

۱-

الف) برای یک مجموعه داده فرمول آنتروپی به صورت زیر است.

$$\text{Entropy}(S) = - \sum_{i=1}^n p_i \log_2(p_i)$$

از ۹ داده موجود ۵ داده متعلق به یک کلاس و ۴ داده متعلق به کلاس دیگر هستند در نتیجه با محاسبه مقدار آنتروپی 0.99 محاسبه می‌شود.

ب) برای محاسبه Information Gain ابتدا باید آنتروپی کل داده را محاسبه کنیم که در مرحله قبل انجام دادیم. سپس به ازای تقسیم براساس فیچر مورد نظر آنتروپی وزن دار فرزندهای نود مورد نظر را محاسبه می‌کنیم. در نهایت اختلاف این دو عدد Information Gain است.

$$\text{Weighted Entropy} = \sum \left(\frac{\text{Count of class}}{\text{Total count}} \times \text{Entropy of class} \right)$$

$$\text{Information Gain} = \text{Total Entropy} - \text{Weighted Entropy}$$

برای a1 میزان $\text{Weighted Entropy} = 0.761$

در نتیجه $\text{Information Gain} = 0.991 - 0.761 = 0.229$

برای a2 میزان $\text{Weighted Entropy} = 0.984$

در نتیجه $\text{Information Gain} = 0.991 - 0.984 = 0.007$

ج) مقدار 5 را که میانه مقادیر a3 است را به عنوان حد آستانه در نظر می‌گیریم. در نتیجه ۵ داده کمتر یا مساوی حد آستانه هستند و ۴ داده بیشتر از حد آستانه هستند و در دسته دوم قرار می‌گیرند حالا آنتروپی وزن دار را برای این داده ها محاسبه می‌کنیم.

$$\text{Weighted Entropy}_{a3} = \left(\frac{4}{9} \times \text{Entropy}_{\text{below}} \right) + \left(\frac{5}{9} \times \text{Entropy}_{\text{above}} \right)$$

$\text{Weighted Entropy} = 0.984$

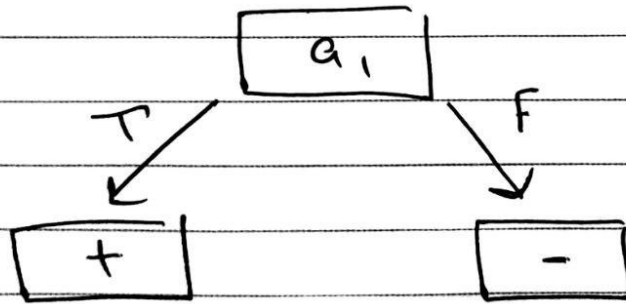
$\text{Information Gain} = 0.991 - 0.984 = 0.007$

(د)

تاریخ: / / ۱۳۹۹

موضوع:

در درخت تصمیم باید ویژگی را انتخاب کنیم که بیشترین 'info gain' را داشته باشد. در اینجا a_1 ویژگی مورد نظر است.



CS Scanned with CamScanner

-۲
(الف)

الهام از $\min \|XB - y\|$ تبدیل می شود.

در مربعات خطا به مقادیر بزرگ انبساط بیشتری داده می شود.

در نتیجه outliers تأثیر بیشتری در پاسخ خواهند داشت.

اگر قدر مطلق خطا را در نظر بگیریم تابع ما نسبت به outliers ها

تأثیر کمتری خواهد داشت اما مشتق پذیر نخواهد بود.

CS Scanned with CamScanner

(ب)

موضوع: تاریخ: / / ۱۳۹۹

منای که از gradient descent استفاده می کنیم باید مشتق بگیریم
اما در قدر مطلق تابع ما مشتق پذیر نیست به همین خاطر نمی توان
از gradient descent استفاده کرد.

(ج)

$$\begin{bmatrix} 1 & x_{A1}^T \\ 1 & x_{A2}^T \\ 1 & x_{A3}^T \\ 1 & x_{A4}^T \\ 1 & x_{A5}^T \end{bmatrix} \quad \begin{bmatrix} 1 & x_{B1}^T \\ \vdots \\ 1 \end{bmatrix} \quad \begin{bmatrix} 1 & x_{C1}^T \\ \vdots \\ 1 \end{bmatrix}$$

$X_A \quad X_B \quad X_C$

بنابراین می توانیم بنویسیم $X_A \beta_A = y_A$ و $X_B \beta_B = y_B$

$X_C \beta_C = y_C$ پس مقادیر β را برای هر ۳ گروه به دست می آوریم

حالا داده جدید خود را مانند y برای y_B ، y_C

y_B می دهیم و کمترین اورد را به عنوان کلاس داده y انتخاب می کنیم
می توان گفت که

$$\text{label}(y) = \text{Arg min} (|y \beta_j - n(j)|)$$

معادله به شکل زیر می شود:

$$y = 1,51x + 45,47$$

برای بدست آوردن a از معادله زیر استفاده می کنیم:

$$a = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

برای بدست آوردن b از معادله زیر استفاده می کنیم:

$$b = \bar{y} - a\bar{x}$$

برای $n = 40$ مقدار y برابر $105,185$ می شود

در حالت اول این احتمال بیشتر است. علت آن هم تفاوت

دقت در داده‌های آموزش و داده‌های تست است.

در حالت دوم با اینکه دقت در داده‌های آموزش کم‌تر است اما

مدل در داده‌های تست هم دقتی در حد داده‌های آموزش دارد.

و شاهد افت دقت نیستیم این نشان می‌دهد که مدل تعمیم‌پذیری

خوب دارد اما در حالت اول مدل دچار overfit شده

و تعمیم‌پذیری خوبی ندارند.

۵- منظم‌سازی یک تکنیک مهم در مدل‌های یادگیری ماشین است که برای جلوگیری از بیش‌برازش به کار می‌رود. بیش‌برازش زمانی رخ می‌دهد که یک مدل به شدت به داده‌های آموزشی خود وابسته شده و قابلیت تعمیم به داده‌های جدید را از دست می‌دهد. دو نوع اصلی منظم‌سازی وجود دارد که به نام‌های $L1$ و $L2$ شناخته می‌شوند.

منظم‌سازی Lasso Regression: منظم‌سازی $L1$ با افزودن جمع مطلق ضرایب مدل به تابع هزینه عمل می‌کند. این روش می‌تواند منجر به صفر شدن برخی ضرایب شود، که به این ترتیب، مدل را به سمت انتخاب ویژگی‌ها سوق می‌دهد. در نتیجه، $Lasso$ می‌تواند به عنوان یک روش انتخاب ویژگی نیز عمل کند و مدل‌های پراکنده‌تری تولید نماید. فرمول منظم‌سازی $L1$ به شکل زیر است:

$$\text{Cost Function} = \text{MSE} + \lambda \sum_{i=1}^n |w_i|$$

که در آن MSE میانگین مربعات خطا، λ پارامتر منظم‌سازی و w_i ضرایب مدل هستند.

منظم‌سازی Ridge Regression: در مقابل، منظم‌سازی $L2$ با اضافه کردن جمع مربعات ضرایب به تابع هزینه عمل می‌کند. این روش موجب می‌شود که همه ضرایب به صورت متعادل کوچک شوند اما به ندرت به صفر می‌رسند. این امر به کاهش واریانس کمک می‌کند بدون آنکه موجب حذف ویژگی‌ها شود. فرمول منظم‌سازی $L2$ به شکل زیر است:

$$\text{Cost Function} = \text{MSE} + \lambda \sum_{i=1}^n w_i^2$$

در اینجا نیز λ پارامتر منظم‌سازی است که میزان تاثیر منظم‌سازی را کنترل می‌کند.

تفاوت اصلی بین این دو روش در نوع جریمه ضرایب است. $L1$ می‌تواند ضرایب را به صفر برساند و ویژگی‌ها را حذف کند، در حالی که $L2$ تمایل دارد ضرایب را به صورت متعادل کاهش دهد. انتخاب بین این دو روش بستگی به داده‌ها و مسئله خاصی دارد که مدل باید حل کند.

تاریخ: / / ۱۳۹۹

موضوع:

ابتدا داده ها را به صورت جدول درمی آوریم.

(white, small, wood)

(white, small, wood)

(white, small, wood)

(white, big, wood)

(white, small, plastic)

(black, big, plastic)

(black, big, plastic)

(black, big, plastic)

ابتدا Entropy برای نور ریشه را به دست می آوریم.

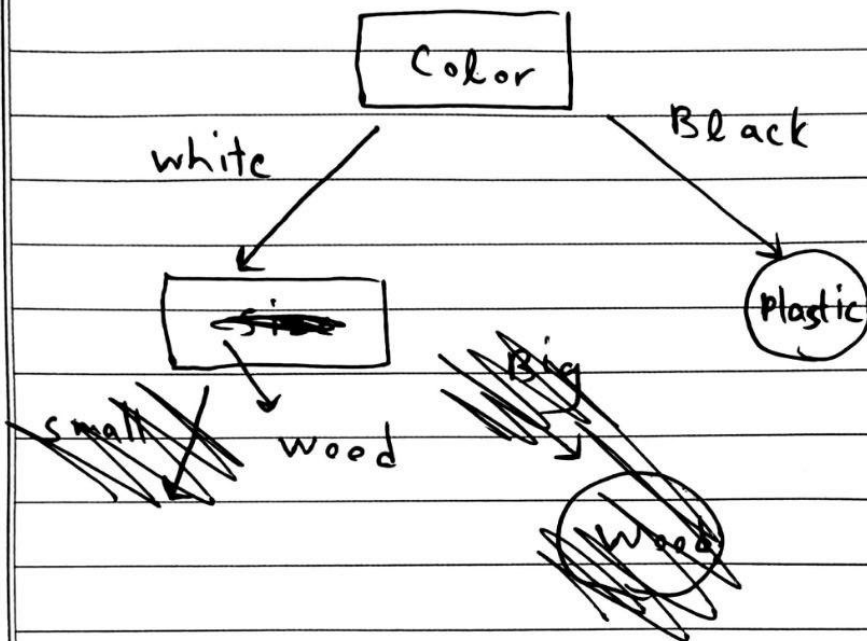
$$\text{Entropy} = -\frac{1}{4} \log \frac{4}{8} - \frac{1}{4} \log \frac{4}{8} = 1$$

gain را برای color و size به دست می آوریم.

$$\text{color: Entropy} = 0.451 \quad \text{gain} = 1 - 0.451 = 0.549$$

$$\text{size: Entropy} = 0.181 \quad \text{gain} = 1 - 0.181 = 0.819$$

انتاده از color ، gain بیشتری دارد. برای همین اذل
از color دپس از size انتخاب می‌کنیم .



در نتیجه تنها با انتخاب از Color میتوان درست تصمیم
را کامل کرد .

$$Gini(Node) = 1 - \sum_{i=1}^n p_i^2$$

این فرمول برای معیار GINI برای یک نود خاص است.
برای اینکه معیار GINI یک Split را پیدا کنیم. از فرمول زیر استفاده می‌کنیم.

$$Gini(Split) = \sum \left(\frac{|Node|}{|Split|} \times Gini(Node) \right)$$

این فرمول در واقع میانگین وزن‌دار معیار GINI برای نودهای حاصل از Split است.

(الف)

تاریخ: / / ۱۳۹۹

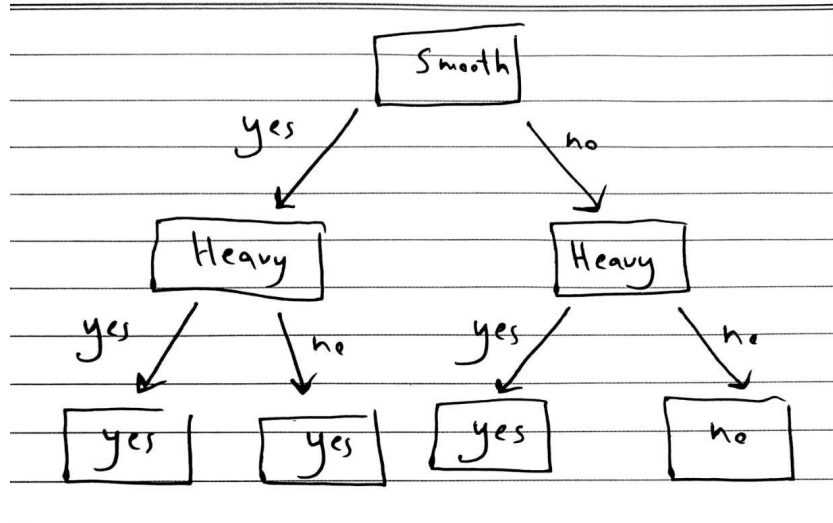
موضوع:

$$\begin{aligned} Gini(Heavy) &= \frac{3}{8} \left(1 - \left(\frac{1}{3} \right)^2 - \left(\frac{2}{3} \right)^2 \right) \\ &+ \frac{5}{8} \left(1 - \left(\frac{2}{5} \right)^2 - \left(\frac{3}{5} \right)^2 \right) = \underline{0.444} \end{aligned}$$

$$\begin{aligned} Gini(Spotted) &= \frac{3}{8} \left(1 - \left(\frac{1}{3} \right)^2 - \left(\frac{2}{3} \right)^2 \right) \\ &+ \frac{5}{8} \left(1 - \left(\frac{2}{5} \right)^2 - \left(\frac{3}{5} \right)^2 \right) = \underline{0.444} \end{aligned}$$

$$\begin{aligned} Gini(Smooth) &= \frac{4}{8} \left(1 - \left(\frac{3}{4} \right)^2 - \left(\frac{1}{4} \right)^2 \right) \\ &+ \frac{4}{8} \left(1 - \left(\frac{2}{4} \right)^2 - \left(\frac{2}{4} \right)^2 \right) = \underline{0.437} \end{aligned}$$

چون معیار Smooth کمترین Gini را دارد از آن
به عنوان best استفاده می‌کنیم.

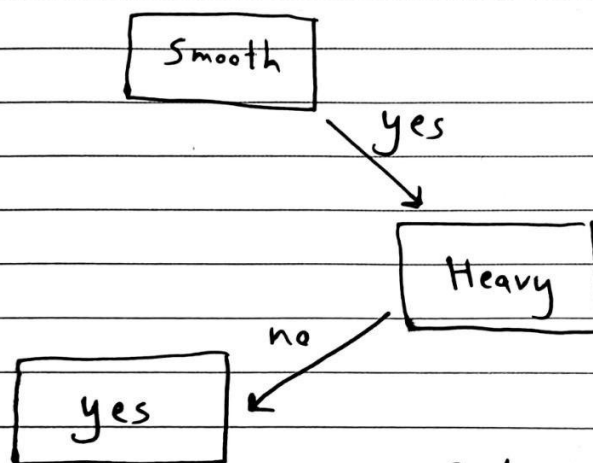


(د)

$$E(\text{Parent}) = -\frac{3}{8} \log \frac{3}{8} - \frac{5}{8} \log \frac{5}{8} = 0.911$$

$$\text{gain} = 0.911 - \frac{4}{8} (0.187) - \frac{4}{8} (1) = 0.424$$

برای داده خامه ۹ از روی درخت به دست آمده در بخش قبل



پایش می کنیم

در نتیجه قارچ می افتد.