دانشگاه صنعتی شریف
دانشکده‌ی مهندسی کامپیوتر

هوش مصنوعی

پاییز ۱۳۹۹

استاد: محمدحسین رهبان

پاسخ تمرین پنجم، بخش دوم — مقدمه ای بر یادگیری ماشین و درخت تصمیم گیری — مهلت ارسال: ـ

سوالات نظری (۸۵ نمره)

۱. (۴۰ نمره)

(آ)

اگر تابع B را به شکل زیر تعریف کنیم،

$$B(q) = -(qlog_2 1 + (1-q)log_2(1-q))$$

می‌توانیم انتروپی را به شکل زیر بنویسیم:

$$H(\text{Goal}) = B(\frac{p}{p+n})$$

حالا اگر تابع Remainder را به شکل زیر تعریف کنیم،

$$Remainder(A) = \sum_{k=1}^{d} \frac{p_k + n_k}{p+n} B(\frac{p_k}{p_k + n_k})$$

تابع Gain به شکل زیر تعریف می‌شود:

$$Gain(A) = B(\frac{p}{p+n}) - Remainder(A)$$

حالا برای هر راس به ترتیب ویژگی‌ها را بررسی و ویژگی‌ای را انتخاب می‌کنیم که بیشترین Gain را داشته باشد.
از آن جا که قسمت $B(\frac{p}{p+n})$ برای همه ویژگی‌های یک راس ثابت است، کافیست ویژگی‌ای انتخاب شود که Remainder کمتری دارد.
برای ریشه قبل از split داریم:

برچسب	تعداد
P	۹
N	۵

ویژگی age:

حالت	P	N	B
≤ 30	۲	۳	$B(\frac{2}{5}) = 0.97$
31...40	۴	۰	$B(\frac{4}{4}) = 0$
> 40	۳	۲	$B(\frac{3}{5}) = 0.97$

$\rightarrow Remainder = 0.6928$

ویژگی income:

حالت	P	N	B
High	۲	۲	$B(\frac{2}{4}) = 1$
Medium	۴	۲	$B(\frac{4}{6}) = 0.92$
Low	۳	۱	$B(\frac{3}{4}) = 0.81$

$\rightarrow Remainder = 0.9114$

ویژگی Student:

حالت	P	N	B
No	۳	۴	$B(\frac{3}{7}) = 0.98$
Yes	۶	۱	$B(\frac{6}{7}) = 0.59$

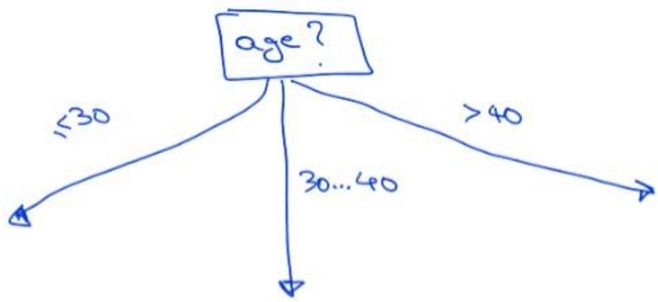
$\rightarrow Remainder = 0.785$

ویژگی Credit Rating:

حالت	P	N	B
Fair	۶	۲	$B(\frac{6}{8}) = 0.81$
Excellent	۳	۳	$B(\frac{3}{6}) = 1$

$\rightarrow Remainder = 0.8914$

به این ترتیب ویژگی age را برای split انتخاب می‌کنیم. درخت تصمیم‌گیری تا به اینجای کار:



برای راس سمت چپ یعنی $age \leq 30$ ویژگی مناسب را پیدا می‌کنیم:

ویژگی income:

حالت	P	N	B
High	۰	۲	$B(\frac{0}{2}) = 0$
Medium	۱	۱	$B(\frac{1}{2}) = 1$
Low	۱	۰	$B(\frac{1}{1}) = 0$

$$\rightarrow Remainder = 0.4$$

ویژگی Student:

حالت	P	N	B
No	۰	۳	$B(\frac{0}{3}) = 0$
Yes	۲	۰	$B(\frac{2}{2}) = 0$

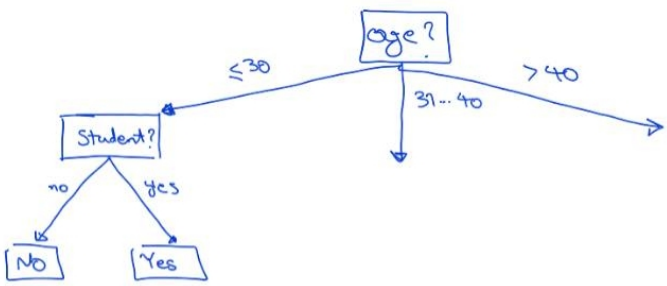
$$\rightarrow Remainder = 0$$

ویژگی Credit Rating:

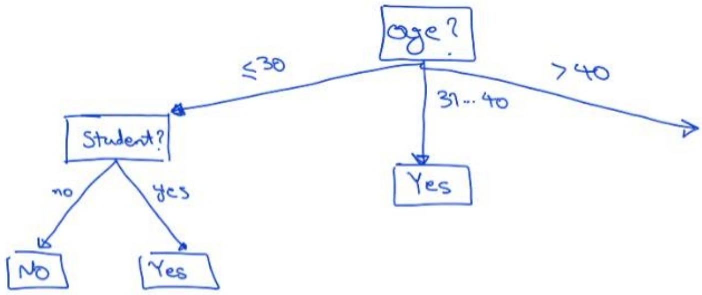
حالت	P	N	B
Fair	۱	۲	$B(\frac{1}{3}) = 0.92$
Excellent	۱	۱	$B(\frac{1}{2}) = 1$

$$\rightarrow Remainder = 0.95$$

به این ترتیب ویژگی student را برای split انتخاب می‌کنیم. این ویژگی در این شاخه به طور کامل نتیجه را مشخص می‌کند. درخت تصمیم‌گیری تا به اینجای کار:



برای راس وسط یعنی حالت $age = 31...40$ چون همه برچسب‌ها مثبت است، نیازی به split بیشتر نیست. درخت تصمیم‌گیری تا اینجای کار:



برای راس سمت راست یعنی حالت $age > 40$ ویژگی مناسب را برای split پیدا می‌کنیم:
ویژگی income:

حالت	P	N	B
Medium	۲	۱	$B(\frac{2}{3}) = 0.92$
Low	۱	۱	$B(\frac{1}{2}) = 1$

$$\rightarrow Remainder = 0.952$$

ویژگی Student:

حالت	P	N	B
No	۱	۱	$B(\frac{1}{2}) = 1$
Yes	۲	۱	$B(\frac{2}{3}) = 0.92$

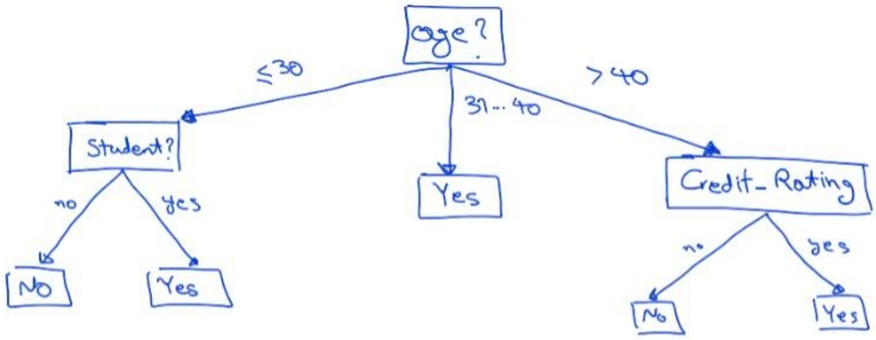
$$\rightarrow Remainder = 0.952$$

ویژگی Credit Rating:

حالت	P	N	B
Fair	۳	۰	$B(\frac{3}{3}) = 0$
Excellent	۰	۲	$B(\frac{0}{2}) = 0$

$$\rightarrow Remainder = 0$$

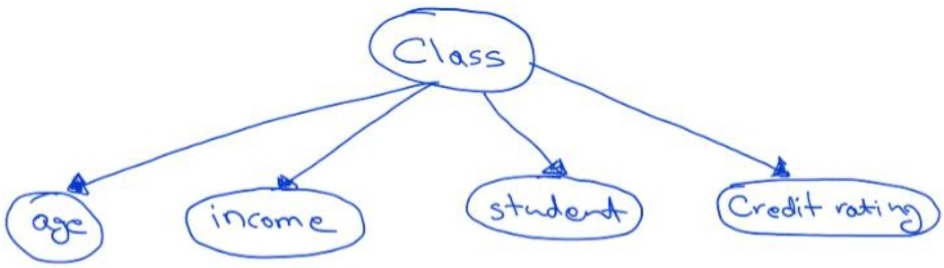
بنابراین split با استفاده از credit rating انجام می‌شود و داده را به طور کامل دسته‌بندی می‌کند. درخت نهایی:



از آن جا که داده به طور کامل دسته‌بندی شده و miss classification نداریم، accuracy صد درصد خواهد بود.

(ب)

گراف Naive Bayes به شکل زیر خواهد بود:



ویژگی‌ها هم شامل age ، income ، student و credit_rating می‌شود.

(ج)

Maximum Likelihood در اینجا برای هر یک از احتمالات شرطی به شکل زیر محاسبه می‌شود. فرض می‌کنیم احتمال هر یک از حالات ویژگی f به شرط داشتن class به شکل زیر تعریف شود:

$$[x_1, x_2, ..., x_k]$$

به طوری که $\sum_{i=1}^{k} x_i = 1$. حالا رابطه Maximum Likelihood را می‌نویسیم:

$$argmax_x P(D|x) = argmax_x \prod_{i=1}^{k} x_i^{c_i}$$

که c_i تعداد تکرار حالت i با داشتن class است:

$$\sum_{i=1}^{k} c_i = n$$

در نتیجه:

$$x_i = \frac{c_i}{n}$$

حالا این مقدار را برای ویژگی‌ها مختلف حساب می‌کنیم: $P(age|class)$:

class	age	احتمال
yes	≤ 30	$\frac{2}{9}$
yes	31..40	$\frac{4}{9}$
yes	> 40	$\frac{3}{9}$
no	≤ 30	$\frac{3}{5}$
no	31...40	$\frac{0}{5}$
no	> 40	$\frac{2}{5}$

برای $P(income|class)$:

class	income	احتمال
yes	high	$\frac{2}{9}$
yes	medium	$\frac{4}{9}$
yes	low	$\frac{3}{9}$
no	high	$\frac{2}{5}$
no	medium	$\frac{2}{5}$
no	low	$\frac{1}{5}$

برای $P(student|class)$:

class	student	احتمال
yes	yes	$\frac{6}{9}$
yes	no	$\frac{3}{9}$
no	yes	$\frac{1}{5}$
no	no	$\frac{4}{5}$

برای $P(credit\_ratin|class)$:

class	credit_rating	احتمال
yes	fair	$\frac{6}{9}$
yes	excellent	$\frac{3}{9}$
no	fair	$\frac{2}{5}$
no	excellent	$\frac{3}{5}$

(د)

برای محاسبه MAP برای احتمال‌های شرطی، فرض می‌کنیم که احتمال هر یک از حالات ویژگی f با داشتن class به شکل زیر است:

$$[x_1, x_2, ..., x_k]$$

از آن‌جا که توزیع prior دیریکله با متغیرهای برابر ۲ است، داریم:

$$
\begin{aligned}
P(X|D) &\sim P(D|X)P(X) \\
&\sim \prod_{i=1}^{k} x_i^{c_i} \prod_{i=1}^{k} x_i^{\alpha_i - 1} \\
&\sim \prod_{i=1}^{k} x_i^{c_i + 1}
\end{aligned}
$$

که مشابه laplace smoothing به اندازه ۱ واحد است.
:$P(age|class)$

class	age	احتمال
yes	≤ 30	$\frac{2}{9}$
yes	31..40	$\frac{4}{9}$
yes	> 40	$\frac{3}{9}$
no	≤ 30	$\frac{3}{5}$
no	31...40	$\frac{0}{5}$
no	> 40	$\frac{2}{5}$

برای $P(income|class)$:

class	income	احتمال
yes	high	$\frac{3}{12}$
yes	medium	$\frac{5}{12}$
yes	low	$\frac{4}{12}$
no	high	$\frac{3}{8}$
no	medium	$\frac{3}{8}$
no	low	$\frac{2}{8}$

برای $P(student|class)$:

class	student	احتمال
yes	yes	$\frac{7}{11}$
yes	no	$\frac{4}{11}$
no	yes	$\frac{2}{7}$
no	no	$\frac{5}{7}$

برای $P(credit\_ratin|class)$:

class	credit_rating	احتمال
yes	fair	$\frac{7}{11}$
yes	excellent	$\frac{4}{11}$
no	fair	$\frac{3}{7}$
no	excellent	$\frac{4}{7}$

(۵)

درخت تصمیم‌گیری هیچ اطلاعاتی در مورد توزیع احتمالاتی ویژگی‌ها ندارد. برای تولید داده جدید، نیاز است نسبت مقدار هر ویژگی متناسب با احتمال وقوع آن مقدار برای ویژگی مورد نظر باشد. چنین اطلاعاتی در درخت تصمیم‌گیری وجود ندارد اما می‌توان این اطلاعات را از توزیع احتمالاتی ویژگی استخراج کرد.

۲. (۲۰ نمره)

ابتدا نشان می‌دهیم زمانی که نسبت $\frac{p_k}{p_k+n_k}$ برای همه k ها ثابت و برابر است، information gain صفر می‌شود.

$$\begin{aligned}
\frac{p_k}{p_k+n_k} = A &\Rightarrow p_k = A(p_k+n_k) \\
&\Rightarrow p_k = -\frac{A}{A-1}n_k \\
&\Rightarrow p = \sum_i p_k = -\frac{A}{A-1}\sum_i n_k \\
&\Rightarrow p = -\frac{A}{A-1}n \\
&\Rightarrow (A-1)p = -An \\
&\Rightarrow A(n+p) = p \\
&\Rightarrow \frac{p}{n+p} = A = \frac{p_k}{p_k+n_k}
\end{aligned}$$

از طرفی Gain به شکل زیر تعریف می‌شد:

$$B(\frac{p}{p+n}) - \sum_{k=1}^{d} \frac{p_k+n_k}{p+n} B(\frac{p_k}{p_k+n_k})$$

با جاگذاری می‌نویسیم:

$$\text{Gain} = B(\frac{p}{p+n}) - B(\frac{p}{p+n})\sum_{k=1}^{d}\frac{p_k+n_k}{p+n} = B(\frac{p}{p+n}) - B(\frac{p}{p+n}) = 0$$

حالا برای نشان‌دادن این‌که Gain تنها در همین نقطه صفر شده و در بقیه نقاط مثبت است، کافیست ثابت کنیم که این نقطه مینیموم تابع Gain با شرط

$$\sum_{i=1}^{k} p_i = p$$

و

$$\sum_{i=1}^{k} n_i = n$$

است. برای این کار تابع زیر را می‌نویسیم و گردایان آن را برابر صفر می‌گذاریم. برابری مشتق بر حسب λ_1 و λ_2 شرط‌های گفته شده را تضمین می‌کند.

$$g = B(\frac{p}{p+n}) - \sum_k \frac{p_k+n_k}{p+n} B(\frac{p_k}{p_k+n_k}) + \lambda_1\left(p - \sum_k p_k\right) + \lambda_2\left(n - \sum_k n_k\right)$$

مشتق‌ها را بر حسب p_k و n_k برابر صفر می‌گذاریم:

$$\frac{\partial g}{\partial p_k} = -\frac{1}{p+n}B(\frac{p_k}{p_k+n_k}) - \frac{p_k+n_k}{p+n}\log\frac{p_k}{n_k}\left(\frac{1}{p_k+n_k} - \frac{p_k}{(p_k+n_k)^2}\right) - \lambda_1 = 0$$

$$\frac{\partial g}{\partial n_k} = -\frac{1}{p+n}B(\frac{p_k}{p_k+n_k}) - \frac{p_k+n_k}{p+n}\log\frac{p_k}{n_k}\left(\frac{-p_k}{(p_k+n_k)^2}\right) - \lambda_2 = 0$$

این دو رابطه را از هم کم می‌کنیم:

$$\log(\frac{p_k}{n_k}) - \frac{p+n}{\lambda_2-\lambda_1} = 0 \Rightarrow \log(\frac{p_k}{n_k}) = \text{Constant}$$

به این ترتیب در نقطه مینیموم نسبت $\frac{p_k}{n_k}$ ثابت است. قبل‌تر هم نشان دادیم با ثابت بودن این نسبت، تابع Gain صفر می‌شود. بنابراین تابع Gain تنها در حالت گفته شده صفر می‌شود و در بقیه نقاط مثبت است.

۳. (۲۵ نمره)

$$P(y|x) \sim p(y)p(x|y) \Rightarrow p(y=i|x) \sim \frac{1}{3}P(x|y=i)$$

(آ)

برای $i=1$:

$$P(y=1|x) \sim \frac{1}{6\pi}(0.49)^{\frac{-1}{2}}exp(-\frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1))$$

از طرفی

$$x-\mu_1 = \begin{bmatrix}-0.5\\0.5\end{bmatrix}$$

و

$$\Sigma^{-1} = \begin{bmatrix}1.42857 & -0.00000\\0.00000 & 1.42857\end{bmatrix}$$

پس:

$$P(y=1|x) = \frac{1}{6\pi}\frac{10}{7}exp(-\frac{1}{2}0.71429) = 0.053027$$

برای $i=2$:

$$P(y=2|x) \sim \frac{1}{6\pi}(0.07)^{\frac{-1}{2}}exp(-\frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2))$$

از طرفی

$$x-\mu_2 = \begin{bmatrix}-1.5\\-0.5\end{bmatrix}$$

و

$$\Sigma^{-1} = \begin{bmatrix}2.8571 & -4.2857\\-4.2857 & 11.4286\end{bmatrix}$$

پس:

$$P(y=2|x) \sim \frac{1}{6\pi}(0.7)^{\frac{-1}{2}}exp(-1.4286) = 0.048054$$

برای $i=3$:

$$P(y=3|x) \sim \frac{1}{6\pi}(0.52)^{\frac{-1}{2}}exp(-\frac{1}{2}(x-\mu_3)^T\Sigma_3^{-1}(x-\mu_3))$$

از طرفی

$$x-\mu_3=\begin{bmatrix}0.5\\-0.5\end{bmatrix}$$

و

$$\Sigma^{-1}=\begin{bmatrix}1.53846 & -0.38462\\-0.38462 & 1.3465\end{bmatrix}$$

پس:

$$P(y=3|x)\sim\frac{1}{6\pi}(0.52)^{\frac{-1}{2}}exp(-0.45673)=0.046595$$

بنابراین برچسب داده ۱ است.

(ب)

برای $i=1$:

$$P(y=1|x)\sim\frac{1}{6\pi}(0.7)^{\frac{-1}{2}}exp(-\frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1))$$

از طرفی

$$x-\mu_1=\begin{bmatrix}0.5\\0.5\end{bmatrix}$$

و

$$\Sigma_1^{-1}=\begin{bmatrix}1.42857 & 0\\0 & 1.42857\end{bmatrix}$$

پس:

$$P(y=1|x)\sim\frac{1}{6\pi}(0.49)^{\frac{-1}{2}}exp(-0.35714)=0.053027$$

برای $i=2$:

$$P(y=2|x)\sim\frac{1}{6\pi}(0.07)^{\frac{-1}{2}}exp(-\frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2))$$

از طرفی

$$x-\mu_2=\begin{bmatrix}-0.5\\-0.5\end{bmatrix}$$

و

$$\Sigma_2^{-1}=\begin{bmatrix}2.8571 & -4.2857\\-4.2857 & 11.4286\end{bmatrix}$$

پس:

$$P(y=2|x)\sim\frac{1}{6\pi}(0.07)^{\frac{-1}{2}}exp(-0.71429)=0.098161$$

برای $i=3$:

$$P(y=3|x)\sim\frac{1}{6\pi}(0.52)^{\frac{-1}{2}}exp(-\frac{1}{2}(x-\mu_3)^T\Sigma_3^{-1}(x-\mu_3))$$

از طرفی

$$x-\mu_3=\begin{bmatrix}1.5\\-0.5\end{bmatrix}$$

و

$$\Sigma_3^{-1}=\begin{bmatrix}1.53846 & -0.38462\\-0.38462 & 1.3465\end{bmatrix}$$

پس:

$$P(y=3|x) \sim \frac{1}{6\pi}(0.52)^{\frac{-1}{2}} exp(-2.1875) = 0.0082543$$

به این ترتیب برچسب این داده ۲ است.