مهدی فیروزبخت

400131027

تمرین سری دوم – شناسایی آماری الگو

(۱

① C_i = کلاس درب ها

ⓐ

x_j = درب انتخاب شده

Prior Probability: $P(C_1)\ ,\ P(C_2)$ ⇒ احتمال اولیه هر کلاس

Posterior Probability: $P(C_i|x_j)$ ⇒ احتمال اینکه پشت درب مشخص شده کدام کلاس است

ⓑ احتمال هر کلاس ⇒ $P(C_1)=\frac{1}{3}\ ,\ P(C_2)=\frac{1}{3}\ ,\ P(C_3)=\frac{1}{3}$

ⓒ y = Host selection

Car → #2 here , $x=2$, $y=3$

$$\Rightarrow P(y=3 \mid c, x=2) \Rightarrow \begin{cases} 1 & \text{if } c=1 \\ \frac{1}{2} & \text{if } c=2 \\ 0 & \text{if } c=3 \end{cases}$$

(D) $$P(C=2 \mid y=3) = \frac{P(y=3 \mid C=2)\, P(C=2)}{P(y=3)}$$

$$\Rightarrow \frac{\frac{1}{2} \times \frac{1}{3}}{P(y=3 \mid C=1)P(C=1) + P(y=3 \mid C=2)P(C=2) + P(y=3 \mid C=3)P(C=3)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

$$P(C=1 \mid y=3) = \frac{P(y=3 \mid C=1)\, P(C=1)}{P(y=3)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

$$P(C=3 \mid y=3) = 0$$

(e) در حالت اولیه احتمال هر کلاس $\frac{1}{3}$ می‌باشد، یعنی احتمال پیروزی ما $\frac{1}{3}$ است. سپس با توجه به تغییراتی که میزبان در انتخاب ما انجام می‌دهد، اگر درب خود را تغییر دهیم احتمال پیروزی $\frac{2}{3}$ می‌شود.

(f) چون میزبان اطلاعی از جایزه پشت درها ندارد احتمال تمام درها در تمام حالات برابر است.

(g) $$P(y=3 \mid C, X=2) \Rightarrow \begin{cases} \frac{1}{2} & C=1 \\ \frac{1}{2} & C=2 \\ \frac{1}{2} & C=3 \end{cases}$$ $\Rightarrow$ احتمال بودن یا نبودن است

$$P(C=1 \mid y=3) = \frac{P(y=3 \mid C=1)\, P(C=1)}{P(y=3)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

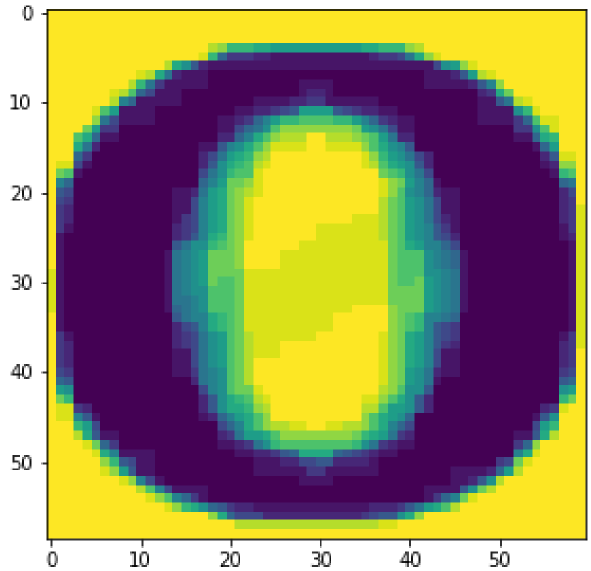
$$P(C=2 \mid y=3) = \frac{P(y=3 \mid C=2)\, P(C=2)}{P(y=3)} = \frac{1}{3}$$

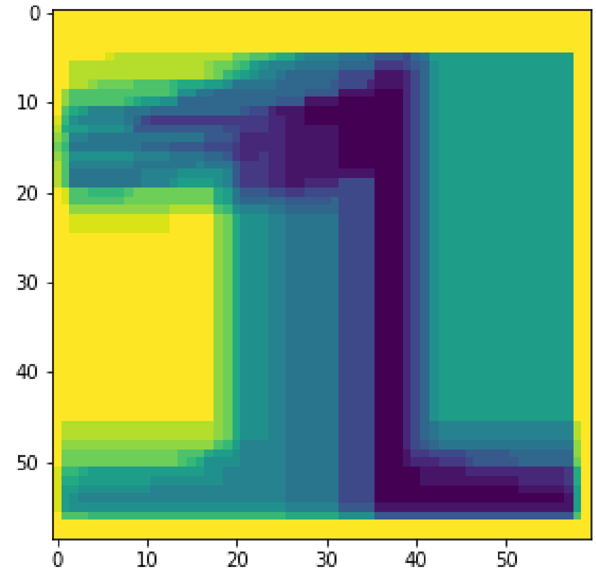
$$P(C=3 \mid y=3) = \frac{P(y=3 \mid C=3)\, P(C=3)}{P(y=3)} = \frac{1}{3}$$

۲)

قسمت نظرات:

در مورد قسمت الف)

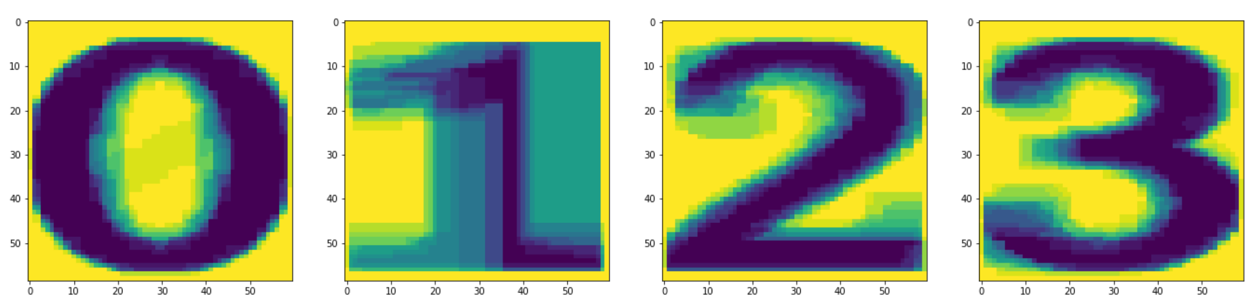




همانطوری که در شکل نیز مشخص است، برای مثال در مورد عدد ۰ ، با توجه به داده های آموزش بیشتر ۰ ها خطی در وسط خود نداشتند اما اکثرا گردی خاص یکسانی داشته اند که برای بعضی از آنها کلفت تر برای بعضی نازک تر بوده برای همین گردی آن در مورد شکل نماینده به صورت پررنگ تشکیل شده است. اما یکی از داده های آموزشی یک خط موربی در خود داشته که در شکل نماینده رنگ بسیار کمرنگی دارد. پس به طور کلی در شکل نماینده سعی شده است از تمام داده ها استفاده شده باشد. در مورد عدد ۱ نیز همین حالت را دارد.

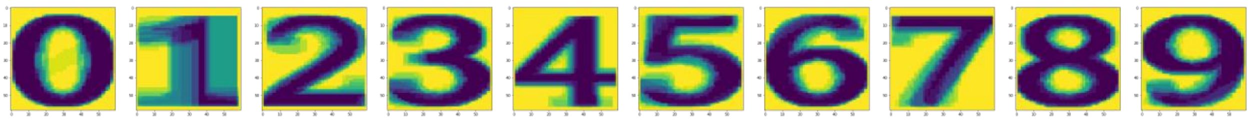
مقدار خطا ها و درست ها تماما در کل آورده شده است.

برای قسمت سوم:



همانند قسمت قبل که برای عدد ۰ بررسی کردیم ، در مورد باقی اعداد ۲ و ۳ نیز میبینیم که سعی شده از تمام داده های آموزشی استفاده شده و بخشی از آن در شکل نماینده آورده شده باشد.

در مورد قسمت چهارم:



همانند قسمت های قبلی ، در مورد این دسته داده ها نیست سعی شده قسمت هایی که بیشترین شباهت را دارند، دارای بیشترین پیکسل های رنگی بوده و بخش هایی که شباهت کمتری دارند در داده های آموزشی ، کمترین رنگ را دارند، برای مثال عدد ۸ که تقریبا تمام حالات آن شباهت زیادی به یکدیگر دارند ، پیکسل کمرنگ کمی دارد.

$\lambda_i = n_i x_i \lambda \qquad y_i \sim \text{Poisson}(\lambda_i)$

$\Rightarrow y_i \sim \text{Poisson}(n_i x_i \lambda)$ (A)

$\Rightarrow P(y_1, y_2, \dots, y_S \mid \lambda_i) = P(y_1, y_2, \dots, y_S \mid n_i x_i \lambda)$

$\Rightarrow \dfrac{P(n_i x_i \lambda \mid y_1, y_2, \dots)\, P(y_i)}{P(n_i x_i \lambda)} = \dfrac{P(n_1 x_1 \lambda \mid y_1) P(y_1) \times P(n_2 x_2 \lambda \mid y_2) P(y_2) \times \dots}{P(n_1 x_1 \lambda) P(n_2 x_2 \lambda) \dots}$

$\lambda_i = n_i x_i \lambda$

$MLE = \underset{\lambda}{\text{argmax}}\left(\prod n_i x_i \lambda\right) \xrightarrow{\ln} \sum_{i=1}^{S} n_i x_i \lambda = \lambda \sum_{i=1}^{S} n_i x_i$ (B)

---

unbias $\Rightarrow E(\hat{\lambda}) = \lambda$ (C)

var $\Rightarrow E\left[(\hat{\lambda} - E(\hat{\lambda}))^2\right]$ (D)

Error = var - bias E

۴)

(پیاده سازی شده است)

تابع likelihood , log_likelihood :

$$likelihood = \prod_{k=1}^{n} P(x|a,b) = a^n b^n \prod_{k=1}^{n} x_k^{a-1} (1-x_k)^{b-1}$$

$$Ln(likelihood) = n\log a + n\log b + \sum_{k=1}^{n} (a-1)\log x_k + \sum_{k=1}^{n} (b-1)\log(1-x_k)$$

۵)

(پیاده سازی شده است)

۶)

(پیاده سازی شده است)

۷)

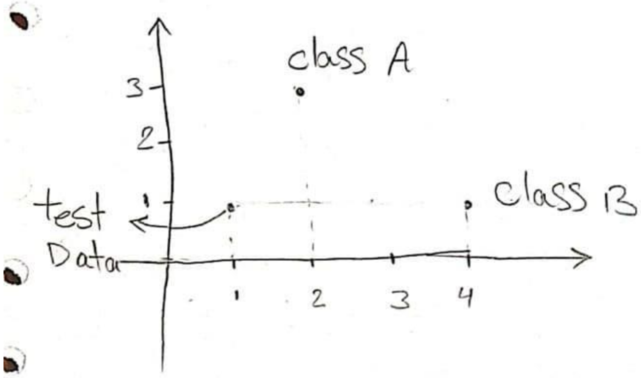
الف)

با توجه به دسته بند بیز ، بیز دقیقا خط وسط بین ۲ کلاس را جدا میکند و ۲ ناحیه برای خطای نوع ۱ و نوع ۲ را ایجاد میکند. هر چقدر این خط جا به جا شود مقدار این خطا ها بیشتر شده و بهترین حالت ، همان خطی است که ناحیه بند بیز ایجاد کرده است. برای هر دسته بند دیگری که دسته بندی میکند خطی بجز خط بیز ایجاد میکند خطایی بیشتر از خطای بیز ایجاد میکند.

ب)

امکان پذیر است. با در نظر گرفتن posteriori probability به روی پارامتر های مدل ، پیشبینی مدل از یک توزیع prior که از قانون بیز به دست می آید استفاده میکند.

پ)



مقایسه فاصله منهتن و اقلیدسی :

$$\text{اقلیدسی} = \begin{cases} \text{Class A} = \sqrt{5} \\ \text{Class B} = 3 \end{cases}$$

به کلاس A تعلق دارد

$$\text{منهتن} = \begin{cases} \text{Class A} = 3 \\ \text{Class B} = 3 \end{cases}$$

ممکن است به کلاس A یا B تعلق بگیرد.

ت)

in Bayes Classifier: $P(x|\omega_i)P(\omega_i) - P(x|\omega_j)P(\omega_j) \underset{\omega_j}{\overset{\omega_i}{\gtrless}}$

(d:۷

$\Rightarrow P(x|\omega_i) =$ توزیع داده‌های کلاس

بر اساس داده های آموزش پارامترهای توزیع و به طور کلی توزیع پیدا می‌شود و بر اساس این توزیع می‌توان کلاس بندی را تشکیل داد.

: MDC

در این مدل، با استفاده از داده های آموزش می توانیم نماینده هر کلاس را یافت و فاصله را تا آن نماینده اندازه گرفت. نماینده اگر میانگین باشد مجموع داده ها بر تعداد آن هاست.

ث)

با توجه به فرمول ۲ دسته بند، در MLE احتمال کلاس را برابر در نظر میگیرد و در تقسیم آن بی تاثیر میشود در حالی که در MAP احتمالات برابر نبوده و تاثیر احتمال هر کلاس در دسته بند مشخص است.

در مسائلی میتوان MLE را از MAP بهتر دانست که در آن مسائل ، احتمال کلاس ها مشخص نباشد یا اهمیتی نداشته باشد. در این موارد MLE بهتر از MAP است زیرا محاسبات زیادی برای به دست آوردن احتمال کلاس ها نیاز است که در بعضی مواقع اهمیت زیادی نیز ندارد پس MLE در این مسائل از MAP بهتر است.