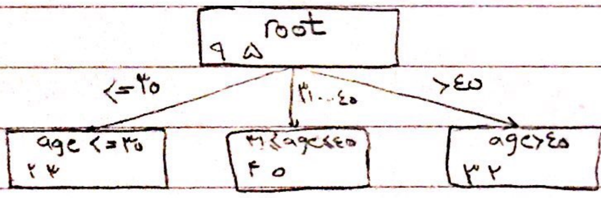
$$H(Y|income) = -\frac{۴}{۱۴}\left(\frac{۱}{۲}\log\frac{۱}{۲} + \frac{۱}{۲}\log\frac{۱}{۲}\right) - \frac{۶}{۱۴}\left(\frac{۱}{۳}\log\frac{۱}{۳} + \frac{۲}{۳}\log\frac{۲}{۳}\right) - \frac{۴}{۱۴}\left(\frac{۳}{۴}\log\frac{۳}{۴} - \frac{۱}{۴}\log\frac{۱}{۴}\right)$$

$$= ۰٫۹۱۱۰۶۳ \quad = ۰٫۰۲۸۹۳$$

$$H(Y|age) = -\frac{۵}{۱۴}\left(\frac{۲}{۵}\log\frac{۲}{۵} + \frac{۳}{۵}\log\frac{۳}{۵}\right) - \frac{۴}{۱۴}(۱\log ۱ + ۰\log ۰) - \frac{۵}{۱۴}\left(\frac{۲}{۵}\log\frac{۲}{۵} + \frac{۳}{۵}\log\frac{۳}{۵}\right)$$

$$= ۰٫۶۹۳۶$$

$$\rightarrow H(Y) - H(Y|age) = ۰٫۲۴۶۴$$



برای age≤۳۰ کار را ادامه می‌دهیم:

$$H(Y') = -\frac{۲}{۵}\log\frac{۲}{۵} - \frac{۳}{۵}\log\frac{۳}{۵} = ۰٫۹۷۰۹۵$$

$$H(Y'|credit\_rating) = -\frac{۲}{۵}\left(\frac{۱}{۲}\log\frac{۱}{۲} + \frac{۱}{۲}\log\frac{۱}{۲}\right) - \frac{۳}{۵}\left(\frac{۲}{۳}\log\frac{۲}{۳} - \frac{۱}{۳}\log\frac{۱}{۳}\right) = ۰٫۹۵۰۹۷۷$$

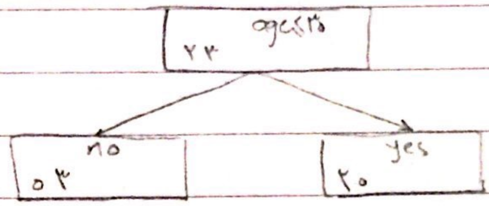
$$\rightarrow H(Y') - H(Y'|credit\_rating) = ۰٫۰۱۹۹۷$$

$$H(Y'|student) = -\frac{۲}{۵}(۱\log ۱ + ۰\log ۰) - \frac{۳}{۵}(۱\log ۱ + ۰\log ۰) = ۰$$

$$\rightarrow H(Y') - H(Y'|student) = ۰٫۹۷۰۹۵$$

$$H(Y'|income) = -\frac{2}{5}(1\log 1 + 0\log 0) - \frac{2}{5}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) - \frac{1}{5}(1\log 1 + 0\log 0) = 0.4$$

$$\rightarrow H(Y') - H(Y'|income) = 0.57095$$



برای age>40 می‌نویسیم:

$$H(Y'') = -\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5} = 0.97095$$

$$H(Y''|credit\_rating) = -\frac{3}{5}(1\log 1 + 0\log 0) - \frac{2}{5}(1\log 1 + 0\log 0) = 0$$

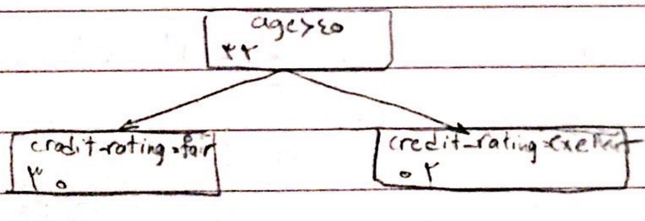
$$\rightarrow H(Y'') - H(Y''|credit\_rating) = 0.97095$$

$$H(Y''|student) = -\frac{2}{5}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) - \frac{3}{5}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) = 0.95097$$

$$\rightarrow H(Y'') - H(Y''|student) = 0.01997$$

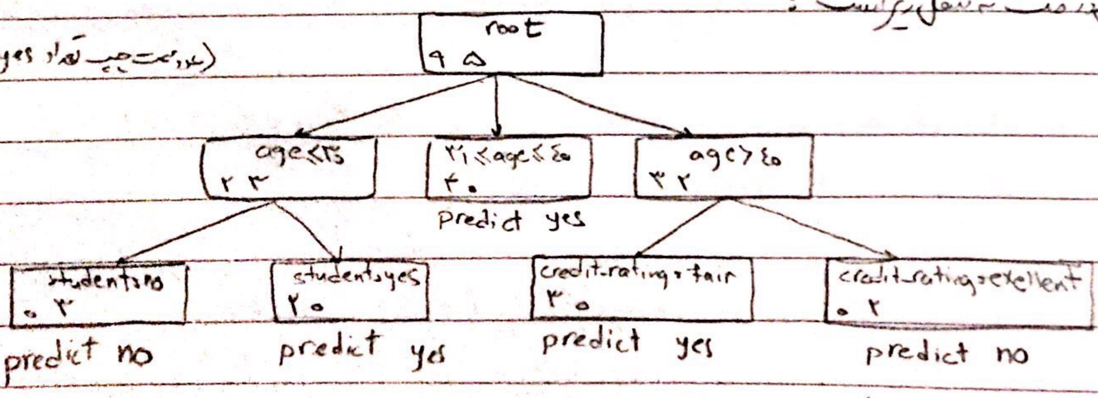
$$H(Y''|income) = -\frac{3}{5}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) - \frac{2}{5}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = 0.95097$$

$$H(Y'') - H(Y''|income) = 0.01997$$



شکل درخت به شکل زیر است:

(عدد سمت چپ تعداد yesها و سمت راست تعداد noها است)

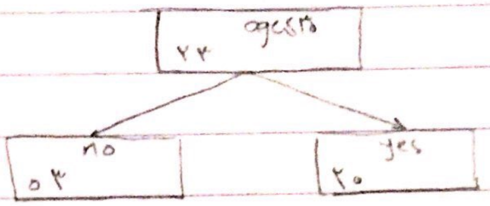


چون درخت را با این داده‌ها کشیدیم دقت آن برای داده‌های استفاده شده همان ٪100 است.

s.a.m

$H(Y'|income) = -\frac{۲}{۵}(1\log 1 + 0\log 0) - \frac{۲}{۵}(\frac{1}{۲}\log\frac{1}{۲} + \frac{1}{۲}\log\frac{1}{۲}) - \frac{1}{۵}(1\log 1 + 0\log 0) = ۰٫۴$

$\rightarrow H(Y') - H(Y'|income) = ۰٫۵۷۰۹۵$



برای age>۴۰ می‌نویسیم:

$H(Y^۴) = -\frac{۲}{۵}\log\frac{۲}{۵} - \frac{۳}{۵}\log\frac{۳}{۵} = ۰٫۹۷۰۹۵$

$H(Y^۴|credit\_rating) = -\frac{۳}{۵}(1\log 1 + 0\log 0) - \frac{۲}{۵}(1\log 1 + 0\log 0) = ۰$

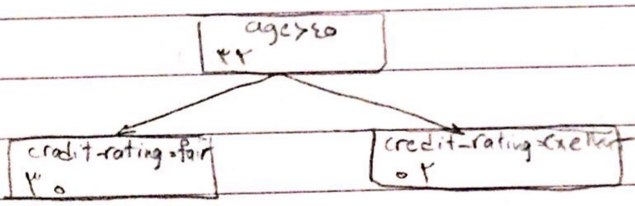
$\rightarrow H(Y^۴) - H(Y^۴|credit\_rating) = ۰٫۹۷۰۹۵$

$H(Y^۴|student) = -\frac{۲}{۵}(\frac{1}{۲}\log\frac{1}{۲} + \frac{1}{۲}\log\frac{1}{۲}) - \frac{۳}{۵}(\frac{۲}{۳}\log\frac{۲}{۳} + \frac{1}{۳}\log\frac{1}{۳}) = ۰٫۹۵۰۹۷$

$\rightarrow H(Y^۴) - H(Y^۴|student) = ۰٫۰۱۹۹۷$

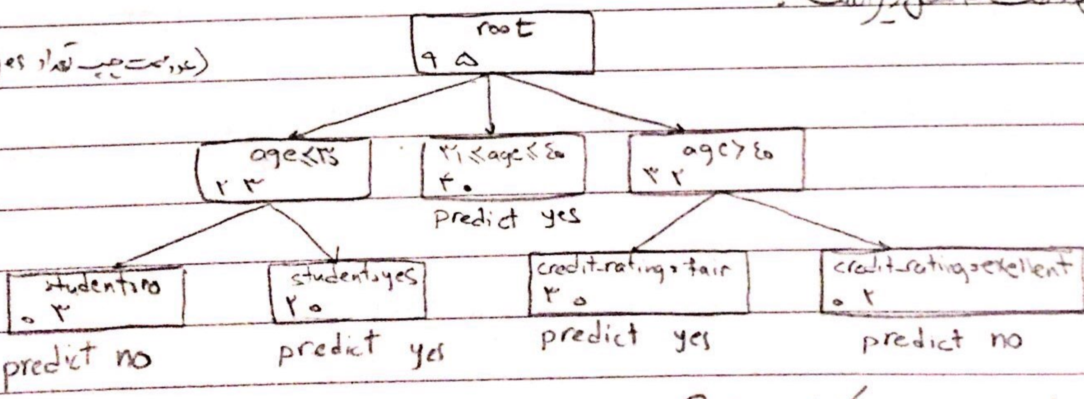
$H(Y^۴|income) = -\frac{۳}{۵}(\frac{۲}{۳}\log\frac{۲}{۳} + \frac{1}{۳}\log\frac{1}{۳}) - \frac{۲}{۵}(\frac{1}{۲}\log\frac{1}{۲} + \frac{1}{۲}\log\frac{1}{۲}) = ۰٫۹۵۰۹۷$

$H(Y^۴) - H(Y^۴|income) = ۰٫۰۱۹۹۷$



شکل درخت به شکل زیر است:

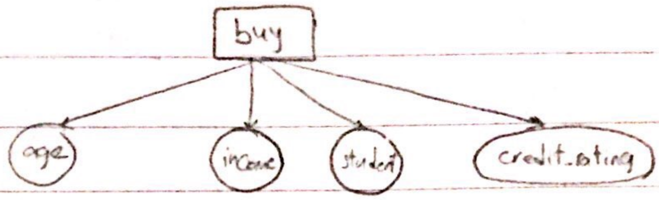
(عدد سمت چپ تعداد yesها و سمت راست تعداد noها است)



چون درخت را روی داده‌ها کشیدیم دقت آن برای داده‌های آموزش‌داده‌شده همان ٪۱۰۰ است.

s.a.m

ب)



ج)

buy	age	p(A|B)
yes	≤۳۰	۲/۹
yes	۳۱...۴۰	۴/۹
yes	>۴۰	۳/۹
No	≤۳۰	۳/۵
No	۳۱...۴۰	۰
No	>۴۰	۲/۵

buy	income	p(I|B)
yes	high	۲/۹
yes	medium	۴/۹
yes	low	۳/۹
NO	high	۲/۵
NO	medium	۲/۵
NO	low	۱/۵

buy	P(B)
yes	۹/۱۴
NO	۵/۱۴

buy	student	P(S|B)
yes	no	۳/۹
yes	yes	۶/۹
NO	no	۴/۵
NO	yes	۱/۵

buy	credit_rating	P(C|B)
yes	fair	۶/۹
yes	exellent	۳/۹
NO	fair	۲/۵
NO	exellent	۳/۵

ب) در این مدل فرض شده است که به شرط داشتن کلاس فیچرهای ما از هم مستقل اند و وقتی داده‌ی جدیدی می‌آید تمام جدول‌ها را ضرب می‌کنیم و تقسیم بر احتمال ضرب فیچرهای می‌کنیم تا احتمال p(buy | features) بدست می‌آید سپس احتمال هر کدام از buy ها بیشتر بود آن لیبل داده‌ی ما می‌شود (شکل بیزنت در بالا نشان داده شده است.)

د)

buy	student	P(S|B)
yes	no	۴/۱۱
yes	yes	۷/۱۱
no	no	۵/۷
no	yes	۲/۷

buy	credit_rating	P(C|B)
yes	fair	۷/۱۱
yes	exellent	۴/۱۱
NO	fair	۳/۷
NO	exellent	۴/۷

buy	P(B)
yes	۱۰/۱۴
NO	۴/۱۴

s.a.m

buy	age	p(A|B)
yes	≤۳۰	۳/۱۲
yes	۳۱...۴۰	۵/۱۲
yes	>۴۰	۴/۱۲
NO	≤۳۰	۴/۸
NO	۳۱...۴۰	۱/۸
NO	>۴۰	۳/۸

buy	income	p(I|B)
yes	high	۳/۱۲
yes	medium	۵/۱۲
yes	low	۴/۱۲
NO	high	۳/۸
NO	medium	۳/۸
NO	low	۲/۸

۵) ما با استفاده از naive bayes توانایی تولید داده های جدید را داریم زیرا که هم p(class) و هم p(feature|class) را داریم پس با انتخاب یک class خاص و سپس با استفاده از p(feature|class) ویژگی های مورد نظرمان را انتخاب می کنیم و به این صورت می توانیم با تولید اعداد رندوم بین (۰و۱) و با توجه به احتمالات، sample های مختلف برای Naive bayes مان تولید کنیم. ولی با decision tree نمی توان این کار را انجام داد زیرا درخت تصمیم سعی ندارد توزیع داده و ارتباط بین متغیرهای مختلف را مدل کند بلکه سعی می کند ارتباط بین متغیرها و class ها را از روی داده ها مدل کند یعنی ما در این مدل درخت تصمیم خودمان را براساس p(class|feature) می سازیم و جلو می بریم یعنی فقط میگوید براساس feature ها class ها چی می شود ولی نمی تواند خودش sample های مختلف تولید کند.

سوال ۲

$$\rightarrow H(Y|E) = \sum_{i=1}^{k} \overbrace{\frac{p_i+n_i}{p+n}}^{P(P_i)} \left[\frac{p_i}{p_i+n_i}\log_2\frac{p_i}{p_i+n_i} + \frac{n_i}{p_i+n_i}\log_2\frac{n_i}{p_i+n_i}\right]$$

اگر مقادیر همه k ها $\frac{p_i}{p_i+n_i}$ برابر باشند و مقدارشان α باشد

$$H(Y|E) = \alpha \underbrace{\sum \frac{p_i+n_i}{p+n}}_{1} = \alpha$$

$$\rightarrow H(Y|E) = \frac{p_i}{p_i+n_i}\log_2\frac{p_i}{p_i+n_i} + \frac{n_i}{p_i+n_i}\log_2\frac{n_i}{n_i+p_i}$$

$$\frac{p_1}{p_1+n_1} = \frac{p_2}{p_2+n_2} = \frac{1}{1+n_1/p_1} = \frac{1}{1+n_2/p_2} \Rightarrow \frac{n_1}{p_1} = \frac{n_2}{p_2} = \frac{n_1+n_2}{p_1+p_2} \rightarrow \frac{n_i}{p_i} = \frac{n}{p}$$

$$\rightarrow \frac{1}{1+n_i/p_i} = \frac{1}{1+n/p} \rightarrow \frac{p_i}{p_i+n_i} = \frac{p}{p+n} \rightarrow H(Y|E) = \frac{p}{p+n}\log_2\frac{p}{p+n} + \frac{n}{n+p}\log_2\frac{n}{p+n}$$

$$= H(Y) \rightarrow IG = 0$$

که این به این معنی است در صورتی که برای k های مختلف $\frac{p_i}{p_i+n_i}$ برابر باشد یعنی نسبت به آن ویژگی مستقل هستیم و داشتن آن ویژگی دانش ما را نسبت به مسئله زیاد نمی کند

s.a.m

حالا برای حالت عادی آن را اثبات می‌کنیم:

طبق نامساوی جنسن می‌دانیم اگر $\alpha_i \geq 0$ و $\sum_{i=1}^{k} \alpha_i = 1$ و $f(x)$ یک تابع محدب باشد آنگاه

$$\sum_{i=1}^{k} \alpha_i f(x) = f\left(\sum_{i=1}^{k} \alpha_i x\right)$$

حالا اگر f را بگیریم: $-x\log_2 x - (1-x)\log_2 (1-x) \rightarrow f'(x) = -\log_2 x + \log_2 (1-x)$

$$\rightarrow f''(x) = -\frac{1}{\ln 2 \times x(1-x)}$$

می‌دانیم x یک احتمال است در نتیجه بین ۰ و ۱ است

در نتیجه عبارت درون کسر همیشه مثبت است پس $f(x)$ محدب است.

$$\frac{p_1+n_1}{p+n} f\left(\frac{p_1}{p_1+n_1}\right) + \dots + \frac{p_k+n_k}{p+n} f\left(\frac{p_k}{p_k+n_k}\right) \leq f\left(\frac{p_1}{p_1+n_1}\frac{p_1+n_1}{p+n} + \dots + \frac{p_k}{p_k+n_k}\frac{p_k+n_k}{p+n}\right)$$

$$= f\left(\frac{p}{p+n}\right) \longrightarrow H(Y|E) \leq H(Y) \rightarrow IG \geq 0$$

سوال ۳:

$$P(Y=i|X) = \frac{P(Y=i)P(X|Y=i)}{P(Y=1)P(X|Y=1)+P(Y=2)P(X|Y=2)+P(Y=3)P(X|Y=3)} =$$

$$\frac{P(X|Y=i)}{P(X|Y=1)+P(X|Y=2)+P(X|Y=3)} = \frac{P(x_1,x_2|Y=i)}{P(x_1,x_2|Y=1)+P(x_1,x_2|Y=2)+P(x_1,x_2|Y=3)}$$

$$\longrightarrow P(x_1,x_2|Y=i) = \frac{1}{2\pi\sqrt{|\Sigma_i|}} e^{-\frac{1}{2}(X-\mu_i)^T\Sigma_i^{-1}(X-\mu_i)}$$

آ) $X=[-0.5, 0.5]$

$$P(x_1,x_2|Y=1) = \frac{1}{2\pi \times 0.7} e^{-\frac{1}{2}[-0.5,\,0.5]^T \begin{bmatrix} 10/7 & 0 \\ 0 & 10/7 \end{bmatrix} [0.5,\,0.5]}$$

که لیبل برابر است با $Y=1$

$$= \frac{e^{-5/14}}{2\pi \times 0.7}$$

$$P(x_1,x_2|Y=2) = \frac{1}{2\pi \times \sqrt{\frac{7}{100}}} \times e^{-\frac{1}{2}[-1.5,\,-0.5]^T \begin{bmatrix} 20/7 & -30/7 \\ -30/7 & 80/7 \end{bmatrix} [-1.5,\,-0.5]} = \frac{e^{-10/7}}{2\pi \sqrt{\frac{7}{100}}}$$

$$P(x_1,x_2|Y=3) = \frac{1}{2\pi \times \sqrt{\frac{13}{25}}} \times e^{-\frac{1}{2}[0.5,\,-0.5]^T \begin{bmatrix} 20/13 & -5/13 \\ -5/13 & 25/26 \end{bmatrix} [0.5,\,-0.5]} = \frac{e^{-\frac{95}{208}}}{2\pi \sqrt{\frac{13}{25}}}$$

$$P(Y=1|X) = \frac{1}{1+0.96422+0.8787} = \boxed{0.3591} \qquad P(Y=3|X) = \frac{1}{1.1380+1.0363+1} = 0.3155$$

$$P(Y=2|X) = \frac{1}{1.1035+1+0.9499} = 0.3254$$

s.a.m

ب) $X=[0.5, 0.5]$

به‌مثل آن ۲و۲ است

$$p(x_1,x_2|Y=1)=\frac{1}{2\pi\cdot 0.7}\, e^{-\frac{1}{2}[0.5,\,0.5]^T \begin{bmatrix} 10/7 & 0 \\ 0 & 10/7 \end{bmatrix} [0.5,\,0.5]}$$

$$=\frac{e^{-5/14}}{2\pi\times 0.7}$$

$$p(x_1,x_2|Y=2)=\frac{1}{2\pi\cdot\sqrt{\frac{7}{100}}}\, e^{-\frac{1}{2}[-0.5,\,-0.5]^T \begin{bmatrix} 20/7 & -30/7 \\ -30/7 & 80/7 \end{bmatrix} [-0.5,\,0.5]} = \frac{e^{-5/7}}{2\pi\times\sqrt{\frac{7}{100}}}$$

$$p(x_1,x_2|Y=3)=\frac{1}{2\pi\cdot\sqrt{\frac{13}{25}}}\times e^{-\frac{1}{2}[1.5,\,-0.5]^T \begin{bmatrix} 20/14 & -5/14 \\ -5/14 & 35/14 \end{bmatrix} [1.5,\,-0.5]} = \frac{e^{-\frac{35}{19}}}{2\pi\cdot\sqrt{\frac{14}{25}}}$$

$$p(Y=1|x)=\frac{1}{1+1.85119+0.155142}=0.3324$$

$$p(Y=2|x)=\frac{1}{0.54056+1+0.0841}=\boxed{0.6154}$$

$$p(Y=3|x)=\frac{1}{6.4242+11.8922+1}=0.0518$$