



تمرین سری اول

پاسخ سوال ۱

امیر حسین محمدی

۹۹۲۰۱۰۸۱



متغیر تصادفی X دارای توزیع

$$f_x(x) = \begin{cases} cx & 0 \leq x \leq 1 \\ c(2-x) & 1 \leq x \leq 2 \\ 0 & \text{elsewhere} \end{cases}$$

است. مقادیر



زیر را برای این توزیع بدست آورید.

- ضریب c
- CDF
- Expected Value



• ضرب c

برای محاسبه ی این ضرب از این نکته استفاده می کنیم که جمع احتمال های یک تابع توزیع احتمال یک است بنابراین داریم:

- $c(2 - x) \quad 1 \leq x \leq 2$ 
تبدیل به فرم
احتمالاتی (مساحت
زیر سطح نمودار)
- cx 
تبدیل به فرم
احتمالاتی (مساحت
زیر سطح نمودار)

$$\int_1^2 c(2 - x) dx + \int_0^1 cx dx = 1$$

$$2cx - \frac{x^2}{2} \Big|_1^2 + \frac{cx^2}{2} \Big|_0^1 = 1 \quad \Rightarrow \quad C=1$$

CDF •



برای محاسبه ی CDF، می دانیم:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

$$\begin{aligned} \bullet \quad CDF &= \int f_x(x) dx \xrightarrow{\text{blue arrow}} \int_1^2 (2-x) dx \xrightarrow{\text{green arrow}} 2x - \frac{x^2}{2} \Big|_1^2 = (4-2) - \left(2 - \frac{1}{2}\right) = \frac{1}{2} \\ &\xrightarrow{\text{blue arrow}} \int_0^1 x dx \xrightarrow{\text{green arrow}} \frac{x^2}{2} \Big|_0^1 = \left(\frac{1}{2} - 0\right) = \frac{1}{2} \end{aligned}$$

$$CDF = \int_1^2 (2-x) dx + \int_0^1 x dx = \frac{1}{2} + \frac{1}{2} = 1$$



برای محاسبه ی امید ریاضی (Expected Value)
داریم:

$$E[X] = \int_{\mathbb{R}} x f(x) dx,$$

$$\bullet \quad E[X] = \int_0^2 x f(x) dx \quad \longrightarrow \quad E[X] = \int_1^2 x(2-x) dx + \int_0^1 x(x) dx =$$

$$\left. \frac{x^3}{3} \right|_0^1 + x^2 - \left. \frac{x^3}{3} \right|_1^2 = 1 \quad \longrightarrow \quad E[X] = 1$$



متغیرهای تصادفی مستقل x_1, x_2, \dots, x_n با توزیع هندسی

$$p(X = k) = \theta(1 - \theta)^{k-1} \quad k \in \{1, 2, 3, \dots\}$$

را در نظر بگیرید. MLE را برای پارامتر θ به دست آورید.



با توجه به تابع MLE برای پارامتر θ داریم:

- $L(\theta) = \prod_{i=0}^n \theta(1 - \theta)^{k-1}$
- $L(\theta) = \theta^n (1 - \theta)^{\sum k - n}$

Ln گیری از طرفین:

$$\ln L(\theta) = n \ln \theta + \left(\sum k - n \right) \ln(1 - \theta)$$

مشتق جزئی نسبت به θ



- $\frac{\partial}{\partial \theta} \ln L(\theta) = n \frac{\partial}{\partial \theta} \ln \theta + (\sum k - n) \frac{\partial}{\partial \theta} \ln(1 - \theta)$
- $\frac{\partial}{\partial \theta} \ln L(\theta) = \frac{n}{\theta} - \frac{1}{1-\theta} (\sum k - n)$
- $\frac{\partial}{\partial \theta} \ln L(\theta) = 0$
- $\frac{n}{\theta} - \frac{1}{1-\theta} (\sum k - n) = 0$
- $\frac{n}{\theta} = \frac{1}{1-\theta} (\sum k - n)$
- $n(1 - \theta) = \theta (\sum k - n)$
- $n - n\theta = \theta \sum k - n\theta$
- $n = \theta \sum k$
- $\theta = \frac{n}{\sum k}$
- $\theta = \frac{1}{\frac{1}{n} \sum k}$



اثبات ماکسیمم بودن رابطه بدست آمده:

$$\theta = \frac{1}{\frac{1}{n} \sum k}$$

مشتق جزئی گیری از رابطه ی زیر :

$$\frac{\partial}{\partial \theta} \ln L(\theta) = \frac{n}{\theta} - \frac{1}{1-\theta} (\sum k - n)$$

$$\bullet \frac{\partial^2}{\partial \theta^2} \ln L(\theta) = n \frac{\partial}{\partial \theta} \left(\frac{1}{\theta} \right) - (\sum k - n) \frac{\partial}{\partial \theta} \left(\frac{1}{1-\theta} \right)$$

$$\bullet \frac{\partial^2}{\partial \theta^2} \ln L(\theta) = \frac{-n}{\theta^2} - \frac{(\sum k - n)}{(1-\theta)^2}$$

$$\bullet \frac{\partial^2}{\partial \theta^2} \ln L(\theta) = \frac{-n(1-\theta)^2 - \theta^2 (\sum k - n)}{\theta^2 (1-\theta)^2}$$

$$\bullet \frac{\partial^2}{\partial \theta^2} \ln L(\theta) = \frac{-n - n\theta^2 + 2n\theta - \theta^2 \sum k + n\theta^2}{\theta^2 (1-\theta)^2}$$



$$\bullet \frac{\partial^2}{\partial \theta^2} \ln L(\theta) = \frac{-n+2(\theta \sum k)\theta-\theta^2 \sum k}{\theta^2(1-\theta)^2}$$

$$\bullet \frac{\partial^2}{\partial \theta^2} \ln L(\theta) = \frac{-n+2\theta^2 \sum k-\theta^2 \sum k}{\theta^2(1-\theta)^2}$$

$$\bullet \frac{\partial^2}{\partial \theta^2} \ln L(\theta) = \frac{-n+\theta^2 \sum k}{\theta^2(1-\theta)^2} \xrightarrow{\text{بدست آوردیم}} \theta = \frac{n}{\sum k} \xrightarrow{\quad} \sum k = \frac{n}{\theta}$$

$$\bullet \frac{\partial^2}{\partial \theta^2} \ln L(\theta) = \frac{-n+\theta^2(\frac{n}{\theta})}{\theta^2(1-\theta)^2}$$

$$\bullet \frac{\partial^2}{\partial \theta^2} \ln L(\theta) = \frac{-n+n\theta}{\theta^2(1-\theta)^2}$$

$$\bullet \frac{\partial^2}{\partial \theta^2} \ln L(\theta) = \frac{n(-1+\theta)}{\theta^2(1-\theta)^2}$$

$$\bullet \frac{\partial^2}{\partial \theta^2} \ln L(\theta) = \frac{-n}{\theta^2(1-\theta)} \xrightarrow{\text{چون } \theta \text{ بین صفر و یک است و چون } n \text{ همیشه مثبت است بنابراین:}} \frac{\partial^2}{\partial \theta^2} \ln L(\theta) = \frac{-n}{\theta^2(1-\theta)} < 0 \xrightarrow{\text{بنابراین:}} \theta = \frac{1}{\frac{1}{n} \sum k}$$



تمرین سری اول

پاسخ سوال ۲

امیرحسین محمدی

۹۹۲۰۱۰۸۱



اگر A ماتریس مربعی، b یک متغیر، a و x بردارهای ستونی باشند، عبارت های زیر را ثابت نمایید:

$$\frac{da^T x}{dx} = \frac{dx^T a}{dx} = a^T \cdot$$

$$\frac{d(x^T a)^2}{dx} = 2x^T a a^T \cdot$$

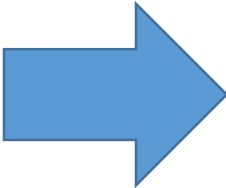
$$\frac{dx^T A x}{dx} = x^T (A + A^T) \cdot$$



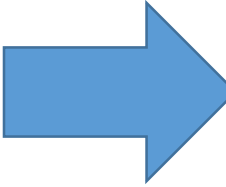
$$\frac{da^T x}{dx} = \frac{dx^T a}{dx} = a^T$$

• برای اثبات این رابطه ابتدا دو بردار زیر را طبق فرضیات مسئله در نظر می گیریم:

$$\begin{aligned} \text{بردار } a &= \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad \text{بردار } x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \end{aligned}$$

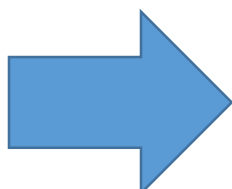


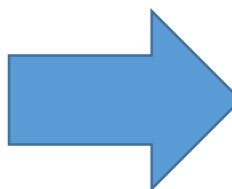
$$a^T x \Rightarrow [a_0 \ a_1 \ a_2 \ a_3] * \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = a_0 x_0 + a_1 x_1 + a_2 x_2 + a_3 x_3$$



$$x^T a \Rightarrow [x_0 \ x_1 \ x_2 \ x_3] * \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = a_0 x_0 + a_1 x_1 + a_2 x_2 + a_3 x_3$$



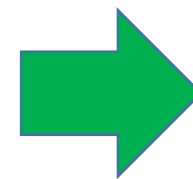

$$a^T x \Rightarrow [a_0 \ a_1 \ a_2 \ a_3] * \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = [a_0 x_0 \ a_1 x_1 \ a_2 x_2 \ a_3 x_3]$$


$$x^T a \Rightarrow [x_0 \ x_1 \ x_2 \ x_3] * \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = [a_0 x_0 \ a_1 x_1 \ a_2 x_2 \ a_3 x_3]$$

X یک بردار است
بنابراین داریم:

$$\bullet \frac{da^T x}{dx} = \frac{d(a_0 x_0 + a_1 x_1 + a_2 x_2 + a_3 x_3)}{dx} = [a_0 \ a_1 \ a_2 \ a_3]$$

$$\bullet \frac{dx^T a}{dx} = \frac{d(a_0 x_0 + a_1 x_1 + a_2 x_2 + a_3 x_3)}{dx} = [a_0 \ a_1 \ a_2 \ a_3]$$



$$\frac{da^T x}{dx} = \frac{dx^T a}{dx} = a^T$$

$$\frac{d(x^T a)^2}{dx} = 2x^T a a^T$$



ابتدا مشتق جزئی می گیریم:

$$\bullet \quad \frac{d(x^T a)^2}{dx} = 2(x^T a) \frac{d(x^T a)}{dx} \xrightarrow{\text{از قسمت یک داریم}} \frac{da^T x}{dx} = \frac{dx^T a}{dx} = a^T \xrightarrow{\text{بنابراین}} \frac{d(x^T a)^2}{dx} = 2x^T a a^T$$



$$\frac{dx^T A x}{dx} = x^T (A + A^T)$$

• فرضیات:

داریم:

$$\frac{\partial(x^T y)}{\partial x} = y$$

فرض می کنیم:

$$y = Ax$$

همچنین بر اساس قاعده ی زنجیری داریم:

$$\frac{d(f(x, y))}{dx} = \frac{\partial(f(x, y))}{\partial x} + \frac{d(y^T(x))}{dx} \frac{\partial(f(x, y))}{\partial y}$$

$$\begin{aligned} \frac{d(x^T A x)}{dx} &= \frac{d(x^T y)}{dx} = \frac{\partial(x^T y)}{\partial x} + \frac{d(y(x)^T)}{dx} \frac{\partial(x^T y)}{\partial y} = y + \frac{d(x^T A^T)}{dx} x = y + A^T x \\ &= (A + A^T)x \end{aligned}$$



مقدار ویژه و بردار ویژه را برای ماتریس HH^T را بدست آورید:

$$H = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$



داریم:

$$H = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

$$HH^T = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$$

محاسبه ی دترمینان برای محاسبه ی
مقدار ویژه:

$$|A - \lambda I| = 0 \quad \longrightarrow \quad |HH^T - \lambda I| = \begin{vmatrix} 5 - \lambda & 2 \\ 2 & 2 - \lambda \end{vmatrix} = 10 - 5\lambda - 2\lambda + \lambda^2$$

$$10 - 5\lambda - 2\lambda + \lambda^2 = 0 \quad \longrightarrow \quad \lambda_1 = 6, \lambda_2 = 1 \quad \text{مقدار ویژه}$$

بردار ویژه



$$\lambda_1 = 6 \rightarrow \begin{vmatrix} 5 - \lambda & 2 \\ 2 & 2 - \lambda \end{vmatrix} \rightarrow \begin{vmatrix} -1 & 2 \\ 2 & -4 \end{vmatrix} * \begin{vmatrix} v_1 \\ v_2 \end{vmatrix} = \begin{vmatrix} 0 \\ 0 \end{vmatrix} \rightarrow \begin{cases} -v_1 + 2v_2 = 0 \\ 2v_1 - 4v_2 = 0 \end{cases}$$

$$\rightarrow v_2 = \frac{v_1}{2}$$

$$\lambda_2 = 1 \rightarrow \begin{vmatrix} 5 - \lambda & 2 \\ 2 & 2 - \lambda \end{vmatrix} \rightarrow \begin{vmatrix} 4 & 2 \\ 2 & 1 \end{vmatrix} * \begin{vmatrix} v_1 \\ v_2 \end{vmatrix} \rightarrow \begin{cases} 4v_1 + 2v_2 = 0 \\ 2v_1 + v_2 = 0 \end{cases}$$

$$\rightarrow v_2 = -2v_1$$



تمرین سری اول

پاسخ سوال ۳

امیرحسین محمدی

۹۹۲۰۱۰۸۱



فرض کنید n داده آموزش به صورت $D = (x^1, y^1), \dots, (x^n, y^n)$ در اختیار داریم که هر کدام از x ها، d بعدی می باشد. می خواهیم از رگرسیون خطی با تابع هزینه SSE استفاده کنیم که به فرم زیر است:

$$j(w) = \sum_{i=0}^n (y^i - w^T x^i)^2$$

فرم w بهینه را بریا تابع هزینه بدست آورید.



- $j(w) = \sum_{i=0}^n (y^i - w^T x^i)^2$

- $j(w) = ||y - Xw||$

- $\bar{\bar{V}}_w j(w) = -2X^T(y - Xw)$

- $\bar{\bar{V}}_w j(w) = 0$

- $X^T X w = X^T y$

- $w = (X^T X)^{-1} X^T y$

$$X = \begin{bmatrix} 1 & x_1^{(1)} & \dots & x_d^{(1)} \\ 1 & x_1^{(2)} & \dots & x_d^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(n)} & \dots & x_d^{(n)} \end{bmatrix} \quad w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} \quad y = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(n)} \end{bmatrix}$$



- $j(w) = \sum_{i=0}^n (y^i - w^T x^i)^2$



بازنویسی
عبارت بالا به
فرم دیگر

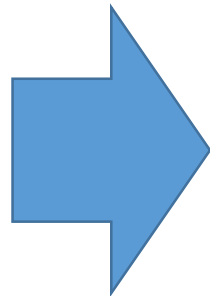
- $Xw = y$



جهت معکوس
کردن X ، ضرب
طرفین در X^T

تا ماتریس مربعی
شود

- $X^T X w = X^T y$



$$X = \begin{bmatrix} 1 & x_1^{(1)} & \dots & x_d^{(1)} \\ 1 & x_1^{(2)} & \dots & x_d^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(n)} & \dots & x_d^{(n)} \end{bmatrix} \quad w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} \quad y = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(n)} \end{bmatrix}$$

- $w = (X^T X)^{-1} X^T y$



مشکلات استفاده مستقیم از رابطه قسمت قبل را بیان کنید و برای آن ها راه حلی ارائه دهید.



❖ از مهمترین مشکلات استفاده از این رابطه ی می توان به سربار محاسباتی بالای آن اشاره کرد. می دانیم مرتبه ی زمانی ضرب دو ماتریس $O(n^3)$ و اگر n از یک حد معین مثلا 10000 باشند این زمان اجرا بسیار قابل ملاحظه بوده و به نوعی گلوگاه محسوب می شود و با توجه به اینکه عمده ی مسائل از حجم قابل توجهی ویژگی برخوردار هستند بنابراین استفاده از رابطه ی مستقیم به عنوان کی مزین محسوب نمی شود. از جمله راه حل هایی که برای رفع این مشکل ارائه می شود کاهش ابعاد و ویژگی هاست که خود این مسئله می تواند و به کاهش دقت منجر شود. از دیگر راه حل های ارائه شده برای رفع این مشکل استفاده از گرادیان کاهشی برای محاسبه ی بهترین w برای مسئله ی مورد نظر است.

❖ از دیگر مشکلات این روش، می توان به معکوس ناپذیری $X^T X$ اشاره کرد که دلیل آن افزونگی ویژگی ها (وابستگی خطی بین آنها) و زیاد بودن ویژگی هاست. که از راه حل های آن می توان به استفاده از ماتریس شبه معکوس اشار کرد.

$$w = \text{pinv}(X.T @ X) @ X.T @ y$$

از دیگر راه حل ها برای رفع مشکل معکوس ناپذیری می توان به حذف برخی از ویژگی ها با استفاده از تنظیم اشاره کرد.



اگر به تابع هزینه جمله منظم ساز $||w^2||$ را بیفزاییم، فرم بسته پاسخ بهینه w را به دست آورید.



- $j(w) = \sum_{i=0}^n (y^i - w^T x^i)^2 + ||w||^2$

- $j(w) = \sum_{i=0}^n (y^i - w^T x^i)^2 + \lambda \times \sum_{j=0}^n w_j^2 = (Xw - y)^T (Xw - y) + \lambda w^T w$

- $\frac{\partial J}{\partial w} = X^T (Xw - y) + \lambda w = X^T Xw - X^T y + \lambda w =$
 $(X^T X + \lambda I)w - X^T y = 0$

- $(X^T X + \lambda I)w = X^T y$

- $w = \underbrace{(X^T X + \lambda I)^{-1}}_{\lambda > 0} X^T y$



رگرسیون خطی وزن دار، تعمیمی از رگرسیون خطی است که در آن به هر یک از داده ها، وزنی اختصاص داده می شود:

$$j(w) = \sum_{i=0}^n f_i (y^i - w^T x^i)^2$$

فرم بهینه w بریا این تابع هزینه بدست آورید.



- $j(w) = \sum_{i=0}^n f_i(y^i - w^T x^i)^2$
- $\sum_{i=0}^n f_i(y^i - w^T x^i)^2 = \sum_{i=0}^n f_i(w^T x^i - y^i)^2 = (Y - XW)^T F (Y - XW)$
- $\frac{\partial J}{\partial w} = -X^T F (Y - XW) = -X^T F Y + X^T F X W$
- $X^T F X W = X^T F Y$
- $W = (X^T F X)^{-1} X^T F Y$



تمرین سری اول

پاسخ سوال ۴

امیرحسین محمدی

۹۹۲۰۱۰۸۱



درخت تصمیم را به صورت دستی بر روی داده های یادگیری آموزش دهید و دقت دسته بند را بر روی داده ای تست بدست آورید.

برای سوالات زیر در صورتی که در درخت تصمیم تغییری ایجاد شود، آن را مجدداً رسم نموده و دقت دسته بند را برای داده های تست حساب نمایید. چنانچه در درخت تصمیم تغییری ایجاد نشود دلیل آن را به طور کامل توضیح دهید.



دیتاست اصلی

نام شخص	ژن ۱	ژن ۲	ژن ۳	ژن ۴	ژن ۵	بیمار بودن
X1	10	30	زیاد	۱	۴۰	بله
X2	101	31	متوسط	۰	۱۴	خیر
X3	20	70	زیاد	۲	۳۳	بله
X4	200	35	کم	۳	۱۰	خیر
X5	37	90	متوسط	۲۲	۱۱	بله
X6	25	25	متوسط	۱۹	۵۰	خیر
X7	40	75	زیاد	۳۰	۵۵	خیر
X8	50	45	کم	۴	۱۲	بله
X9	120	80	متوسط	۲۳	۸	خیر



دیتاست پس از تغییر بر اساس فرضیات مسئله

نام شخص	ژن ۱	ژن ۲	ژن ۳	ژن ۴	ژن ۵	بیمار بودن
X1	کم	کم	زیاد	کم	زیاد	بله
X2	زیاد	کم	متوسط	کم	کم	خیر
X3	کم	زیاد	زیاد	کم	زیاد	بله
X4	زیاد	کم	کم	کم	کم	خیر
X5	کم	زیاد	متوسط	زیاد	کم	بله
X6	کم	کم	متوسط	زیاد	زیاد	خیر
X7	کم	زیاد	زیاد	زیاد	زیاد	خیر
X8	کم	کم	کم	کم	کم	بله
X9	زیاد	زیاد	متوسط	زیاد	کم	خیر

مقادیر: کم، زیاد → ویژگی: ژن ۱



$S[4+ \quad 5-]$

$$\text{Entropy}(s) = -\frac{4}{9} * \log_2 \frac{4}{9} - \frac{5}{9} * \log_2 \frac{5}{9} = 0.989$$

$S_{\text{کم}}[۴+ \quad ۲-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{4}{6} * \log_2 \frac{4}{6} - \frac{2}{6} * \log_2 \frac{2}{6} = 0.916$$

$S_{\text{زیاد}}[۰+ \quad ۳-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{0}{3} * \log_2 \frac{0}{3} - \frac{3}{3} * \log_2 \frac{3}{3} = 0$$

- $\text{Gain}(s, \text{ژن ۱}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۱}) = \text{entropy}(s) - \frac{6}{9} * \text{entropy}(s_{\text{کم}}) - \frac{3}{9} * \text{entropy}(s_{\text{زیاد}}) = 0.379$

مقادیر: کم، زیاد → ویژگی: ژن ۲



$S[4+ \quad 5-]$

$$\text{Entropy}(s) = -\frac{4}{9} * \log_2 \frac{4}{9} - \frac{5}{9} * \log_2 \frac{5}{9} = 0.989$$

$S_{\text{کم}}[۲+ \quad ۳-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{2}{5} * \log_2 \frac{2}{5} - \frac{3}{5} * \log_2 \frac{3}{5} = 0.96$$

$S_{\text{زیاد}}[۲+ \quad ۲-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{4} * \log_2 \frac{2}{4} - \frac{2}{4} * \log_2 \frac{2}{4} = 1$$

- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \frac{5}{9} * \text{entropy}(s_{\text{کم}}) - \frac{4}{9} * \text{entropy}(s_{\text{زیاد}}) = 0.019$



مقادیر: کم، زیاد و متوسط \longrightarrow ویژگی: ۳ زن

$S[4+ \quad 5-]$

$$\text{Entropy}(s) = -\frac{4}{9} * \log_2 \frac{4}{9} - \frac{5}{9} * \log_2 \frac{5}{9} = 0.989$$

$S_{\text{کم}}[1+ \quad 1-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

$S_{\text{زیاد}}[2+ \quad 1-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.913$$

$S_{\text{متوسط}}[1+ \quad 3-]$

$$\text{Entropy}(s_{\text{متوسط}}) = -\frac{1}{4} * \log_2 \frac{1}{4} - \frac{3}{4} * \log_2 \frac{3}{4} = 0.811$$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \frac{2}{9} * \text{entropy}(s_{\text{کم}}) - \frac{3}{9} * \text{entropy}(s_{\text{زیاد}}) - \frac{4}{9} * \text{entropy}(s_{\text{متوسط}}) = 0.103$

مقادیر: کم، زیاد → ویژگی: ژن ۴



$S[4+ \quad 5-]$

$$\text{Entropy}(s) = -\frac{4}{9} * \log_2 \frac{4}{9} - \frac{5}{9} * \log_2 \frac{5}{9} = 0.989$$

$S_{\text{کم}}[3+ \quad 2-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{3}{5} * \log_2 \frac{3}{5} - \frac{2}{5} * \log_2 \frac{2}{5} = 0.96$$

$S_{\text{زیاد}}[1+ \quad 3-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{4} * \log_2 \frac{1}{4} - \frac{3}{4} * \log_2 \frac{3}{4} = 0.811$$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \frac{5}{9} * \text{entropy}(s_{\text{کم}}) - \frac{4}{9} * \text{entropy}(s_{\text{زیاد}}) = 0.091$

مقادیر: کم، زیاد → ویژگی: ژن ۵



$S[4+ \quad 5-]$

$$\text{Entropy}(s) = -\frac{4}{9} * \log_2 \frac{4}{9} - \frac{5}{9} * \log_2 \frac{5}{9} = 0.989$$

$S_{\text{کم}}[۲+ \quad ۳-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{2}{5} * \log_2 \frac{2}{5} - \frac{3}{5} * \log_2 \frac{3}{5} = 0.969$$

$S_{\text{زیاد}}[۲+ \quad ۲-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{4} * \log_2 \frac{2}{4} - \frac{2}{4} * \log_2 \frac{2}{4} = 0.007$$

- $\text{Gain}(s, \text{ژن ۵}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$

- $\text{Gain}(s, \text{ژن ۵}) = \text{entropy}(s) - \frac{5}{9} * \text{entropy}(s_{\text{کم}}) - \frac{4}{9} * \text{entropy}(s_{\text{زیاد}}) = 0.091$

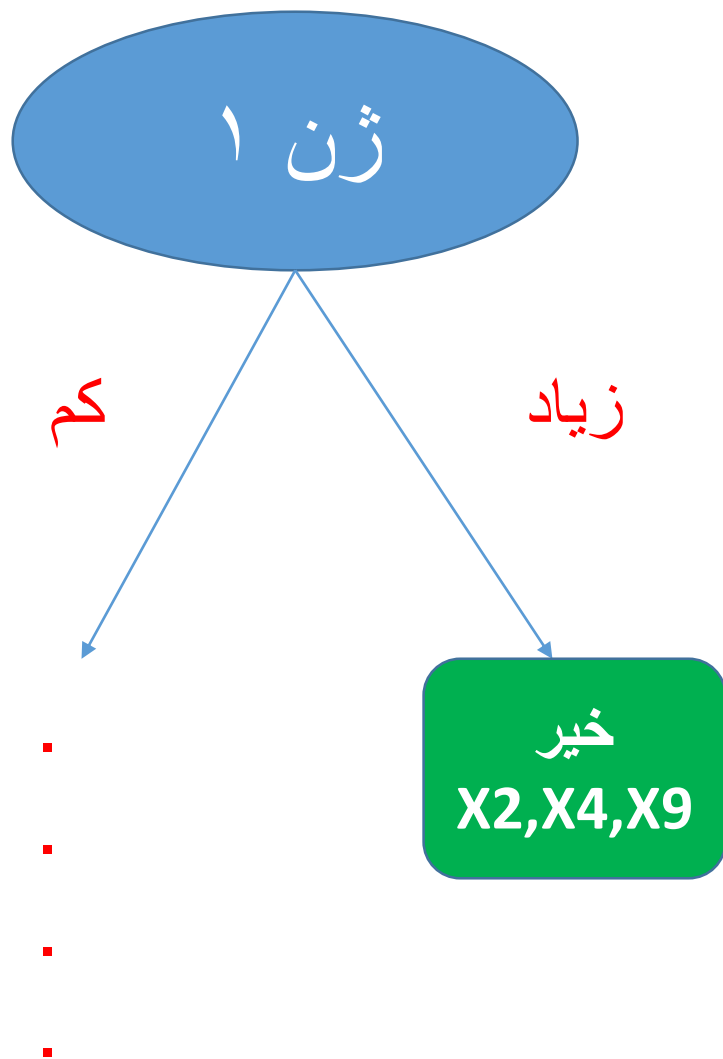


- $Gain(s, \text{ژن}^1) = entropy(s) - \frac{6}{9} * entropy(s_{\text{کم}}) - \frac{3}{9} * entropy(s_{\text{زیاد}}) = 0.379$
- $Gain(s, \text{ژن}^2) = entropy(s) - \frac{5}{9} * entropy(s_{\text{کم}}) - \frac{4}{9} * entropy(s_{\text{زیاد}}) = 0.019$
- $Gain(s, \text{ژن}^3) = entropy(s) - \frac{2}{9} * entropy(s_{\text{کم}}) - \frac{3}{9} * entropy(s_{\text{زیاد}}) - \frac{4}{9} * entropy(s_{\text{متوسط}}) = 0.103$
- $Gain(s, \text{ژن}^4) = entropy(s) - \frac{5}{9} * entropy(s_{\text{کم}}) - \frac{4}{9} * entropy(s_{\text{زیاد}}) = 0.091$
- $Gain(s, \text{ژن}^5) = entropy(s) - \frac{5}{9} * entropy(s_{\text{کم}}) - \frac{4}{9} * entropy(s_{\text{زیاد}}) = 0.091$



$$Gain(s, \text{ژن}^1) = entropy(s) - \frac{6}{9} * entropy(s_{\text{کم}}) - \frac{3}{9} * entropy(s_{\text{زیاد}}) = 0.379$$

انتخاب شد





دیتاست پس از انتخاب ژن ۱

نام شخص	ژن ۲	ژن ۳	ژن ۴	ژن ۵	بیمار بودن
X1	کم	زیاد	کم	زیاد	بله
X3	زیاد	زیاد	کم	زیاد	بله
X5	زیاد	متوسط	زیاد	کم	بله
X6	کم	متوسط	زیاد	زیاد	خیر
X7	زیاد	زیاد	زیاد	زیاد	خیر
X8	کم	کم	کم	کم	بله

مقادیر: کم، زیاد → ویژگی: ژن ۲



$S[4+ \quad 2-]$

$$\text{Entropy}(s) = -\frac{4}{6} * \log_2 \frac{4}{6} - \frac{2}{6} * \log_2 \frac{2}{6} = 0.916$$

$S_{\text{کم}}[۲+ \quad ۱-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.916$$

$S_{\text{زیاد}}[۲+ \quad ۱-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.916$$

- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \frac{3}{6} * \text{entropy}(s_{\text{کم}}) - \frac{3}{6} * \text{entropy}(s_{\text{زیاد}}) = 0$



مقادیر: کم، زیاد و متوسط \longrightarrow ویژگی: ژن ۳

$S[4+ \quad 2-]$

$$\text{Entropy}(s) = -\frac{4}{9} * \log_2 \frac{4}{9} - \frac{5}{9} * \log_2 \frac{5}{9} = 0.989$$

$S_{\text{کم}}[1+ \quad 0-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{1}{1} * \log_2 \frac{1}{1} - \frac{0}{1} * \log_2 \frac{0}{1} = 0$$

$S_{\text{زیاد}}[2+ \quad 1-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.916$$

$S_{\text{متوسط}}[1+ \quad 1-]$

$$\text{Entropy}(s_{\text{متوسط}}) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

- $\text{Gain}(s, \text{ژن ۳}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۳}) = \text{entropy}(s) - \frac{1}{6} * \text{entropy}(s_{\text{کم}}) - \frac{3}{6} * \text{entropy}(s_{\text{زیاد}}) - \frac{2}{6} * \text{entropy}(s_{\text{متوسط}}) = 0.125$

مقادیر: کم، زیاد \longrightarrow ویژگی: ژن ۴



$S[4+ \quad 2-]$

$$\text{Entropy}(s) = -\frac{4}{6} * \log_2 \frac{4}{6} - \frac{2}{6} * \log_2 \frac{2}{6} = 0.916$$

$S_{\text{کم}}[3+ \quad 0-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{3}{3} * \log_2 \frac{3}{3} - \frac{0}{3} * \log_2 \frac{0}{3} = 0$$

$S_{\text{زیاد}}[1+ \quad 2-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{3} * \log_2 \frac{1}{3} - \frac{2}{3} * \log_2 \frac{2}{3} = 0.916$$

- $\text{Gain}(s, f_{\text{ژن}}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, f_{\text{ژن}}) = \text{entropy}(s) - \frac{3}{6} * \text{entropy}(s_{\text{کم}}) - \frac{3}{6} * \text{entropy}(s_{\text{زیاد}}) = 0.458$

مقادیر: کم، زیاد \longrightarrow ویژگی: زن ۵



$S[4+ \quad 2-]$

$$\text{Entropy}(s) = -\frac{4}{6} * \log_2 \frac{4}{6} - \frac{2}{6} * \log_2 \frac{2}{6} = 0.916$$

$S_{\text{کم}}[0+ \quad 2-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{0}{2} * \log_2 \frac{0}{2} - \frac{2}{2} * \log_2 \frac{2}{2} = 0$$

$S_{\text{زیاد}}[2+ \quad 2-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{4} * \log_2 \frac{2}{4} - \frac{2}{4} * \log_2 \frac{2}{4} = 1$$

- $\text{Gain}(s, \Delta_{\text{زن}}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \Delta_{\text{زن}}) = \text{entropy}(s) - \frac{2}{6} * \text{entropy}(s_{\text{کم}}) - \frac{4}{6} * \text{entropy}(s_{\text{زیاد}}) = 0.249$

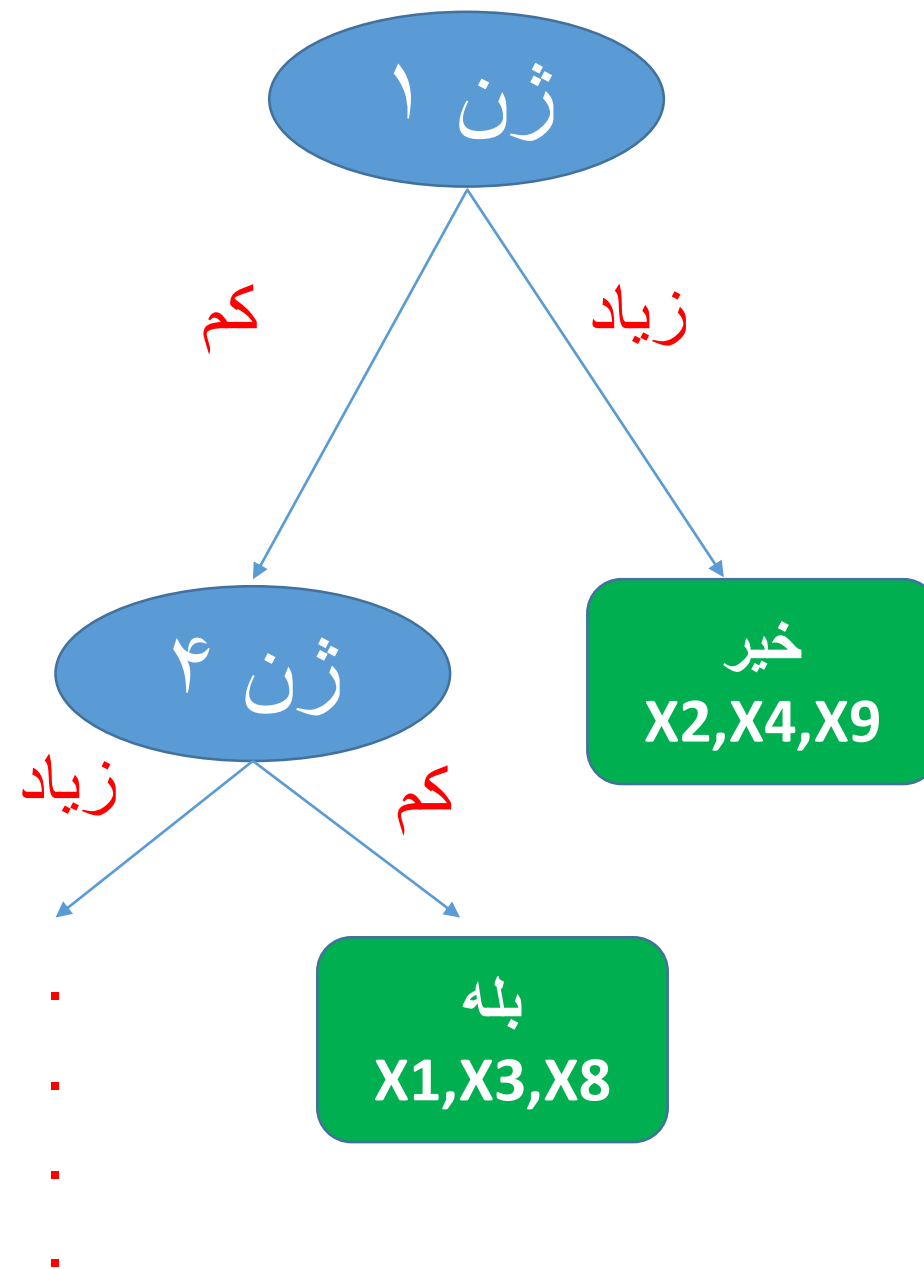


- $Gain(s, \text{ژن ۲}) = entropy(s) - \frac{3}{6} * entropy(s_{\text{کم}}) - \frac{3}{6} * entropy(s_{\text{زیاد}}) = 0$
- $Gain(s, \text{ژن ۳}) = entropy(s) - \frac{1}{6} * entropy(s_{\text{کم}}) - \frac{3}{6} * entropy(s_{\text{زیاد}}) - \frac{2}{6} * entropy(s_{\text{متوسط}}) = 0.125$
- $Gain(s, \text{ژن ۴}) = entropy(s) - \frac{3}{6} * entropy(s_{\text{کم}}) - \frac{3}{6} * entropy(s_{\text{زیاد}}) = 0.458$
- $Gain(s, \text{ژن ۵}) = entropy(s) - \frac{2}{6} * entropy(s_{\text{کم}}) - \frac{4}{6} * entropy(s_{\text{زیاد}}) = 0.249$



$$Gain(s, \text{ژن ۴}) = entropy(s) - \frac{3}{6} * entropy(s_{\text{کم}}) - \frac{3}{6} * entropy(s_{\text{زیاد}}) = 0.458$$

انتخاب شد





دیتاست پس از انتخاب ژن ۴

نام شخص	ژن ۲	ژن ۳	ژن ۵	بیمار بودن
X5	زیاد	متوسط	کم	بله
X6	کم	متوسط	زیاد	خیر
X7	زیاد	زیاد	زیاد	خیر

مقادیر: کم، زیاد → ویژگی: ژن ۲



$S[1+ \quad 2-]$

$$\text{Entropy}(s) = -\frac{1}{3} * \log_2 \frac{1}{3} - \frac{2}{3} * \log_2 \frac{2}{3} = 0.916$$

$S_{\text{کم}}[0+ \quad 1-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{0}{1} * \log_2 \frac{0}{1} - \frac{1}{1} * \log_2 \frac{1}{1} = 0$$

$S_{\text{زیاد}}[1+ \quad 1-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \frac{1}{3} * \text{entropy}(s_{\text{کم}}) - \frac{2}{3} * \text{entropy}(s_{\text{زیاد}}) = 0.249$



مقادیر: کم، زیاد و متوسط \longrightarrow ویژگی: ژن ۳

S[1+ 2-]

$$\text{Entropy}(s) = -\frac{1}{3} * \log_2 \frac{1}{3} - \frac{2}{3} * \log_2 \frac{2}{3} = 0.916$$

S[۰+ ۱-] زیاد

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{0}{1} * \log_2 \frac{0}{1} - \frac{1}{1} * \log_2 \frac{1}{1} = 0$$

S[۱+ ۱-] متوسط

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \frac{1}{3} * \text{entropy}(s_{\text{زیاد}}) - \frac{2}{3} * \text{entropy}(s_{\text{متوسط}}) = 0.249$

مقادیر: کم، زیاد → ویژگی: ژن ۵



S[1+ 2-]

$$\text{Entropy}(s) = -\frac{1}{3} * \log_2 \frac{1}{3} - \frac{2}{3} * \log_2 \frac{2}{3} = 0.916$$

S_{کم}[۱+ ۰-]

$$\text{Entropy}(s_{\text{کم}}) = -\frac{0}{1} * \log_2 \frac{0}{1} - \frac{1}{1} * \log_2 \frac{1}{1} = 0$$

S_{زیاد}[۰+ ۲-]

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{0}{2} * \log_2 \frac{0}{2} - \frac{2}{2} * \log_2 \frac{2}{2} = 0$$

- $\text{Gain}(s, \text{ژن ۵}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۵}) = \text{entropy}(s) - \frac{1}{3} * \text{entropy}(s_{\text{کم}}) - \frac{2}{3} * \text{entropy}(s_{\text{زیاد}}) = 0.916$



- $Gain(s, ۲ن) = entropy(s) - \frac{1}{3} * entropy(s_{کم}) - \frac{2}{3} * entropy(s_{زیاد}) = 0.249$

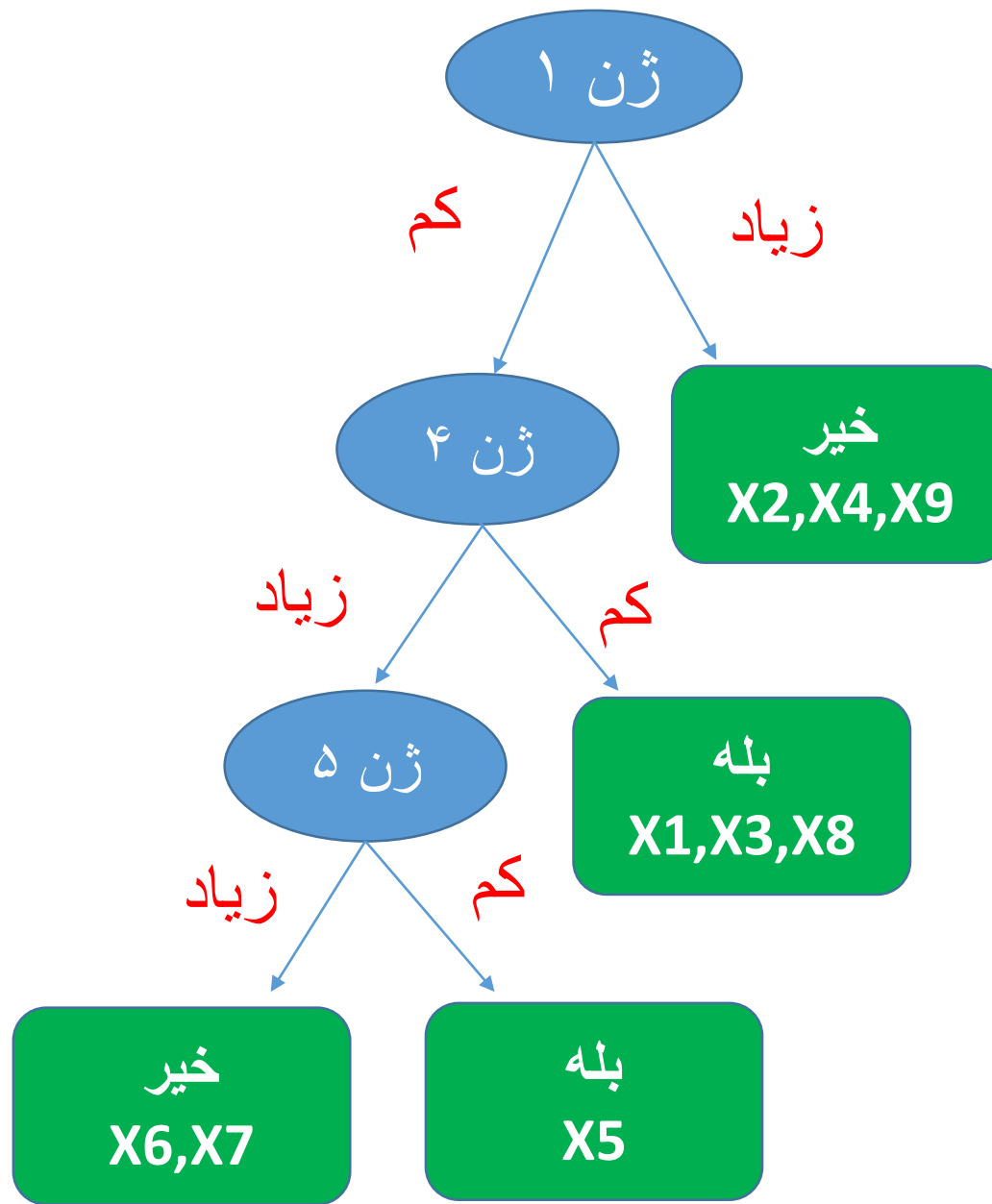
- $Gain(s, ۳ن) = entropy(s) - \frac{1}{3} * entropy(s_{زیاد}) - \frac{2}{3} * entropy(s_{متوسط}) = 0.249$

- $Gain(s, ۵ن) = entropy(s) - \frac{1}{3} * entropy(s_{کم}) - \frac{2}{3} * entropy(s_{زیاد}) = 0.916$



$$Gain(s, ۵ن) = entropy(s) - \frac{1}{3} * entropy(s_{کم}) - \frac{2}{3} * entropy(s_{زیاد}) = 0.916$$

انتخاب شد





داده های تست

نام شخص	ژن ۱	ژن ۲	ژن ۳	ژن ۴	ژن ۵	بیمار بودن
Y1	42	20	زیاد	۰	۳۷	بله
Y2	139	85	کم	۱۵	۵۰	خیر
Y3	59	41	زیاد	۴	۵۰	بله
Y4	22	37	متوسط	۱	۸۰	خیر
Y5	76	65	متوسط	۲	۶۰	بله
Y6	30	83	زیاد	۱	۱۷	خیر



تغییر داده های تست با توجه به فرضیات مسئله

نام شخص	ژن ۱	ژن ۲	ژن ۳	ژن ۴	ژن ۵	بیمار بودن
Y1	کم	کم	زیاد	کم	زیاد	بله
Y2	زیاد	زیاد	کم	زیاد	زیاد	خیر
Y3	کم	کم	زیاد	کم	زیاد	بله
Y4	کم	کم	متوسط	کم	زیاد	خیر
Y5	کم	زیاد	متوسط	کم	زیاد	بله
Y6	کم	زیاد	زیاد	کم	کم	خیر



پیش بینی	بیمار بودن (لیبل واقعی)	نام شخص
بله	بله	Y1
خیر	خیر	Y2
خیر	بله	Y3
خیر	خیر	Y4
خیر	بله	Y5
خیر	خیر	Y6

دقت = ۶۶/۶۶
درصد



اگر ژن شماره ۲ وجود نداشت تغییری در درخت ایجاد می شد؟ در صورتی که
ژن شماره ۳ نباشد آیا باز هم تغییری در درخت تصمیم ایجاد نمی شود؟



اگر ژن شماره ی دو و ژن شماره سه در این دیتا ست وجود نداشت، هیچ تغییری در درخت به وجود نمی آمد. زیرا همانطور که در قسمت یک مشاهده شد، درخت تصمیم تشکیل شده مستقل از ژن دو و سه است و صرفا به ژن یک و چهار و پنج وابستگی دارد. دلیل استفاده نکردن از ژن دو و سه در درخت تصمیم **information gain** بسیار کم بدست آمده در هر مرحله است. دلیل **information gain** کم این دو ژن پراکندگی زیاد بین این دو ژن است به این معنا که مثلا اگر ۹ داده در مرحله ی آموزش وجود داشته باشد، تفکیک داده ها بر اساس لیبل ها معین شده تقریبا ۵۰ درصد به ازای هر لیبل است و این مسئله سبب می شود که **information gain** در هر مرحله کمترین میزان را به خود بگیرد (زیرا ۵۰ ۵۰ شدن داده ها کمکی به تصمیم گیری نمی کند) و این مسئله در ژن دو و سه وجود دارد. همچنین از دیگر دلایل انتخاب نشدن ژن سه می توان به بیشتر بودن تعداد ویژگی ها در این ژن اشاره کرد که این مسئله به همراه ۵۰ ۵۰ شدن لیبل ها به نوعی خود سبب کاهش **information gain** می شود.



اگر فرد X5 به عنوان فرد سالم در نظر گرفته می شد، تغییری در درخت تصمیم رخ می داد؟



دیتاست پس از تغییر بر اساس فرضیات مسئله

نام شخص	ژن ۱	ژن ۲	ژن ۳	ژن ۴	ژن ۵	بیمار بودن
X1	کم	کم	زیاد	کم	زیاد	بله
X2	زیاد	کم	متوسط	کم	کم	خیر
X3	کم	زیاد	زیاد	کم	زیاد	بله
X4	زیاد	کم	کم	کم	کم	خیر
X5	کم	زیاد	متوسط	زیاد	کم	خیر
X6	کم	کم	متوسط	زیاد	زیاد	خیر
X7	کم	زیاد	زیاد	زیاد	زیاد	خیر
X8	کم	کم	کم	کم	کم	بله
X9	زیاد	زیاد	متوسط	زیاد	کم	خیر

مقادیر: کم، زیاد → ویژگی: ژن ۱



$S[3+ \quad 6-]$

$$\text{Entropy}(s) = -\frac{3}{9} * \log_2 \frac{3}{9} - \frac{6}{9} * \log_2 \frac{6}{9} = 0.916$$

$S_{\text{کم}}[۳+ \quad ۳-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{3}{6} * \log_2 \frac{3}{6} - \frac{3}{6} * \log_2 \frac{3}{6} = 1$$

$S_{\text{زیاد}}[۰+ \quad ۳-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{0}{3} * \log_2 \frac{0}{3} - \frac{3}{3} * \log_2 \frac{3}{3} = 0$$

- $\text{Gain}(s, \text{ژن ۱}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۱}) = \text{entropy}(s) - \frac{6}{9} * \text{entropy}(s_{\text{کم}}) - \frac{3}{9} * \text{entropy}(s_{\text{زیاد}}) = 0.249$

مقادیر: کم، زیاد → ویژگی: ژن ۲



S[3+ 6-]

$$\text{Entropy}(s) = -\frac{3}{9} * \log_2 \frac{3}{9} - \frac{6}{9} * \log_2 \frac{6}{9} = 0.916$$

S_{کم}[۲+ ۳-]

$$\text{Entropy}(s_{\text{کم}}) = -\frac{2}{5} * \log_2 \frac{2}{5} - \frac{3}{5} * \log_2 \frac{3}{5} = 0.969$$

S_{زیاد}[۱+ ۳-]

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{4} * \log_2 \frac{1}{4} - \frac{3}{4} * \log_2 \frac{3}{4} = 0.811$$

- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \frac{5}{9} * \text{entropy}(s_{\text{کم}}) - \frac{4}{9} * \text{entropy}(s_{\text{زیاد}}) = 0.018$



مقادیر: کم، زیاد، متوسط \longrightarrow ویژگی: ژن ۳

$S[3+ \quad 6-]$

$$\text{Entropy}(s) = -\frac{3}{9} * \log_2 \frac{3}{9} - \frac{6}{9} * \log_2 \frac{6}{9} = 0.916$$

$S_{\text{کم}}[1+ \quad 1-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

$S_{\text{زیاد}}[2+ \quad 1-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.916$$

$S_{\text{متوسط}}[0+ \quad 4-]$

$$\text{Entropy}(s_{\text{متوسط}}) = -\frac{0}{4} * \log_2 \frac{0}{4} - \frac{4}{4} * \log_2 \frac{4}{4} = 0$$

- $\text{Gain}(s, \text{ژن ۳}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$

- $\text{Gain}(s, \text{ژن ۳}) = \text{entropy}(s) - \frac{2}{9} * \text{entropy}(s_{\text{کم}}) - \frac{3}{9} * \text{entropy}(s_{\text{زیاد}}) - \frac{4}{9} * \text{entropy}(s_{\text{متوسط}})$
 $= 0.389$

مقادیر: کم، زیاد → ویژگی: ژن ۴



S[3+ 6-]

$$\text{Entropy}(s) = -\frac{3}{9} * \log_2 \frac{3}{9} - \frac{6}{9} * \log_2 \frac{6}{9} = 0.916$$

S_{کم}[۳+ ۲-]

$$\text{Entropy}(s_{\text{کم}}) = -\frac{3}{5} * \log_2 \frac{3}{5} - \frac{2}{5} * \log_2 \frac{2}{5} = 0.969$$

S_{زیاد}[۰+ ۴-]

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{0}{4} * \log_2 \frac{0}{4} - \frac{4}{4} * \log_2 \frac{4}{4} = 0$$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \frac{5}{9} * \text{entropy}(s_{\text{کم}}) - \frac{4}{9} * \text{entropy}(s_{\text{زیاد}}) = 0.378$



مقادیر: کم، زیاد → ویژگی: ژن ۵

S[3+ 6-]

$$\text{Entropy}(s) = -\frac{3}{9} * \log_2 \frac{3}{9} - \frac{6}{9} * \log_2 \frac{6}{9} = 0.916$$

S[۱+ ۴-] کم

$$\text{Entropy}(s_{\text{کم}}) = -\frac{1}{5} * \log_2 \frac{1}{5} - \frac{4}{5} * \log_2 \frac{4}{5} = 0.72$$

S[۲+ ۲-] زیاد

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{4} * \log_2 \frac{2}{4} - \frac{2}{4} * \log_2 \frac{2}{4} = 1$$

- $\text{Gain}(s, \text{ژن ۵}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۵}) = \text{entropy}(s) - \frac{5}{9} * \text{entropy}(s_{\text{کم}}) - \frac{4}{9} * \text{entropy}(s_{\text{زیاد}}) = 0.071$

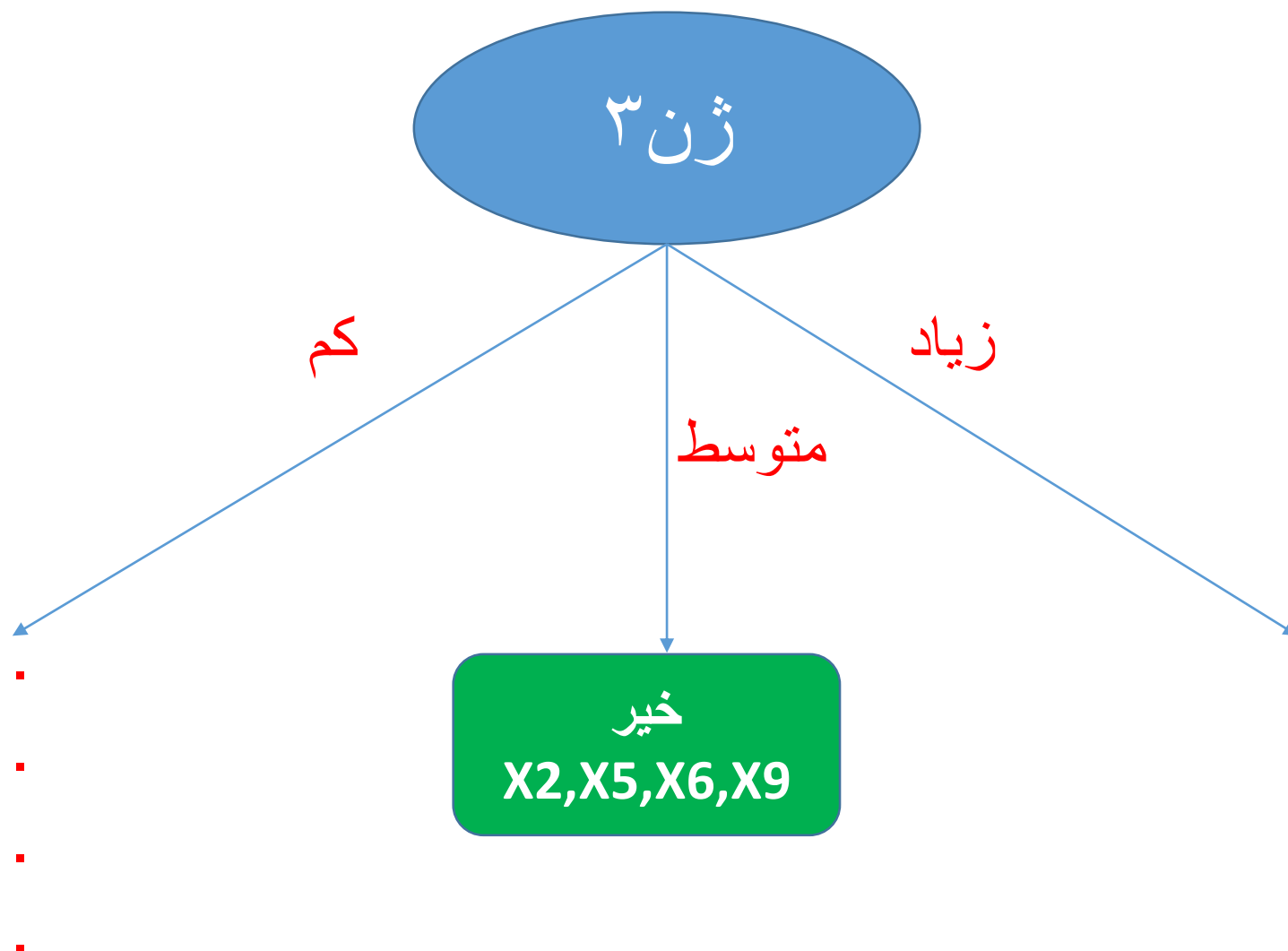


- $Gain(s, \text{ژن}^1) = entropy(s) - \frac{6}{9} * entropy(s_{\text{کم}}) - \frac{3}{9} * entropy(s_{\text{زیاد}}) = 0.249$
- $Gain(s, \text{ژن}^2) = entropy(s) - \frac{5}{9} * entropy(s_{\text{کم}}) - \frac{4}{9} * entropy(s_{\text{زیاد}}) = 0.018$
- $Gain(s, \text{ژن}^3) = entropy(s) - \frac{2}{9} * entropy(s_{\text{کم}}) - \frac{3}{9} * entropy(s_{\text{زیاد}}) - \frac{4}{9} * entropy(s_{\text{متوسط}}) = 0.389$
- $Gain(s, \text{ژن}^4) = entropy(s) - \frac{5}{9} * entropy(s_{\text{کم}}) - \frac{4}{9} * entropy(s_{\text{زیاد}}) = 0.378$
- $Gain(s, \text{ژن}^5) = entropy(s) - \frac{5}{9} * entropy(s_{\text{کم}}) - \frac{4}{9} * entropy(s_{\text{زیاد}}) = 0.071$



$$Gain(s, \text{ژن}^3) = entropy(s) - \frac{2}{9} * entropy(s_{\text{کم}}) - \frac{3}{9} * entropy(s_{\text{زیاد}}) - \frac{4}{9} * entropy(s_{\text{متوسط}}) = 0.389$$

انتخاب شد





داده هایی که ویژگی ژن ۳ آنها مقدار زیاد را داراست

نام شخص	ژن ۱	ژن ۲	ژن ۴	ژن ۵	بیمار بودن
X1	کم	کم	کم	زیاد	بله
X3	کم	زیاد	کم	زیاد	بله
X7	کم	زیاد	زیاد	زیاد	خیر

مقادیر: کم، زیاد → ویژگی: زن ۱



$S[2+ \quad 1-]$

$$\text{Entropy}(s) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.916$$

$S_{\text{کم}}[2+ \quad 1-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.916$$

- $\text{Gain}(s, \text{زن ۱}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{زن ۱}) = \text{entropy}(s) - \frac{3}{3} * \text{entropy}(s_{\text{کم}}) - \frac{0}{3} * \text{entropy}(s_{\text{زیاد}}) = 0$

مقادیر: کم، زیاد → ویژگی: ژن ۲



$S[2+ \quad 1-]$

$$\text{Entropy}(s) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.916$$

$S_{\text{کم}}[1+ \quad 0-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{1}{1} * \log_2 \frac{1}{1} - \frac{0}{1} * \log_2 \frac{0}{1} = 0$$

$S_{\text{زیاد}}[1+ \quad 1-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \frac{1}{3} * \text{entropy}(s_{\text{کم}}) - \frac{2}{3} * \text{entropy}(s_{\text{زیاد}}) = 0.249$

مقادیر: کم، زیاد → ویژگی: ژن ۴



$S[2+ \quad 1-]$

$$\text{Entropy}(s) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.916$$

$S_{\text{کم}}[۲+ \quad ۰-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{2}{2} * \log_2 \frac{2}{2} - \frac{0}{2} * \log_2 \frac{0}{2} = 0$$

$S_{\text{زیاد}}[۰+ \quad ۱-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{1} * \log_2 \frac{1}{1} - \frac{0}{2} * \log_2 \frac{0}{2} = 0$$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \frac{2}{3} * \text{entropy}(s_{\text{کم}}) - \frac{1}{3} * \text{entropy}(s_{\text{زیاد}}) = 0.916$

مقادیر: کم، زیاد \longrightarrow ویژگی: ژن ۵



$S[2+ \quad 1-]$

$$\text{Entropy}(s) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.916$$

$S[2+ \quad 1-]$ زیاد

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.916$$

- $\text{Gain}(s, \text{ژن ۵}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۵}) = \text{entropy}(s) - \frac{3}{3} * \text{entropy}(s_{\text{زیاد}}) = 0$

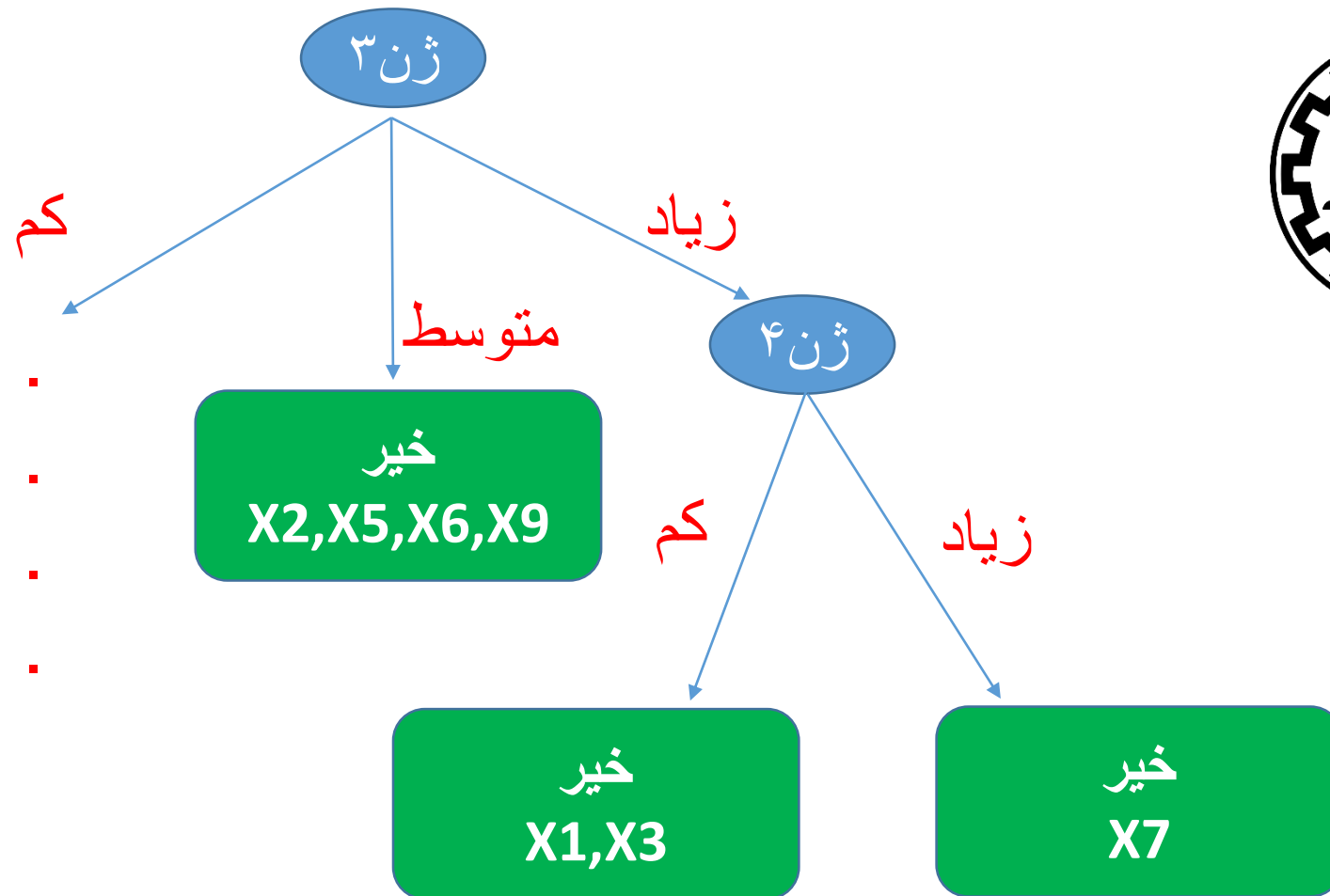


- $Gain(s, \text{ژن}^1) = entropy(s) - \frac{3}{3} * entropy(s_{\text{کم}}) - \frac{0}{3} * entropy(s_{\text{زیاد}}) = 0$
- $Gain(s, \text{ژن}^2) = entropy(s) - \frac{1}{3} * entropy(s_{\text{کم}}) - \frac{2}{3} * entropy(s_{\text{زیاد}}) = 0.249$
- $Gain(s, \text{ژن}^4) = entropy(s) - \frac{2}{3} * entropy(s_{\text{کم}}) - \frac{1}{3} * entropy(s_{\text{زیاد}}) = 0.916$
- $Gain(s, \text{ژن}^5) = entropy(s) - \frac{3}{3} * entropy(s_{\text{زیاد}}) = 0$



$$Gain(s, \text{ژن}^4) = entropy(s) - \frac{2}{3} * entropy(s_{\text{کم}}) - \frac{1}{3} * entropy(s_{\text{زیاد}}) = 0.916$$

انتخاب شد





داده هایی که ویژگی ژن ۳ آنها مقدار کم را داراست

نام شخص	ژن ۱	ژن ۲	ژن ۴	ژن ۵	بیمار بودن
X4	زیاد	کم	کم	کم	خیر
X8	کم	کم	کم	کم	بله

مقادیر: کم، زیاد → ویژگی: ژن ۱



$S[1+ \quad 1-]$

$$\text{Entropy}(s) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

$S_{\text{کم}}[1+ \quad 0-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{1}{1} * \log_2 \frac{1}{1} - \frac{0}{1} * \log_2 \frac{0}{1} = 0$$

$S_{\text{زیاد}}[0+ \quad 1-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{1} * \log_2 \frac{1}{1} - \frac{0}{1} * \log_2 \frac{0}{1} = 0$$

- $\text{Gain}(s, \text{ژن ۱}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۱}) = \text{entropy}(s) - \frac{1}{2} * \text{entropy}(s_{\text{کم}}) - \frac{1}{2} * \text{entropy}(s_{\text{زیاد}}) = 1$

مقادیر: کم، زیاد → ویژگی: ژن ۲



$S[1+ \quad 1-]$

$$\text{Entropy}(s) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

$S_{\text{کم}}[1+ \quad 1-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \frac{2}{2} * \text{entropy}(s_{\text{کم}}) = 0$

مقادیر: کم، زیاد → ویژگی: ژن ۴



$S[1+ \quad 1-]$

$$\text{Entropy}(s) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

$S_{\text{کم}}[1+ \quad 1-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \frac{2}{2} * \text{entropy}(s_{\text{کم}}) = 0$

مقادیر: کم، زیاد → ویژگی: زن ۵



$S[1+ \quad 1-]$

$$\text{Entropy}(s) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

$S[\text{زیاد} \quad ۱+ \quad ۱-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

- $\text{Gain}(s, \text{زن}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{زن}) = \text{entropy}(s) - \frac{2}{2} * \text{entropy}(s_{\text{زیاد}}) = 0$

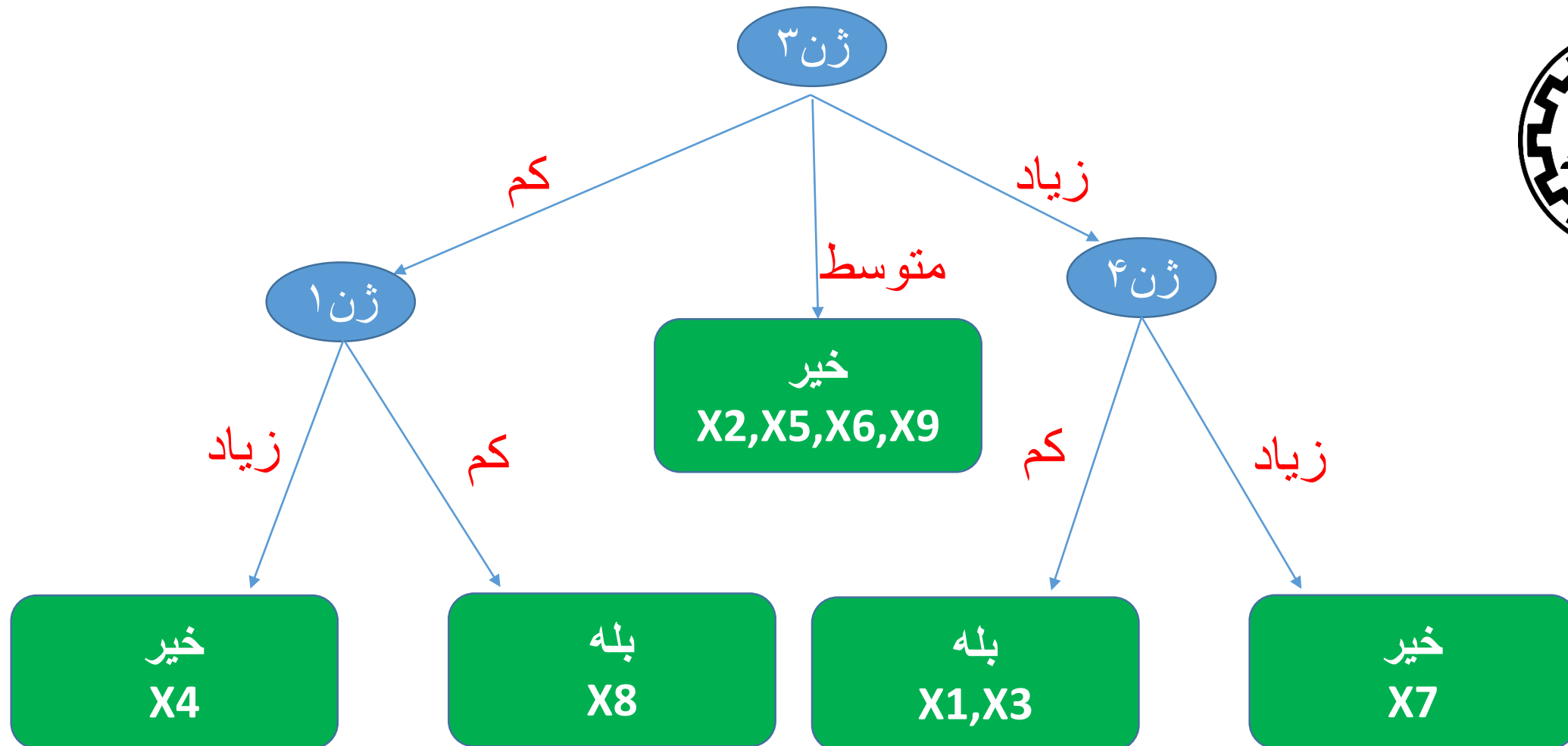


- $Gain(s, \text{ژن}^1) = entropy(s) - \frac{1}{2} * entropy(s_{\text{کم}}) - \frac{1}{2} * entropy(s_{\text{زیاد}}) = 1$
- $Gain(s, \text{ژن}^2) = entropy(s) - \frac{2}{2} * entropy(s_{\text{کم}}) = 0$
- $Gain(s, \text{ژن}^3) = entropy(s) - \frac{2}{2} * entropy(s_{\text{کم}}) = 0$
- $Gain(s, \text{ژن}^4) = entropy(s) - \frac{2}{2} * entropy(s_{\text{زیاد}}) = 0$



$$Gain(s, \text{ژن}^1) = entropy(s) - \frac{1}{2} * entropy(s_{\text{کم}}) - \frac{1}{2} * entropy(s_{\text{زیاد}}) = 1$$

انتخاب شد





بنابر این تغییر لیبل داده ی X5 سبب تغییر درخت تصیم و انتخاب ژن ها شد.

داده های تست



نام شخص	ژن ۱	ژن ۲	ژن ۳	ژن ۴	ژن ۵	بیمار بودن
Y1	کم	کم	زیاد	کم	زیاد	بله
Y2	زیاد	زیاد	کم	زیاد	زیاد	خیر
Y3	کم	کم	زیاد	کم	زیاد	بله
Y4	کم	کم	متوسط	کم	زیاد	خیر
Y5	کم	زیاد	متوسط	کم	زیاد	بله
Y6	کم	زیاد	زیاد	کم	کم	خیر



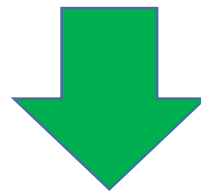
دقت = $66/66$
درصد

پیش بینی	بیمار بودن (لیبل واقعی)	نام شخص
بله	بله	Y1
خیر	خیر	Y2
بله	بله	Y3
خیر	خیر	Y4
خیر	بله	Y5
بله	خیر	Y6

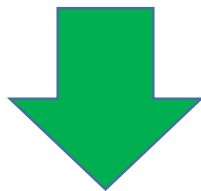


اگر فرد جدیدی به نام X10 به داده های آموزش اضافه می شد و علاوه بر اطلاعات زیر، شامل چند ژن دیگر با مقادیر تصادفی کم یا زیاد بود، تغییری در درخت تصمیم ایجاد می کرد؟

X10	17	80	کم	27	20	خیر
-----	----	----	----	----	----	-----



X10	17	80	کم	27	20	کم	زیاد	خیر
-----	----	----	----	----	----	----	------	-----



X10	کم	زیاد	کم	زیاد	کم	کم	زیاد	خیر
-----	----	------	----	------	----	----	------	-----



بیمار بودن	ژن ۷	ژن ۶	ژن ۵	ژن ۴	ژن ۳	ژن ۲	ژن ۱	نام شخص
بله	-	-	زیاد	کم	زیاد	کم	کم	X1
خیر	-	-	کم	کم	متوسط	کم	زیاد	X2
بله	-	-	زیاد	کم	زیاد	زیاد	کم	X3
خیر	-	-	کم	کم	کم	کم	زیاد	X4
خیر	-	-	کم	زیاد	متوسط	زیاد	کم	X5
خیر	-	-	زیاد	زیاد	متوسط	کم	کم	X6
خیر	-	-	زیاد	زیاد	زیاد	زیاد	کم	X7
بله	-	-	کم	کم	کم	کم	کم	X8
خیر	-	-	کم	زیاد	متوسط	زیاد	زیاد	X9
خیر	زیاد	کم	کم	زیاد	کم	زیاد	کم	X10

مقادیر: کم، زیاد → ویژگی: ژن ۱



$S[4+ \quad 6-]$

$$\text{Entropy}(s) = -\frac{4}{10} * \log_2 \frac{4}{10} - \frac{6}{10} * \log_2 \frac{6}{10} = 0.969$$

$S_{\text{کم}}[۴+ \quad ۳-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{4}{7} * \log_2 \frac{4}{7} - \frac{3}{7} * \log_2 \frac{3}{7} = 0.984$$

$S_{\text{زیاد}}[۰+ \quad ۳-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{0}{3} * \log_2 \frac{0}{3} - \frac{3}{3} * \log_2 \frac{3}{3} = 0$$

- $\text{Gain}(s, \text{ژن ۱}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$

- $\text{Gain}(s, \text{ژن ۱}) = \text{entropy}(s) - \frac{7}{10} * \text{entropy}(s_{\text{کم}}) - \frac{3}{10} * \text{entropy}(s_{\text{زیاد}}) = 0.280$

مقادیر: کم، زیاد → ویژگی: ژن ۲



$S[4+ \quad 6-]$

$$\text{Entropy}(s) = -\frac{4}{10} * \log_2 \frac{4}{10} - \frac{6}{10} * \log_2 \frac{6}{10} = 0.984$$

$S_{\text{کم}}[۲+ \quad ۳-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{2}{5} * \log_2 \frac{2}{5} - \frac{3}{5} * \log_2 \frac{3}{5} = 0.969$$

$S_{\text{زیاد}}[۲+ \quad ۳-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{5} * \log_2 \frac{2}{5} - \frac{3}{5} * \log_2 \frac{3}{5} = 0.969$$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \frac{5}{10} * \text{entropy}(s_{\text{کم}}) - \frac{5}{10} * \text{entropy}(s_{\text{زیاد}}) = 0$



مقادیر: کم، زیاد و متوسط \longrightarrow ویژگی: ۳ زن

$S[4+ \quad 6-]$

$$\text{Entropy}(s) = -\frac{4}{10} * \log_2 \frac{4}{10} - \frac{6}{10} * \log_2 \frac{6}{10} = 0.984$$

$S_{\text{کم}}[1+ \quad 2-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{1}{3} * \log_2 \frac{1}{3} - \frac{2}{3} * \log_2 \frac{2}{3} = 0.916$$

$S_{\text{زیاد}}[2+ \quad 1-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.916$$

$S_{\text{متوسط}}[1+ \quad 3-]$

$$\text{Entropy}(s_{\text{متوسط}}) = -\frac{1}{4} * \log_2 \frac{1}{4} - \frac{3}{4} * \log_2 \frac{3}{4} = 0.811$$

- $\text{Gain}(s, \text{ژن ۳}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$

- $\text{Gain}(s, \text{ژن ۳}) = \text{entropy}(s) - \frac{3}{10} * \text{entropy}(s_{\text{کم}}) - \frac{3}{10} * \text{entropy}(s_{\text{زیاد}}) - \frac{4}{10} * \text{entropy}(s_{\text{متوسط}}) = 0.096$

مقادیر: کم، زیاد → ویژگی: ژن ۴



$S[4+ \quad 6-]$

$$\text{Entropy}(s) = -\frac{4}{10} * \log_2 \frac{4}{10} - \frac{6}{10} * \log_2 \frac{6}{10} = 0.984$$

$S_{\text{کم}}[۲+ \quad ۳-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{2}{5} * \log_2 \frac{2}{5} - \frac{3}{5} * \log_2 \frac{3}{5} = 0.969$$

$S_{\text{زیاد}}[۱+ \quad ۴-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{5} * \log_2 \frac{1}{5} - \frac{4}{5} * \log_2 \frac{4}{5} = 0.72$$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \frac{5}{10} * \text{entropy}(s_{\text{کم}}) - \frac{5}{10} * \text{entropy}(s_{\text{زیاد}}) = 0.125$

مقادیر: کم، زیاد → ویژگی: ژن ۵



$S[2+ \quad 4-]$

$$\text{Entropy}(s) = -\frac{4}{10} * \log_2 \frac{4}{10} - \frac{6}{10} * \log_2 \frac{6}{10} = 0.984$$

$S_{\text{کم}}[۲+ \quad ۴-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{2}{6} * \log_2 \frac{2}{6} - \frac{4}{6} * \log_2 \frac{4}{6} = 0.916$$

$S_{\text{زیاد}}[۲+ \quad ۲-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{4} * \log_2 \frac{2}{4} - \frac{2}{4} * \log_2 \frac{2}{4} = 1$$

- $\text{Gain}(s, \text{ژن ۵}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$

- $\text{Gain}(s, \text{ژن ۵}) = \text{entropy}(s) - \frac{6}{10} * \text{entropy}(s_{\text{کم}}) - \frac{4}{10} * \text{entropy}(s_{\text{زیاد}}) = 0.034$



- $Gain(s, ١\text{ ژن}) = entropy(s) - \frac{7}{10} * entropy(s_{کم}) - \frac{3}{10} * entropy(s_{زیاد}) = 0.280$

- $Gain(s, ٢\text{ ژن}) = entropy(s) - \frac{5}{10} * entropy(s_{کم}) - \frac{5}{10} * entropy(s_{زیاد}) = 0$

- $Gain(s, ٣\text{ ژن}) = entropy(s) - \frac{3}{10} * entropy(s_{کم}) - \frac{3}{10} * entropy(s_{زیاد}) - \frac{4}{10} * entropy(s_{متوسط}) = 0.096$

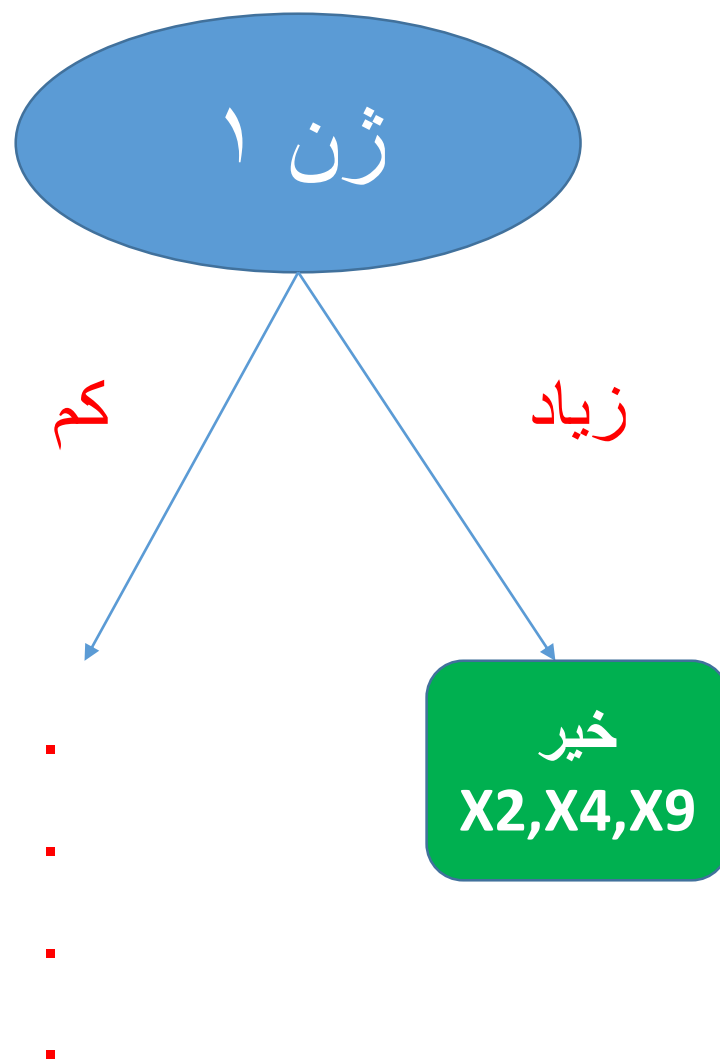
- $Gain(s, ٤\text{ ژن}) = entropy(s) - \frac{5}{10} * entropy(s_{کم}) - \frac{5}{10} * entropy(s_{زیاد}) = 0.125$

- $Gain(s, ٥\text{ ژن}) = entropy(s) - \frac{6}{10} * entropy(s_{کم}) - \frac{4}{10} * entropy(s_{زیاد}) = 0.034$



$$Gain(s, ١\text{ ژن}) = entropy(s) - \frac{7}{10} * entropy(s_{کم}) - \frac{3}{10} * entropy(s_{زیاد}) = 0.280$$

انتخاب شد





دیتاست پس از انتخاب ژن ۱

نام شخص	ژن ۲	ژن ۳	ژن ۴	ژن ۵	ژن ۶	ژن ۷	بیمار بودن
X1	کم	زیاد	کم	زیاد	-	-	بله
X3	زیاد	زیاد	کم	زیاد	-	-	بله
X5	زیاد	متوسط	زیاد	کم	-	-	خیر
X6	کم	متوسط	زیاد	زیاد	-	-	خیر
X7	زیاد	زیاد	زیاد	زیاد	-	-	خیر
X8	کم	کم	کم	کم	-	-	بله
X10	زیاد	کم	زیاد	کم	کم	زیاد	خیر

مقادیر: کم، زیاد → ویژگی: ژن ۲



$S[4+ \quad 3-]$

$$\text{Entropy}(s) = -\frac{4}{7} * \log_2 \frac{4}{7} - \frac{3}{7} * \log_2 \frac{3}{7} = 0.984$$

$S_{\text{کم}}[۲+ \quad ۱-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.916$$

$S_{\text{زیاد}}[۲+ \quad ۲-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{4} * \log_2 \frac{2}{4} - \frac{2}{4} * \log_2 \frac{2}{4} = 1$$

- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \frac{3}{7} * \text{entropy}(s_{\text{کم}}) - \frac{4}{7} * \text{entropy}(s_{\text{زیاد}}) = 0.021$



مقادیر: کم، زیاد و متوسط \longrightarrow ویژگی: ۳ زن

$S[4+ \quad 3-]$

$$\text{Entropy}(s) = -\frac{4}{7} * \log_2 \frac{4}{7} - \frac{3}{7} * \log_2 \frac{3}{7} = 0.984$$

$S_{\text{کم}}[1+ \quad 1-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

$S_{\text{زیاد}}[2+ \quad 1-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.916$$

$S_{\text{متوسط}}[1+ \quad 1-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

- $\text{Gain}(s, \text{ژن ۳}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۳}) = \text{entropy}(s) - \frac{2}{7} * \text{entropy}(s_{\text{کم}}) - \frac{3}{7} * \text{entropy}(s_{\text{زیاد}}) - \frac{2}{7} * \text{entropy}(s_{\text{متوسط}}) = 0.022$



مقادیر: کم، زیاد → ویژگی: ژن ۴

$S[4+ \quad 3-]$

$$\text{Entropy}(s) = -\frac{4}{7} * \log_2 \frac{4}{7} - \frac{3}{7} * \log_2 \frac{3}{7} = 0.984$$

$S_{\text{کم}}[3+ \quad 0-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{3}{3} * \log_2 \frac{3}{3} - \frac{0}{3} * \log_2 \frac{0}{3} = 0$$

$S_{\text{زیاد}}[1+ \quad 3-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{4} * \log_2 \frac{1}{4} - \frac{3}{4} * \log_2 \frac{3}{4} = 0.811$$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \frac{3}{7} * \text{entropy}(s_{\text{کم}}) - \frac{4}{7} * \text{entropy}(s_{\text{زیاد}}) = 0.501$



مقادیر: کم، زیاد → ویژگی: ژن ۵

$S[4+ \quad 3-]$

$$\text{Entropy}(s) = -\frac{4}{7} * \log_2 \frac{4}{7} - \frac{3}{7} * \log_2 \frac{3}{7} = 0.984$$

$S_{\text{کم}}[۲+ \quad ۱-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{2}{3} * \log_2 \frac{2}{3} - \frac{1}{3} * \log_2 \frac{1}{3} = 0.916$$

$S_{\text{زیاد}}[۲+ \quad ۲-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{2}{4} * \log_2 \frac{2}{4} - \frac{2}{4} * \log_2 \frac{2}{4} = 1$$

- $\text{Gain}(s, \text{ژن ۵}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۵}) = \text{entropy}(s) - \frac{3}{7} * \text{entropy}(s_{\text{کم}}) - \frac{4}{7} * \text{entropy}(s_{\text{زیاد}}) = 0.005$

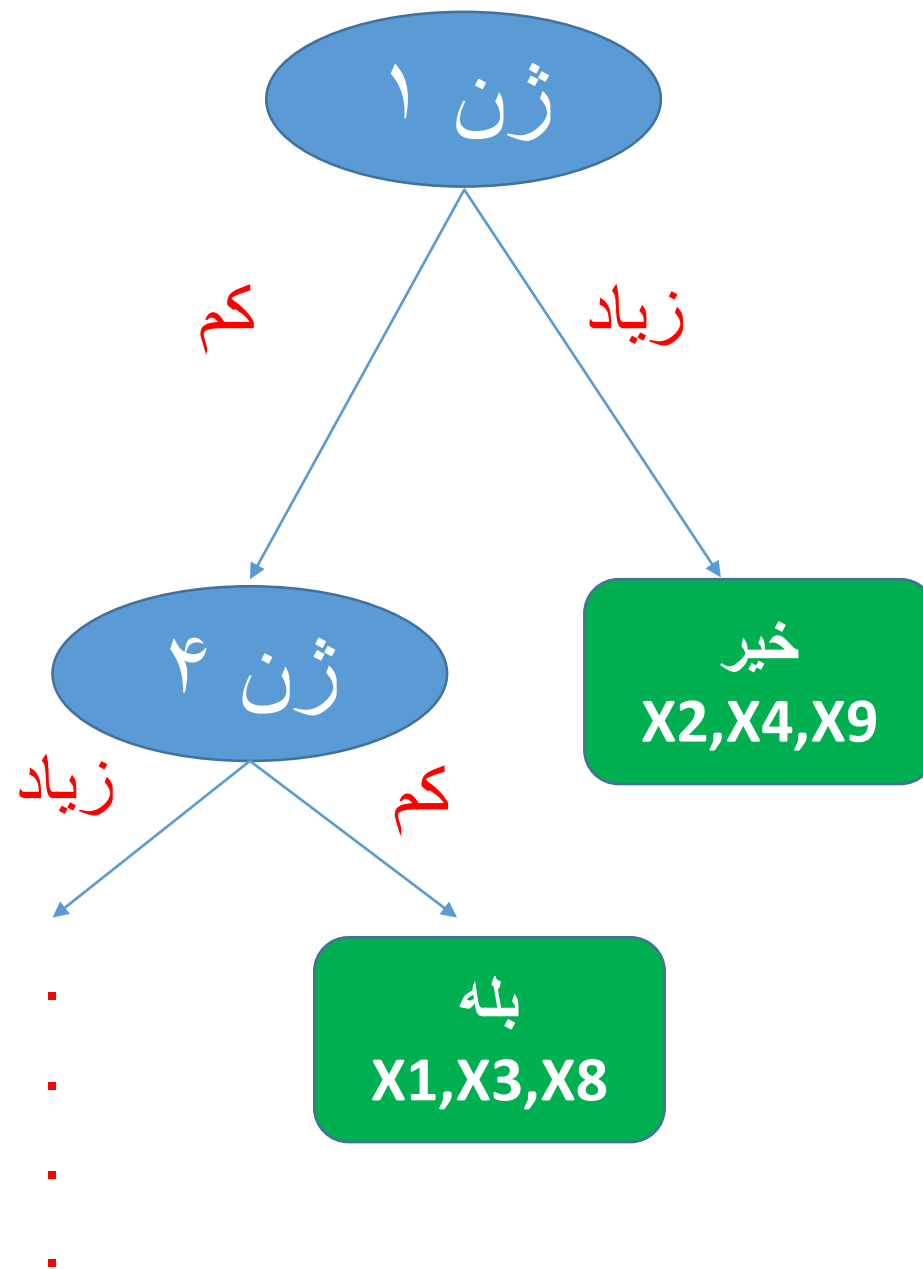


- $Gain(s, \text{ژن}^2) = entropy(s) - \frac{3}{7} * entropy(s_{\text{کم}}) - \frac{4}{7} * entropy(s_{\text{زیاد}}) = 0.021$
- $Gain(s, \text{ژن}^3) = entropy(s) - \frac{2}{7} * entropy(s_{\text{کم}}) - \frac{3}{7} * entropy(s_{\text{زیاد}}) - \frac{2}{7} * entropy(s_{\text{متوسط}}) = 0.022$
- $Gain(s, \text{ژن}^4) = entropy(s) - \frac{3}{7} * entropy(s_{\text{کم}}) - \frac{4}{7} * entropy(s_{\text{زیاد}}) = 0.501$
- $Gain(s, \text{ژن}^5) = entropy(s) - \frac{3}{7} * entropy(s_{\text{کم}}) - \frac{4}{7} * entropy(s_{\text{زیاد}}) = 0.005$



$$Gain(s, \text{ژن}^2) = entropy(s) - \frac{3}{7} * entropy(s_{\text{کم}}) - \frac{4}{7} * entropy(s_{\text{زیاد}}) = 0.021$$

انتخاب شد





دیتاست پس از انتخاب ژن ۴

نام شخص	ژن ۲	ژن ۳	ژن ۵	ژن ۶	ژن ۷	بیمار بودن
X5	زیاد	متوسط	کم	-	-	خیر
X6	کم	متوسط	زیاد	-	-	خیر
X7	زیاد	زیاد	زیاد	-	-	خیر
X10	زیاد	کم	کم	کم	زیاد	خیر

مقادیر: کم، زیاد → ویژگی: ژن ۲



$S[1+ \quad 3-]$

$$\text{Entropy}(s) = -\frac{1}{4} * \log_2 \frac{1}{4} - \frac{3}{4} * \log_2 \frac{3}{4} = 0.811$$

$S_{\text{کم}}[0+ \quad 1-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{0}{1} * \log_2 \frac{0}{1} - \frac{1}{1} * \log_2 \frac{1}{1} = 0$$

$S_{\text{زیاد}}[1+ \quad 2-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{3} * \log_2 \frac{1}{3} - \frac{2}{3} * \log_2 \frac{2}{3} = 0.916$$

- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۲}) = \text{entropy}(s) - \frac{1}{4} * \text{entropy}(s_{\text{کم}}) - \frac{3}{4} * \text{entropy}(s_{\text{زیاد}}) = 0.124$



مقادیر: کم، زیاد → ویژگی: ژن ۳
و متوسط

$S[1+ \quad 3-]$

$$\text{Entropy}(s) = -\frac{1}{4} * \log_2 \frac{1}{4} - \frac{3}{4} * \log_2 \frac{3}{4} = 0.811$$

$S[\text{کم} \quad 0+ \quad 1-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{0}{1} * \log_2 \frac{0}{1} - \frac{1}{1} * \log_2 \frac{1}{1} = 0$$

$S[\text{زیاد} \quad 0+ \quad 1-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{0}{1} * \log_2 \frac{0}{1} - \frac{1}{1} * \log_2 \frac{1}{1} = 0$$

$S[\text{متوسط} \quad 1+ \quad 1-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \frac{1}{4} * \text{entropy}(s_{\text{کم}}) - \frac{1}{4} * \text{entropy}(s_{\text{زیاد}}) - \frac{2}{4} * \text{entropy}(s_{\text{متوسط}}) = 0.311$

مقادیر: کم، زیاد → ویژگی: ژن ۵



$S[1+ \quad 3-]$

$$\text{Entropy}(s) = -\frac{1}{4} * \log_2 \frac{1}{4} - \frac{3}{4} * \log_2 \frac{3}{4} = 0.811$$

$S_{\text{کم}}[1+ \quad 1-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

$S_{\text{زیاد}}[0+ \quad 2-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{0}{2} * \log_2 \frac{0}{2} - \frac{2}{2} * \log_2 \frac{2}{2} = 0$$

- $\text{Gain}(s, \text{ژن ۵}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن ۵}) = \text{entropy}(s) - \frac{2}{4} * \text{entropy}(s_{\text{کم}}) - \frac{1}{4} * \text{entropy}(s_{\text{زیاد}}) = 0.311$

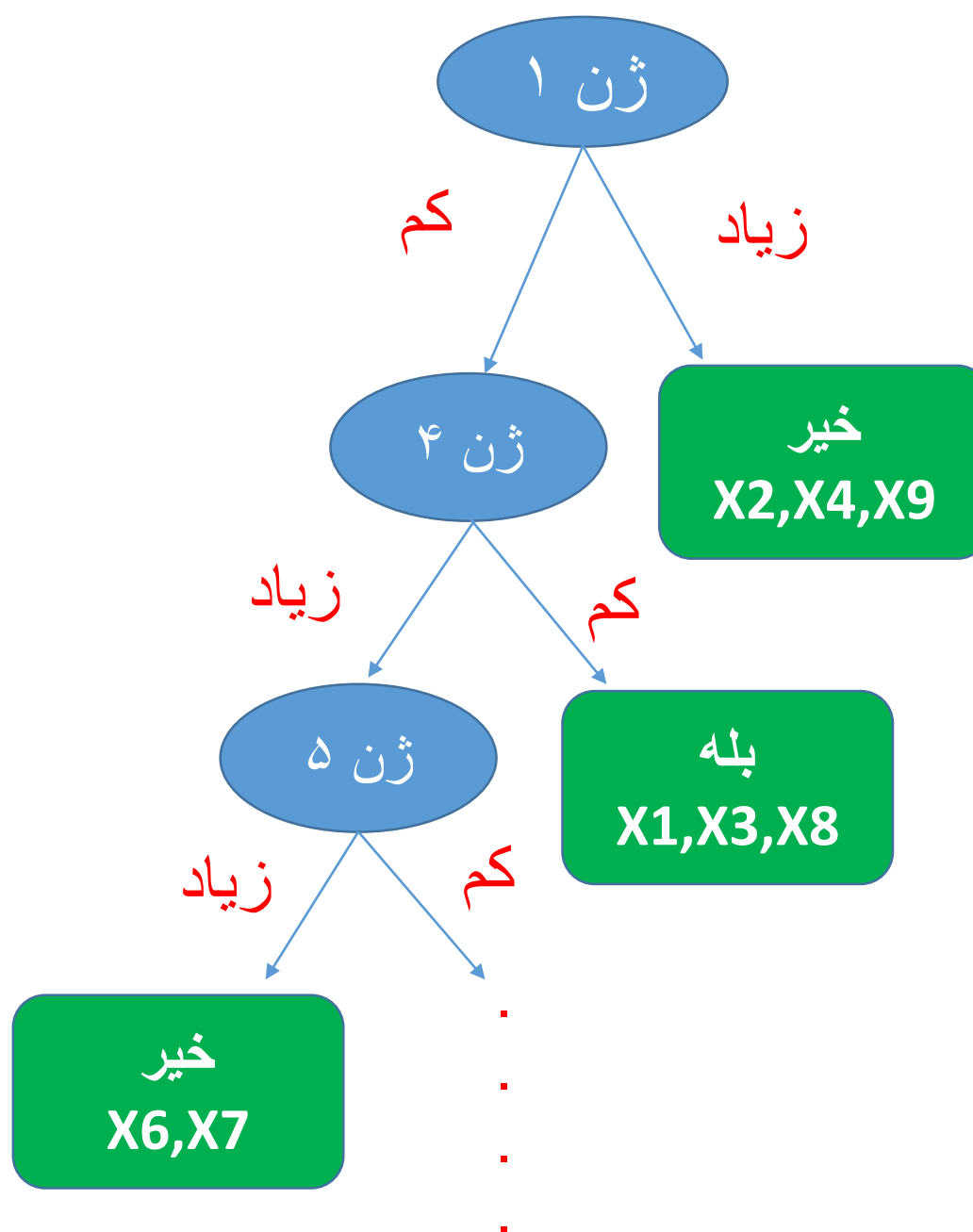


- $Gain(s, \text{ژن}^2) = entropy(s) - \frac{1}{4} * entropy(s_{\text{کم}}) - \frac{3}{4} * entropy(s_{\text{زیاد}}) = 0.124$
- $Gain(s, \text{ژن}^3) = entropy(s) - \frac{1}{4} * entropy(s_{\text{کم}}) - \frac{1}{4} * entropy(s_{\text{زیاد}}) - \frac{2}{4} * entropy(s_{\text{متوسط}}) = 0.311$
- $Gain(s, \text{ژن}^4) = entropy(s) - \frac{2}{4} * entropy(s_{\text{کم}}) - \frac{1}{4} * entropy(s_{\text{زیاد}}) = 0.311$



$$Gain(s, \text{ژن}^5) = entropy(s) - \frac{2}{4} * entropy(s_{\text{کم}}) - \frac{1}{4} * entropy(s_{\text{زیاد}}) = 0.311$$

انتخاب شد





دیتاست پس از انتخاب ژن ۵

نام شخص	ژن ۲	ژن ۳	ژن ۶	ژن ۷	بیمار بودن
X5	زیاد	متوسط	-	-	خیر
X10	زیاد	کم	کم	زیاد	خیر

مقادیر: کم، زیاد → ویژگی: ژن ۲



$S[1+ \quad 1-]$

$$\text{Entropy}(s) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

$S[\text{زیاد} \quad ۱-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{1}{2} * \log_2 \frac{1}{2} - \frac{1}{2} * \log_2 \frac{1}{2} = 1$$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \frac{2}{2} * \text{entropy}(s_{\text{زیاد}}) = 0$



مقادیر: کم، زیاد و متوسط \longrightarrow ویژگی: ژن ۳

$S[1+ \quad 1-]$

$$\text{Entropy}(s) = -\frac{1}{4} * \log_2 \frac{1}{4} - \frac{3}{4} * \log_2 \frac{3}{4} = 0.811$$

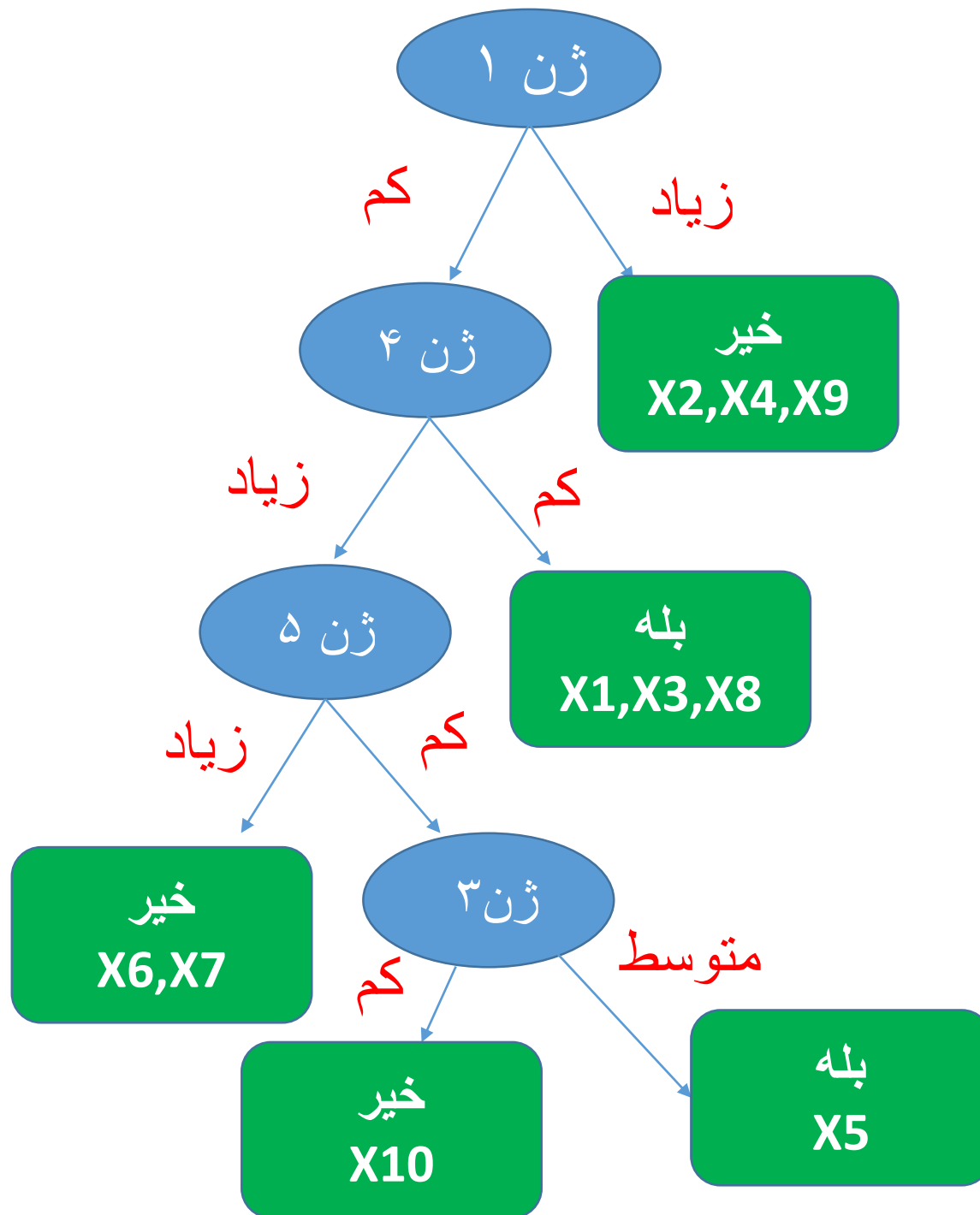
$S_{\text{کم}}[0+ \quad 1-]$

$$\text{Entropy}(s_{\text{کم}}) = -\frac{0}{1} * \log_2 \frac{0}{1} - \frac{1}{1} * \log_2 \frac{1}{1} = 0$$

$S_{\text{زیاد}}[1+ \quad 0-]$

$$\text{Entropy}(s_{\text{زیاد}}) = -\frac{0}{1} * \log_2 \frac{0}{1} - \frac{1}{1} * \log_2 \frac{1}{1} = 0$$

- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \sum \frac{|S_V|}{|S|} * \text{entropy}(S_V)$
- $\text{Gain}(s, \text{ژن}) = \text{entropy}(s) - \frac{1}{2} * \text{entropy}(s_{\text{کم}}) - \frac{1}{2} * \text{entropy}(s_{\text{زیاد}}) = 1$





بنابراین اضافه کردن ویژگی X10 سبب تغییر درخت شد اما ژن ها
اضافی داده ی X10 برای Pure شدن برگ ها مورد نیاز قرار نگرفت.

داده های تست



نام شخص	ژن ۱	ژن ۲	ژن ۳	ژن ۴	ژن ۵	بیمار بودن
Y1	کم	کم	زیاد	کم	زیاد	بله
Y2	زیاد	زیاد	کم	زیاد	زیاد	خیر
Y3	کم	کم	زیاد	کم	زیاد	بله
Y4	کم	کم	متوسط	کم	زیاد	خیر
Y5	کم	زیاد	متوسط	کم	زیاد	بله
Y6	کم	زیاد	زیاد	کم	کم	خیر



پیش بینی	بیمار بودن (لیبل واقعی)	نام شخص
بله	بله	Y1
خیر	خیر	Y2
بله	بله	Y3
خیر	خیر	Y4
خیر	بله	Y5
بله	خیر	Y6

دقت = ۶۶/۶۶

درصد