



هوش مصنوعی

پاییز ۱۴۰۰

استاد: محمدحسین رهبان

گردآورندگان: کیان باختری، کیمیا نوربخش

بررسی و بازبینی: محمد مهدی ماهری

دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

Intro to Machine Learning

پاسخ تمرین پنجم سری دوم

مهلت ارسال: ۱۸ آذر

سوالات (۱۰۰ نمره)

۱. (۱۰۰ نمره)

(آ) در مسائل حوزه یادگیری ماشین ممکن است یک مشاهده در دادگان آزمون وجود داشته باشد که در دادگان یادگیری وجود ندارد. برای جلوگیری از صفر شدن احتمال آن مشاهده، فرض می‌کنیم در دادگان یادگیری یک مشاهده مشابه وجود دارد.

(ب) بله، به عنوان مثال در داده‌هایی که برچسب آن‌ها Go است، هیچگاه فیچر ستون اول (سمت چپ) برابر با ۱ نشده است. این در حالی است که در کوثری خواسته شده فیچر ستون اول ۱ است و مال احتمال ۱ شدن این فیچر به شرطی که برچسب داده Go باشد را عملاً نداریم.

برای رفع این مشکل از smoothing استفاده می‌کنیم. به این معنی که فرض می‌کنیم به شرط هر برچسب، از هر فیچر یکی بیشتر از تعداد واقعی‌اش وجود دارد. یعنی برای مثال، اگر در واقع پنج داده داریم که برچسبشان Go است و در هیچ یک فیچر ستون اول ۱ نشده است، فرض می‌کنیم که شش داده با برچسب Go داریم و یکی از آن‌ها فیچر ستون اول‌اش ۱ شده است. این کار عملاً معادل با این است که فرض کنیم از هر برچسب یک داده (همان مشاهده یا سطر) داریم که همه‌ی فیچرهایش ۱ شده است. علاوه بر این، با داده‌های واقعی احتمال صفر شدن فیچر چهارم به شرطی که برچسب داده Stop باشد نیز صفر می‌شود. پس لازم است داده‌ای که تمام فیچرهایش صفر است را به شرط هر برچسب به صورت فرضی به داده‌ها اضافه کنیم. در جدول زیر احتمالات شرطی را پیش و پس از smoothing مشاهده می‌کنید.

before smoothing					after smoothing				
f1	f2	f3	f4	label	f1	P(f1 Go)	P(f1 Stop)	P(f1 Go)	P(f1 Stop)
1	1	0	1	Stop	f1	0	3/4	1/7	4/6
0	1	0	1	Stop	f2	3/5	3/4	4/7	4/6
0	1	1	0	Go	f3	1	0	6/7	1/6
1	0	0	1	Stop	f4	2/5	1	3/7	5/6
0	0	1	0	Go					
1	1	0	1	Stop					
0	1	1	1	Go					
0	1	1	0	Go					
0	0	1	1	Go					

L	P(L)
Go	7/13
Stop	6/13

1	1	1	0	?
---	---	---	---	---

با استفاده از رابطه‌ی بیز و استقلال شرطی بین فیچرها به شرط برجسب‌ها، احتمال برجسب کوئری خواسته شده محاسبه می‌شود:

$$P(L \mid Q) \propto P(Q \mid L)P(L)$$

$$P(Q \mid L) = \prod_{i=1}^4 P(f_i \mid L) \rightarrow \begin{cases} P(Q \mid L = Go) = \frac{1}{7} \times \frac{4}{7} \times \frac{6}{7} \times \left(1 - \frac{3}{7}\right) = \frac{96}{2401} \\ P(Q \mid L = Stop) = \frac{4}{6} \times \frac{4}{6} \times \frac{1}{6} \times \left(1 - \frac{5}{6}\right) = \frac{1}{81} \end{cases}$$

$$P(L) = \frac{\text{Count}(L)}{N}$$

$$\Rightarrow \begin{cases} P(L = Go \mid Q) \propto \frac{6}{13} \times \frac{96}{2401} \\ P(L = Stop \mid Q) \propto \frac{7}{13} \times \frac{1}{81} \end{cases}$$

$$\xrightarrow{\text{Normalization}} \begin{cases} P(L = Go \mid Q) = 0.73 \\ P(L = Stop \mid Q) = 0.26 \end{cases}$$