

$$\text{mean}(\bar{x}) = \frac{x_1 + x_2 + \dots + x_n}{n}, \quad s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (\text{سوال ۱ الف})$$

$$GA = \sum_{i=1}^n |x_i - \bar{x}|$$

سوال در صد جی بی:

$$\mu = \frac{9,0 + 24,0 + 11,1 + \dots + 30,1}{11} = 21,11$$

$$s = \sqrt{\frac{1}{11-1} ((9,0 - 21,11)^2 + (24,0 - 21,11)^2 + (11,1 - 21,11)^2 + \dots + (30,1 - 21,11)^2)}$$

$$= \sqrt{10,68} = 3,28$$

$$GA = |9,0 - 21,11| + |24,0 - 21,11| + |11,1 - 21,11| + \dots + |30,1 - 21,11| = 127,42$$

برای سن:

$$\mu = \frac{23 + 23 + \dots + 61}{11} = 44,55$$

$$s = \sqrt{\frac{1}{11-1} ((23 - 44,55)^2 + (23 - 44,55)^2 + \dots + (61 - 44,55)^2)} = \sqrt{\frac{1}{10} \times 2970,44}$$

$$= 17,22$$

$$GA = |23 - 44,55| + |23 - 44,55| + \dots + |61 - 44,55| = 190,44$$

die je zu den anderen $z' = \frac{z - z_{\min}}{z_{\max} - z_{\min}}$ (Jeweils: 0,9 bis 1,0) : $\frac{1}{10}$

$$z'_1 = \frac{91,0 - 71,1}{82,0 - 71,1} = 0,8199 \quad z'_{14} = \frac{42,9 - 71,1}{82,0 - 71,1} = 0,1773$$

$$z'_2 = \frac{19,0 - 71,1}{82,0 - 71,1} = 0,04190 \quad z'_{15} = \frac{41,2 - 71,1}{82,0 - 71,1} = 0,1920$$

$$z'_3 = \frac{71,1 - 71,1}{82,0 - 71,1} = 0 \quad z'_{16} = \frac{30,7 - 71,1}{82,0 - 71,1} = 0,1020$$

$$z'_4 = \frac{18,1 - 71,1}{82,0 - 71,1} = 0,1811$$

$$z'_5 = \frac{31,8 - 71,1}{82,0 - 71,1} = 0,1101$$

$$z'_6 = \frac{20,9 - 71,1}{82,0 - 71,1} = 0,1191$$

$$z'_7 = \frac{26,2 - 71,1}{82,0 - 71,1} = 0,0412$$

$$z'_8 = \frac{26,2 - 71,1}{82,0 - 71,1} = 0,0412$$

$$z'_9 = \frac{41,2 - 71,1}{82,0 - 71,1} = 0,1920$$

$$z'_{10} = \frac{42,9 - 71,1}{82,0 - 71,1} = 0,1773$$

$$z'_{11} = \frac{42,0 - 71,1}{82,0 - 71,1} = 0$$

$$z'_{12} = \frac{28,1 - 71,1}{82,0 - 71,1} = 0,0011$$

$$z'_{13} = \frac{41,2 - 71,1}{82,0 - 71,1} = 0,1920$$

$$z'_{14} = \frac{40,2 - 71,1}{82,0 - 71,1} = 0,1920$$

$$z'_{15} = \frac{42,1 - 71,1}{82,0 - 71,1} = 0,1920$$

$$z' = \frac{z - z_{\min}}{z_{\max} - z_{\min}}$$

بجایگزینی

$$z'_1 = \frac{0 - 24}{91 - 24} = 0$$

$$z'_{14} = \frac{0.5 - 24}{38} = 0,1415$$

$$z'_2 = \frac{24 - 24}{41 - 24} = 0$$

$$z'_{15} = \frac{0.5 - 24}{38} = 0,1415$$

$$z'_3 = \frac{27 - 24}{38} = 0,1052$$

$$z'_{16} = \frac{0.5 - 24}{38} = 0,1415$$

$$z'_4 = \frac{27 - 24}{38} = 0,1052$$

$$z'_{17} = \frac{0.5 - 24}{38} = 0,1415$$

$$z'_5 = \frac{29 - 24}{38} = 0,1315$$

$$z'_{18} = \frac{0.5 - 24}{38} = 0,1415$$

$$z'_6 = \frac{31 - 24}{38} = 0,1842$$

$$z'_{19} = \frac{0.5 - 24}{38} = 0,1415$$

$$z'_7 = \frac{37 - 24}{38} = 0,3421$$

$$z'_{20} = \frac{29 - 24}{38} = 0,1315$$

$$z'_8 = \frac{40 - 24}{38} = 0,4210$$

$$z'_{21} = \frac{0.5 - 24}{38} = 0,1415$$

$$z'_{22} = \frac{0.5 - 24}{38} = 0,1415$$

$$z'_{23} = \frac{0.5 - 24}{38} = 0,1415$$

استاندارد سازی داده ها در روش بارش: با انحراف معیار و میانگین: $F4, FF$

$$Z' = \frac{(Z - \bar{Z})}{\sigma_A} \rightarrow \sigma_A = 190, 44$$

$$Z'_1 = -0,1118$$

$$Z'_2 = -0,1118$$

$$Z'_3 = -0,1985$$

$$Z'_4 = -0,1985$$

$$Z'_5 = -0,1377$$

$$Z'_6 = -0,1027$$

$$Z'_7 = 0,1027$$

$$Z'_8 = 0,1029$$

$$Z'_9 = 0,1018$$

$$Z'_{10} = 0,1028$$

$$Z'_{11} = 0,1038$$

$$Z'_{12} = 0,1038$$

$$Z'_{13} = 0,1038$$

$$Z'_{14} = 0,1054$$

$$Z'_{15} = 0,1051$$

$$Z'_{16} = 0,1051$$

$$Z'_{17} = 0,1061$$

$$Z'_{18} = -0,1024$$

استاندارد سازی داده ها در روش بارش: با انحراف معیار و میانگین: $28, 78$

$$Z' = \frac{(Z - \bar{Z})}{\sigma_A} \rightarrow \sigma_A = 127, 42$$

$$Z'_1 = -0,1010$$

$$Z'_2 = -0,1017$$

$$Z'_3 = -0,1145$$

$$Z'_4 = -0,10840$$

$$Z'_5 = -0,1020$$

$$Z'_6 = -0,10240$$

$$Z'_7 = 0,1010$$

$$Z'_8 = -0,1012$$

$$Z'_9 = 0,10844$$

$$Z'_{10} = 0,10504$$

$$Z'_{11} = 0,1075$$

$$Z'_{12} = 0,10015$$

$$Z'_{13} = 0,10342$$

$$Z'_{14} = 0,10111$$

$$Z'_{15} = 0,10314$$

$$Z'_{16} = 0,10322$$

$$Z'_{17} = 0,10522$$

$$Z'_{18} = 0,10973$$

استاندارد سازی داده‌های ویژگی‌ها بر اساس انحراف معیار:

$$Z' = \frac{Z - \bar{X}}{S}$$

$$Z'_1 = \frac{23 - 46.44}{13.22} = -1.773$$

$$Z'_{16} = \frac{58 - 46.44}{13.22} = 0.874$$

$$Z'_2 = \frac{23 - 46.44}{13.22} = -1.773$$

$$Z'_{17} = \frac{60 - 46.44}{13.22} = 1.025$$

$$Z'_3 = \frac{27 - 46.44}{13.22} = -1.4709$$

$$Z'_{18} = \frac{61 - 46.44}{13.22} = 1.1011$$

$$Z'_4 = \frac{27 - 46.44}{13.22} = -1.4709$$

$$Z'_5 = \frac{39 - 46.44}{13.22} = -0.563$$

$$Z'_6 = \frac{41 - 46.44}{13.22} = -0.411$$

$$Z'_7 = \frac{47 - 46.44}{13.22} = 0.042$$

$$Z'_8 = \frac{49 - 46.44}{13.22} = 0.193$$

$$Z'_9 = \frac{50 - 46.44}{13.22} = 0.268$$

$$Z'_{10} = \frac{52 - 46.44}{13.22} = 0.4202$$

$$Z'_{11} = \frac{54 - 46.44}{13.22} = 0.5715$$

$$Z'_{12} = \frac{54 - 46.44}{13.22} = 0.5715$$

$$Z'_{13} = \frac{56 - 46.44}{13.22} = 0.7228$$

$$Z'_{14} = \frac{57 - 46.44}{13.22} = 0.7985$$

$$Z'_{15} = \frac{58 - 46.44}{13.22} = 0.874$$

استاندارد ساری داده های ویرنجی در محدوده ی: بر اساس اعراض معیار

$$Z' = \frac{(Z - \bar{Z})}{s}$$

$$s = 9.25 \quad \bar{Z} = 28.78$$

$$Z'_1 = \frac{9.5 - 28.78}{9.25} = -2.08$$

$$Z'_{15} = \frac{34.1 - 28.78}{9.25} = 0.575$$

$$Z'_2 = \frac{26.5 - 28.78}{9.25} = 0.24$$

$$Z'_{16} = \frac{32.9 - 28.78}{9.25} = 0.445$$

$$Z'_3 = \frac{7.8 - 28.78}{9.25} = -2.26$$

$$Z'_{17} = \frac{41.2 - 28.78}{9.25} = 1.342$$

$$Z'_4 = \frac{17.8 - 28.78}{9.25} = -1.18$$

$$Z'_{18} = \frac{35.7 - 28.78}{9.25} = 0.748$$

$$Z'_5 = \frac{31.4 - 28.78}{9.25} = 0.283$$

$$Z'_6 = \frac{25.9 - 28.78}{9.25} = -0.311$$

$$Z'_7 = \frac{27.4 - 28.78}{9.25} = -0.149$$

$$Z'_8 = \frac{27.2 - 28.78}{9.25} = -0.170$$

$$Z'_9 = \frac{31.2 - 28.78}{9.25} = 0.261$$

$$Z'_{10} = \frac{34.6 - 28.78}{9.25} = 0.629$$

$$Z'_{11} = \frac{42.5 - 28.78}{9.25} = 1.483$$

$$Z'_{12} = \frac{28.8 - 28.78}{9.25} = 0.002$$

$$Z'_{13} = \frac{33.4 - 28.78}{9.25} = 0.499$$

$$Z'_{14} = \frac{30.2 - 28.78}{9.25} = 0.153$$

$$\left[\frac{z_{min} - \bar{z}}{\sigma}, \frac{z_{max} - \bar{z}}{\sigma} \right] \quad (1)$$

$$\left[\frac{z_{\min} - \bar{z}}{\sigma_A}, \frac{z_{\max} - \bar{z}}{\sigma_A} \right] \quad (2)$$

$$z' - \frac{z - z_{\min}}{z_{\max} - z_{\min}} \times (b - a) + a \quad \text{مثال ٣}$$

۲. روش سیستمی، روش مبتنی بر کس :

$$r = \left[\frac{n}{k} \right] = 4 \rightarrow [5, 15], [15, 42], [42, 215]$$

→ {5, 10, 11, 13}, {15, 40, 50, 55}, {72, 92, 200, 210}

(ب) سیستم های سه عنصری :

$z = \frac{210 - 5}{3} = 70$ $[5, 70), [70, 140), [140, 210]$

$\{5, 10, 11, 13, 15, 30, 50, 55, 72, 92\}$ $\{2, 215\}$

(طبق جزوه)

مثال (۵) حذف کردن: این روش معمولاً زمانی استفاده می شود که رگورهای دارای فیلد از دست رفته
تعداد بسیار کمی از کل رگورهای داده باشد.

تخمین زدن : این روش زمانی استفاده می شود که در شرایط نامعین و نامشخص داشته باشیم و به تعادیر مورد تخمین از نظام خاصی تبعیت کنند.

نابیده گشتن : در این روش بر خلاف حذف کردن کل مورد را که یک ویژگی از دست نرفته دارد حذف نمی کنیم و در مورد ویژگی های از دست رفته هم اقدامی انجام نمی دهیم و آن ها را نابیده می بینیم .
در واقع بخش که حذف شده باشد را داشته داشته می کنیم و بخش از دست رفته را نابیده می بینیم .
معمولاً زمانی که مقدار مقادیر Null داده ها در سطح پایگاه داده زیاد باشد از این روش استفاده می کنیم .

جائزین کردن: بدین روش قادر Null به تمام قادر امکان پذیر جائزین می گردد و معادل از این روش زمانی که هم ظاهر باشد جمع کم داده با ارزش زیر دترم است و در هر صورت

سؤال 6: برای انتخاب درختی جای ششیم از استندارتی استفاده می‌کنیم:

$$H(X) = -\left(\frac{7}{13} \log_2 \left(\frac{7}{13}\right) + \frac{6}{13} \log_2 \left(\frac{6}{13}\right)\right) = -\left(\frac{7}{13} (-0.8430) + \frac{6}{13} (-1.1547)\right) = 0.9957$$

بر اساس درختی که داده‌ها را به 3 گروه زیر افراز می‌کنیم:

سن	سابقه	سرطان	سن	سابقه	سرطان	سن	سابقه	سرطان
جوان	+	+	میانسال	+	+	پیر	-	+
جوان	-	-	میانسال	-	+	پیر	+	+
جوان	+	-	میانسال	-	+	پیر	-	+
جوان	-	-	میانسال	-	+			
جوان	+	+	میانسال	-	+			
			میانسال	-	+			

(1) (2) (3)

$$H(C_1) = -\left(\frac{2}{5} \log_2 \left(\frac{2}{5}\right) + \frac{3}{5} \log_2 \left(\frac{3}{5}\right)\right) = 0.9709$$

$$H(C_2) = -\left(\frac{2}{5} \log_2 \left(\frac{2}{5}\right) + \frac{3}{5} \log_2 \left(\frac{3}{5}\right)\right) = 0.9709$$

$$H(C_3) = 0$$

$$\text{Gain (سن)} = H(X) - \frac{3}{13} (H(C_1)) - \frac{5}{13} (H(C_2)) - \frac{3}{13} (H(C_3))$$

پس داده‌ها را بر اساس سابقه سرطان به 2 بخش افراز می‌کنیم:

سن	سابقه	سرطان	سن	سابقه	سرطان
پیر	-	+	میانسال	+	+
میانسال	-	+	پیر	+	+
میانسال	-	-	جوان	+	+
جوان	-	-	جوان	+	-
پیر	-	+	جوان	+	+
جوان	-	-			
میانسال	-	-			
میانسال	-	-			

(1)

$$H(C_1) = -\left(\frac{3}{8} \log_2 \left(\frac{3}{8}\right) + \frac{5}{8} \log_2 \left(\frac{5}{8}\right)\right) = 0.95443$$

$$H(C_2) = -\left(\frac{4}{5} \log_2 \left(\frac{4}{5}\right) + \frac{1}{5} \log_2 \left(\frac{1}{5}\right)\right) = 0.7239$$

$$\text{Gain (سابقه)} = H(X) - \frac{8}{13} (H(C_1)) - \frac{5}{13} (H(C_2)) = 0.1299$$

$$\text{Gain (سن)} > \text{Gain (سابقه)}$$

پس درختی که درختی سن بر حسب سنی داریم.