

به نام خدا



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



درس سیستم‌های هوشمند

تمرین شماره 5

نام و نام خانوادگی : محمدرضا بختیاری

شماره دانشجویی : 810197468

دی 1400

فهرست سوالات

- سوال 1 : بیز ساده انگارانه 3
- 1 3
- 2 4
- سوال 2 : یادگیری تقویتی مبتنی بر مدل 5
- الف: پیاده سازی 5
- ب: تحلیلی 6
- سوال 3 : یادگیری تقویتی غیر مبتنی بر مدل 10
- (امتیازی) : 10
- پیوست: 10

سوال 1: بیز ساده انگارانه

در این سوال با استفاده از الگوریتم بیز ساده انگارانه¹ داده های موجود را طبقه بندی می کنیم و در انتها دقت² و ماتریس آشفتگی³ را گزارش می کنیم .

1.

ابتدا احتمال پیشین⁴ را برای هر کلاس طبق رابطه ی زیر به دست می آوریم :

$$\hat{P}(c) = \frac{N_c}{N}$$

سپس برای هر ویژگی ، احتمال به شرط کلاس موجود را از طریق Laplace smoothing محاسبه می کنیم تا در مواقعی که یک ویژگی در یک کلاس وجود نداشت ، احتمال کل برابر صفر نشود :

$$\hat{P}(w_i | c) = \frac{\text{count}(w_i, c) + 1}{\sum_{w \in V} (\text{count}(w, c) + 1)}$$

در انتها با توجه به مستقل بودن ویژگی ها می توان با ضرب کردن احتمال پیشین در احتمال هر ویژگی به شرط کلاس ، داده را به کلاسی اختصاص دهیم که احتمال بیشتری دارد :

$$C_{MAP} = \operatorname{argmax}_c \hat{P}(c) \prod_i \hat{P}(x_i | c)$$

در آخر درصد دقت را برای دادگان آزمون⁵ گزارش می کنیم :

Accuracy is :
99.80916030534351 %

شکل 1-1: درصد دقت برای دادگان آزمون

¹ Naïve Bayes

² Accuracy

³ Confusion Matrix

⁴ Priors Probability

⁵ Test Data

2.

در این قسمت , ماتریس آشفتگی را برای دادگان آزمون گزارش می کنیم :

```
Confusion Matrix is :  
[249, 1]  
[0, 274]
```

شکل 1-2 : ماتریس آشفتگی برای دادگان آزمون

مشاهده می کنیم فقط یکی از قارچ های سمی به اشتباه قارچ خوراکی تشخیص داده شده است .

سوال 2: یادگیری تقویتی مبتنی بر مدل

در این قسمت با استفاده از روش های مبتنی بر مدل پس از جست و جوی محیط و به دست آوردن مدل محیط یک بار پیاده سازی کردی و یک بار نیز به صورت تحلیلی دو مرحله از این الگوریتم را نشان می دهیم .

الف: پیاده سازی

ب: تحلیلی

با استفاده از الگوریتم مبتنی بر تکرار^۱، دو مرحله را برای این مساله جلو می بریم .

گام اول : ابتدا یک سیاست دلخواه^۲ (در اینجا به صورت رندم در گام اول جهت حرکت در هر خانه را مشخص می کنیم) انتخاب می کنیم .

سپس گام دوم و سوم را انقدر انجام می دهیم تا به همگرایی رسیده و در انتخاب بعدی جهت هیچ خانه ای تغییر نکند (در اینجا فقط دو مرحله تکرار را جلو می بریم)

گام دوم : مقدار^۳ هر خانه را با استفاده از رابطه زیر تغییر می دهیم (مقادیر اولیه خانه ها را به صورت پیش فرض ، صفر در نظر می گیریم)

$$V_{\pi}(s) = R_s^{\pi(s)} + \gamma \sum_{s'} P_{ss'}^{\pi(s)} V_{\pi}(s')$$

که در آن $R_s^{\pi(s)}$ مقدار پاداش (در اینجا مجازات) در هر مرحله است که صفر در نظر گرفته شده است و مقدار μ نیز برابر 0.2 در نظر می گیریم .

گام سوم : پس از مشخص شدن مقدار هر خانه ، سیاست قبلی را بهبود بخشیده و با رابطه ی زیر سیاست جدیدی را برای حرکت در نظر می گیریم .

$$\pi'(s) := \arg \max_a R_s^a + \gamma \sum_{s'} P_{ss'}^a V_{\pi}(s')$$

سیاست اولیه را به این شکل زیر در نظر می گیریم و سپس مقدار هر خانه را به روز رسانی می کنیم :

0	0	0.12	3
0	0	-0.08	-2
0		0	-0.24

←	↓	↓	3
→	→	↑	-2
↓		→	↑

¹ Policy Iteration

²Arbitrary policy π'

³ Value

ارزش های به دست آمده در قسمت قبل را به این صورت محاسبه کردیم :

$$V(1,4) = 0.2*0.6*(-2) = -0.24$$

$$V(3,2) = 0.2*0.2*(-2) = -0.08$$

$$V(3,2) = 0.2*0.2*(3) = 0.12$$

در ادامه وقت آن می رسد که با ارزش های به دست آمده ، سیاست خود را به روز رسانی کنیم .

برای خانه (2.3) داریم :

$$V(2,3)(\downarrow) = 0.2*0.2*(0.12) = 0.0048$$

$$V(2,3)(\rightarrow) = 0.2*0.6*(0.12) = 0.0144$$

$$V(2,3)(\leftarrow) = 0$$

$$V(2,3)(\uparrow) = 0.2*0.2*(0.12) = 0.0048$$

پس سیاست جدید برای این خانه ، سمت راست است که بیشترین ارزش را دارد .

برای خانه (2.2) داریم :

$$V(2,2)(\downarrow) = 0.2*0.2*(-0.08) = -0.0032$$

$$V(2,2)(\rightarrow) = 0.2*0.6*(-0.08) = -0.0096$$

$$V(2,2)(\leftarrow) = 0$$

$$V(2,2)(\uparrow) = 0.2*0.2*(-0.08) = -0.0032$$

پس سیاست جدید برای این خانه ، سمت بالا یا پایین است که هر بیشترین ارزش را دارند (ما به صورت دلخواه بالا را انتخاب می کنیم)

برای خانه (3.1) داریم :

$$V(3,1)(\downarrow) = 0.2*0.2*(-0.24) = -0.0096$$

$$V(3,1)(\rightarrow) = 0.2*0.6*(-0.24) + 0.2*0.2*(-0.08) = -0.032$$

$$V(3,1)(\leftarrow) = 0.2*0.2*(-0.08) = -0.0032$$

$$V(3,1)(\uparrow) = 0.2*0.6*(-0.08) + 0.2*0.2*(-0.24) = -0.0192$$

پس سیاست جدید برای این خانه ، سمت چپ است که بیشترین ارزش را دارد .

خانه های $V(1.1)$, $V(1.2)$, $V(1.3)$ نیز بدون تغییر سیاست به کار خود ادامه می دهند .

همچنین سیاست بهینه برای حرکت در خانه های $V(3,3)$, $V(3,2)$, $V(4,1)$ به وضوح به ترتیب برابر با : پایین ، چپ و راست خواهد بود .

در نتیجه سیاست به روز رسانی شده و مقادیر جدید نیز به شکل زیر خواهند بود :

0	0.0144	0.12	3
0	-0.0032	-0.08	-2
0		-0.0032	-0.24

←	→	→	3
→	↑	←	-2
↓		←	↓

تکرار اول تمام شده و در شروع تکرار دوم ، سیاست 3 خانه ی $V(1.1)$, $V(1.2)$, $V(1.3)$ را به روز رسانی می کنیم :

برای خانه (1.3) داریم :

$$V(1,3)(\downarrow) = 0.2*0.2*(0.0144) = 0.000567$$

$$V(1,3)(\rightarrow) = 0.2*0.6*(0.0144) = 0.001728$$

$$V(1,3)(\leftarrow) = 0$$

$$V(1,3)(\uparrow) = 0.2*0.2*(0.0144) = 0.000567$$

پس سیاست جدید برای این خانه ، سمت راست است که بیشترین ارزش را دارد .

برای خانه (1.2) داریم :

$$V(1,2)(\downarrow) = 0.2*0.2*(-0.0032) = -0.000128$$

$$V(1,2)(\rightarrow) = 0.2*0.6*(-0.0032) = -0.000384$$

$$V(1,2)(\leftarrow) = 0$$

$$V(1,2)(\uparrow) = 0.2*0.2*(-0.0032) = -0.000128$$

پس سیاست جدید برای این خانه ، سمت چپ است که بیشترین ارزش را دارد .

پس از دو تکرار ، سیاست برای نقطه ی $V(1,1)$ کماکان تغییر نکرده و ارزش آن خانه کماکان صفر می باشد .

در نهایت پس از دو تکرار , سیاست نهایی به شکل زیر می باشد که نشان می دهد حرکت خانه ها نسبت به مقدار اولیه ی خود بهبود یافته است :

→	→	→	3
←	↑	←	-2
↓		←	↓

سوال 3: یادگیری تقویتی غیر مبتنی بر مدل

(امتیازی):

پیوست: