**به نام خدا**

# درس یادگیری

# ماشین تمرین اول

حدیثه مصباح

810102253

# پاسخ I. سوالاتی راجب M-Class Classifier

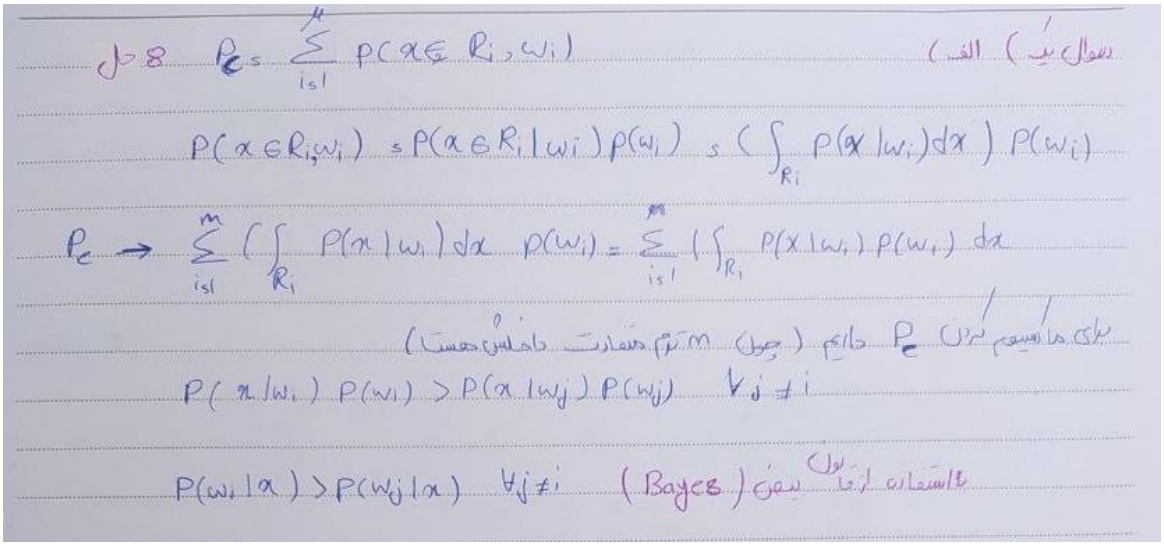
در یک مسئله طبقه‌بندی چند کلاسه:

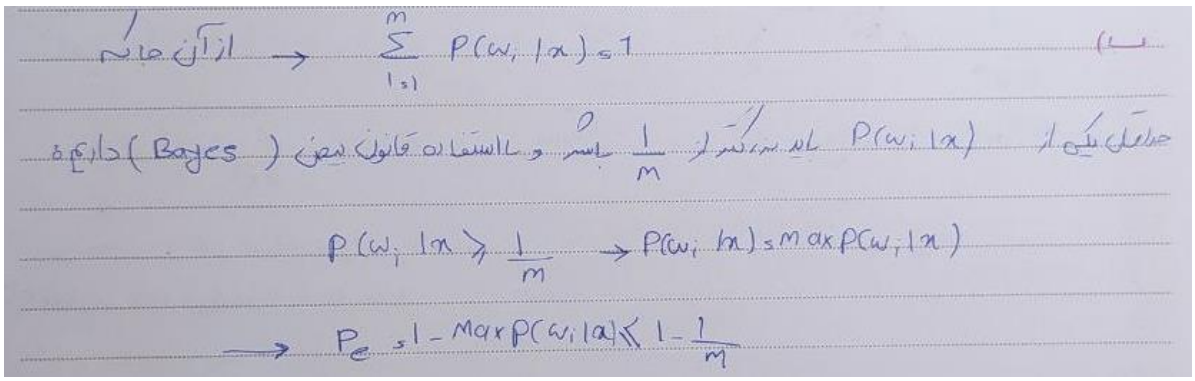
الف) نشان دهید که تصمیم‌گیری به کمک روش Bayes احتمال خطا را کمینه می‌کند.

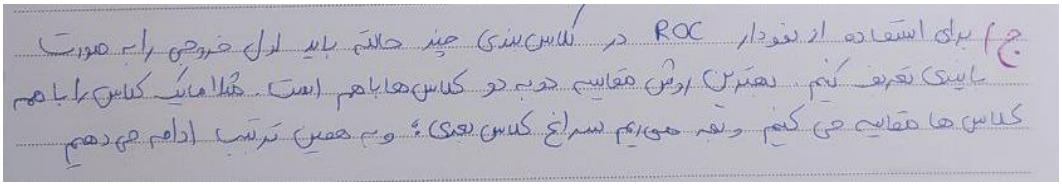
ب) ثابت کنید اگر M کلاس داشته باشیم، حد بالای خطا به صورت $p_e \leq \frac{M-1}{M}$ خواهد بود.

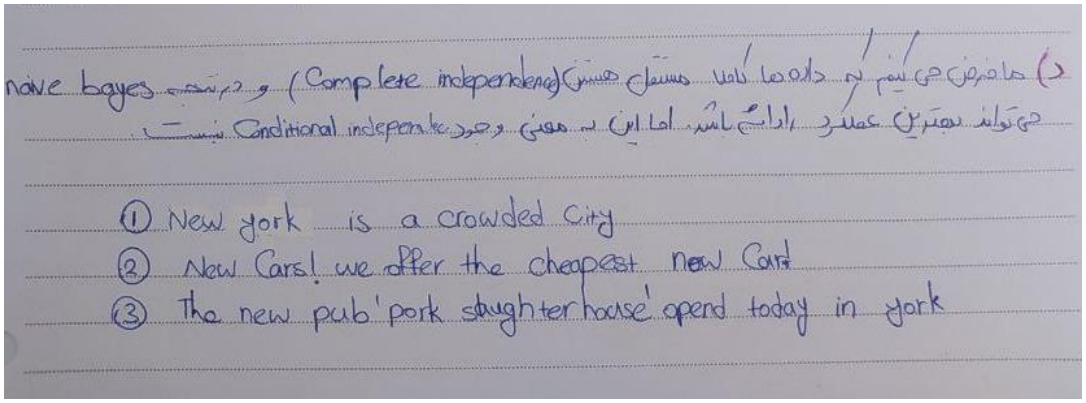
ج) راهی برای رسم نمودار ROC در حالت چند کلاسه پیشنهاد کنید.

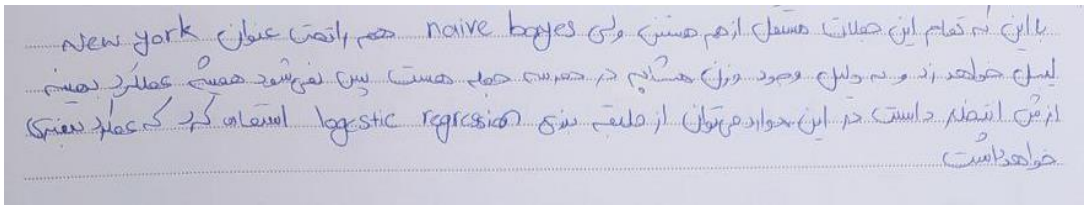
د) توضیح دهید که در چه مجموعه داده‌هایی naïve Bayes عملکرد بهینه خواهد داشت. علت را به تفصیل شرح دهید.







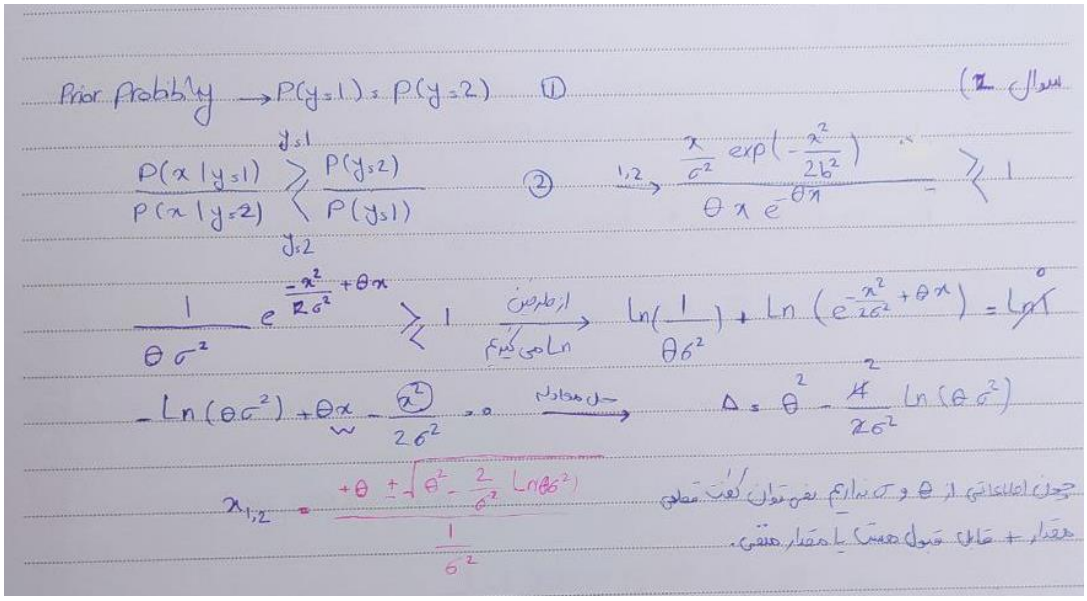


**پاسخ 2.** The corresponding area of the two classes in the Bayes class.

یک طبقه‌بند دو کلاسه با احتمال پیشین مساوی را در نظر بگیرید. فرض کنید داده‌های دو کلاس بر اساس توزیع‌های زیر تولید می‌شوند:

$$p(x|y=1) = \frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right) \quad x \geq 0$$
$$p(x|y=2) = \theta x exp(-\theta x) \quad x \leq 0$$

که $\sigma > 0$ و $\theta > 0$ پارامترهای مدل هستند. ناحیه مربوط به دو کلاس را در طبقه‌بند بیز به دست آورید.



## پاسخ 3. RISK MATRIX

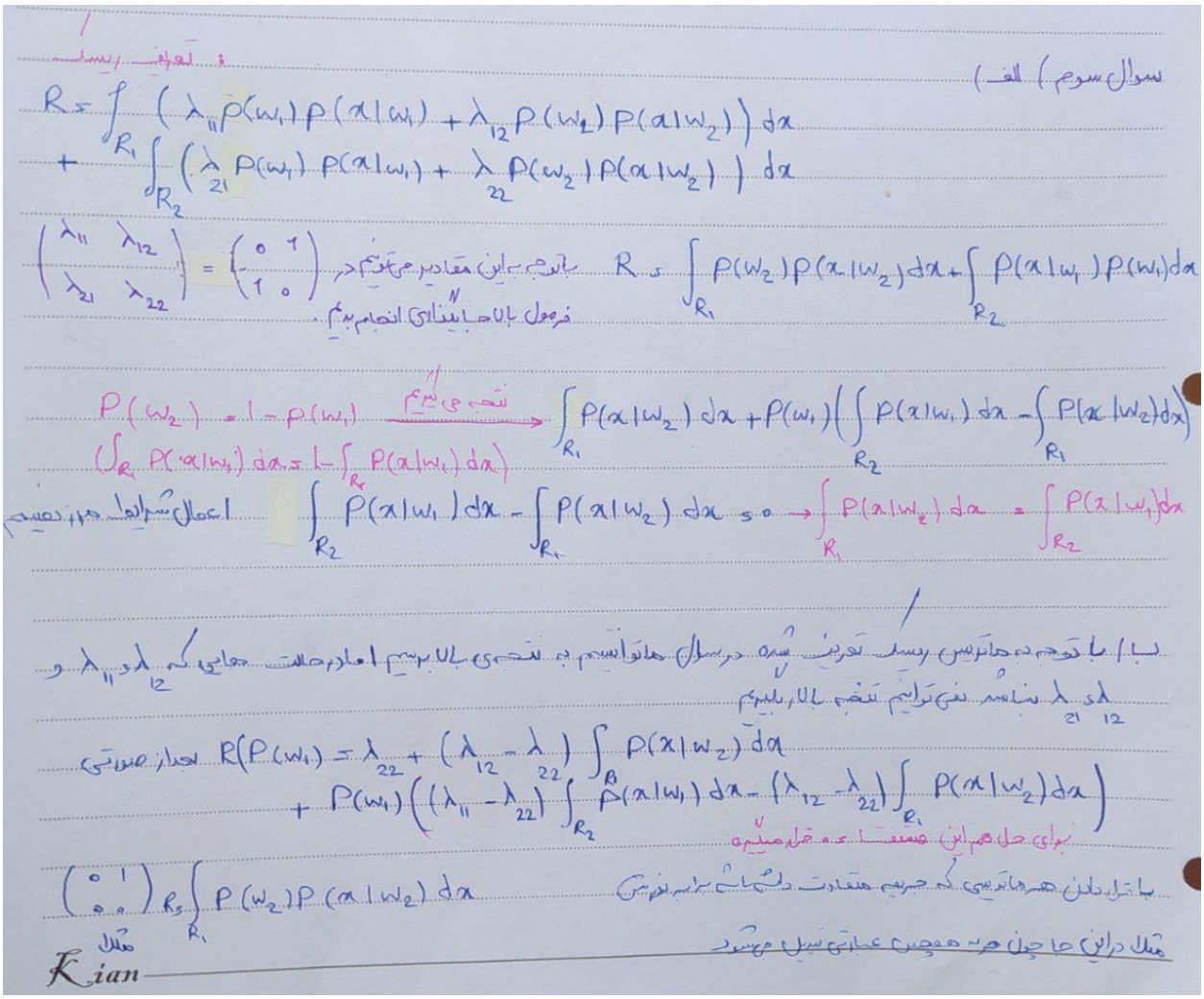
ماتریس ریسک زیر را در نظر بگیرید.

$$\begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

الف) نشان دهید با این شرایط مرز تصمیم شرط زیر را ارضا می‌کند.

$$\int_{R_2} p(x|\omega_1)\, dx = \int_{R_1} p(x|\omega_2)\, dx$$

ب)آیا این پاسخ همواره یکتاست؟ در غیر این صورت یک مثال نقض بزنید.



## پاسخ 4.$\mu$ Show the MAP estimate of the parameter $\mu$

متغیر تصادفی $x$ از توزیع $N(\mu, \sigma^2)$ است که $pdf$ مربوط به پارامتر $\mu$ به شکل زیر است:

$$p(\mu) = \frac{\mu \exp\left(-\mu^2 / 2\sigma_\mu^2\right)}{\sigma_\mu^2}$$

نشان دهید که تخمین $MAP$ پارامتر $\mu$ برابر است با:

$$\hat{\mu}_{MAP} = \frac{Z}{2R}\left(1 + \sqrt{1 + \frac{4R}{Z^2}}\right), \qquad Z = \frac{1}{\sigma^2}\sum_{k=1}^{N} x_k \qquad R = \frac{N}{\sigma^2} + \frac{1}{\sigma_\mu^2}$$

سوال چهارم)

$$P(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

نمونه تصادفی از توزیع $N(\mu, \sigma^2)$

* ماکسیمم پسین MAP مستقل از $P(x)$ چون

$$\mu_{MAP} = \arg\max_{\mu} P(\mu|x) \quad \frac{P(x|\mu)P(\mu)}{P(x)}$$

$$\rightarrow \arg\max_{\mu} P(x|\mu)P(\mu)$$

با استفاده از likelihood اندازه‌گیری تصادفی $x_i$ ها را خواهیم داشت:

$$l(\mu) = \left(\prod_{i=1}^{N} P(x_i|\mu,\sigma^2)\right) P(\mu)$$

log likelihood

$$L(\mu) = \ln(l(\mu)) = \ln(P(\mu)) + \sum_{k=1}^{N} \ln(p(x_k|\mu,\sigma^2))$$

$$\frac{\mu \exp\left(\frac{-\mu^2}{2\sigma_\mu^2}\right)}{\sigma_\mu^2}$$

محاسبه $\ln(P(\mu))$ و $\ln\mu - \ln(\sigma_\mu^2) - \frac{\mu^2}{2\sigma_\mu^2}$

محاسبه $\ln(p(x_k|\mu,\sigma^2)) = \sum_{k=1}^{N} \ln\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma}}\right) = N\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2}\sum_{k=1}^{N}(x_k-\mu)^2$

حال که عبارات بالا ساده شده حال برای پیدا کردن max پسین می‌توانیم از $\frac{dL(\mu)}{d\mu}$ استفاده کنیم

$$\frac{1}{\mu} - \frac{\mu}{\sigma_\mu^2} + \frac{1}{\sigma^2}\sum_{k=1}^{N}(x_k-\mu) = 0 \qquad -\frac{N\mu^2}{\sigma^2} - \frac{\mu^2}{\sigma\mu^2} + \frac{\mu}{\sigma^2}\sum_{k=1}^{N} x_k + 1 = 0$$

$$\sum_{k=1}^{N} x_k - \sum_{k=1}^{N} \mu \rightarrow N\mu$$

$$\mu^2\left(\frac{1}{\sigma_\mu^2} + \frac{N}{\sigma^2}\right) - \left(\frac{1}{\sigma^2}\sum_{k=1}^{N} x_k\right)\mu - 1 = 0$$

$R$ $\quad z$

$\Delta = b^2 - 4ac \rightarrow \Delta_r = z^2 + 4R \qquad x_1, x_2 \quad \frac{+z \pm \sqrt{z^2+4R}}{2R}$

طبق شرط $z$ $\quad z\left(1+\sqrt{1+\frac{4R}{z^2}}\right)$ ✓

$2R$

## پاسخ 5. Log likelihood Vs. Maximum A Posteriori

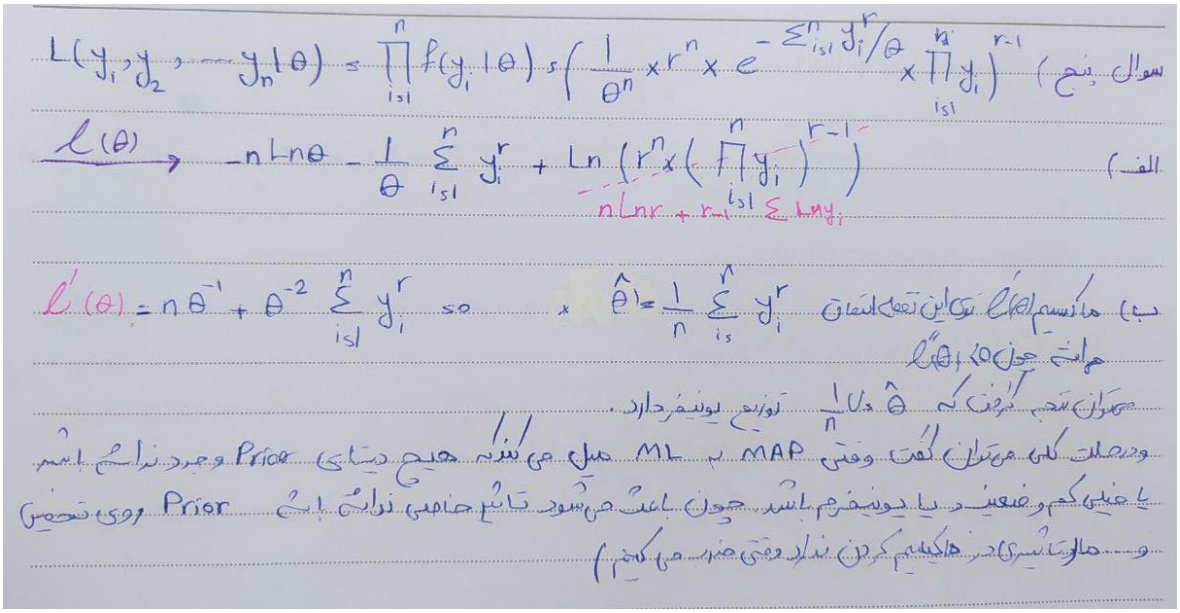
فرض کنید تابع چگالی احتمال متغیر تصادفی Y به شکل زیر است.

$$f_Y(y|\theta) = \begin{cases} \frac{1}{\theta} r y^{r-1} e^{-\frac{y^r}{\theta}}, & \theta > 0, y > 0 \\ 0, & elsewhere \end{cases}$$

که r یک ثابت مثبت است.

الف) تابع log likelihood را به دست بیاورید.

ب) توضیح دهید در چه حالتی و چرا تخمین‌گر MAP به ML میل می‌کند.



## پاسخ 6. Naïve bayes classification Vs. optimal classification

هدف از این سوال آشنایی و پیاده سازی طبقه‌بند naïve bayes است.

آ) در ابتدا در مورد طبقه‌بند naïve bayes توضیح دهید و تفاوت ساختاری آن‌را با یک طبقه‌بند بیزی بیان کنید. توضیح دهید که چرا به جای طبقه‌بند بیز از این طبقه‌بند استفاده می‌کنیم، هزینه‌ای که می‌دهیم چیست و در چه زمان‌هایی استفاده از این طبقه‌بند کاری منطقی است.

مجموعه داده Breast Cancer Wisconsin را از لینک زیر دانلود کنید.

https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data

توضیحات مربوط به این مجموعه داده به طور کامل در صفحه فوق وجود دارد؛ لطفا قبل از شروع به انجام تمرین توضیحات را مطالعه نمایید.

در ابتدا در صورت نیاز روی داده‌ها پیش‌پردازش انجام دهید.(هر پیش‌پردازشی که روی داده‌ها انجام می‌دهید را باید با ذکر دلیل توضیح دهید).

ب) این مجموعه داده شامل دو کلاس است. یک طبقه‌بند naïve bayes را از پایه و بدون استفاده از کتاب‌خانه پیاده‌سازی کنید. و طبقه‌بندی که طراحی کردید استفاده کنید. دقت، precision، Recall و ماتریس آشفتگی[1] را بررسی و تحلیل نمایید.

پ) مورد ب را به کمک کتاب‌خانه SKLEARN انجام دهید. نتایج دو بخش را مقایسه کنید.

**قسمت الف)**

Naïve bayes: **روشی برای دسته بندی است که از قضیه احتمالی بیز استفاده می‌کند که فرض می‌کند** feature **ها از هم مستقل هستند و هیچ ترتیب یا تعاملی میان آن‌ها وجود ندارد. این تبسیط، محاسبات را ساده‌تر می‌کند. و به طور کلی برچسبی را به داده اختصاص می‌دهد که احتمالش بیشتر باشد.**

$$y = \arg\max_{y} P(y) \prod_{i=1}^{n} P(x_i|y)$$

Optimal bayes: مثل Naïve bayes است فقط فرض مستقل بودن feature ها از هم را ندارد، و کواریانس بین آن‌ها را هم حساب می‌کند. پس نمی‌توان آن را مثل فرمول بالا نوشت. در این طبقه بند معمولاً از توزیع‌های احتمالی پیچیده‌تری برای مدل‌سازی استفاده می‌شود. این بدین معناست که طبقه‌بند بیزی می‌تواند تعامل‌ها و ترتیب‌های پیچیده‌تری میان ویژگی‌ها را مدل کند.

دلایل استفاده از Naïve bayes

سرعت و کارآیی: طبقه‌بند Naïve bayes به دلیل ساختار ساده‌تر و محاسبات سریع‌تر معمولاً در مسائلی که دارای تعداد زیادی ویژگی هستند، مانند پردازش متن و معنایی، بسیار مؤثر و سریع است.

مقاومت در برابر ابعاد بالا: در مسائل با فضای ویژگی‌های بزرگ، نیاز به تعداد زیادی داده آموزشی ندارد و با دیتای کم هم نتیجه قابل قبولی میدهد.

مثال هایی برای استفاده از Naïve bayes

طبقه‌بندی متن: معمولاً برای شناسایی ایمیل‌های هرزنامه، تجزیه و تحلیل احساسات و طبقه‌بندی اسناد استفاده می‌شود، جایی که ویژگی‌ها (کلمات) می‌توانند به‌عنوان مستقل مشروط در نظر گرفته شوند.

فضاهای ویژگی بزرگ: زمانی که یک فضای ویژگی با ابعاد بالا دارید و الگوریتمی می خواهید که بتواند آن را به طور موثر مدیریت کند.

نمونه سازی سریع: برای نمونه سازی سریع و به عنوان یک مدل پایه برای مقایسه با الگوریتم های پیچیده تر مفید است.

طبقه‌بند Naïve bayes فرض‌های ساده‌تری نسبت به واقعیت دارد و در مواردی که تعاملات پیچیده‌تری بین ویژگی‌ها وجود دارد، عملکرد ضعیف‌تری داشته دارد با این حال، محدودیت هایی وجود دارد و کم پیش میاد که داده های ما به طور کلی و تماما از هم مستقل باشند. در چنین مواردی، مدل‌های پیچیده‌تر مانند طبقه‌بند بیزی، درخت‌های تصمیم‌گیری یا شبکه‌های عصبی ممکن است مناسب‌تر باشند و با استفاده از Naïve bayes احتمال خطا بسیار بالا میرودو احتمال دسته بندی اشتباه وجود دارد .

ب)

Recall score: 0.9166666666666666

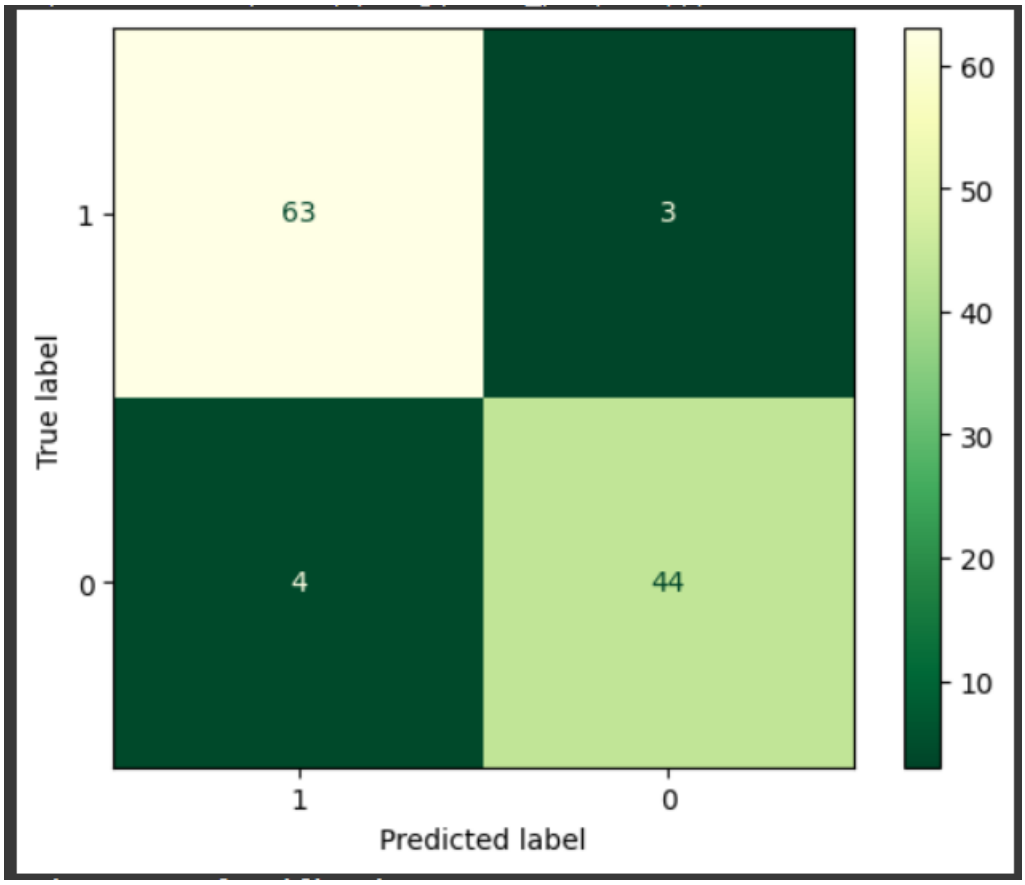
یعنی در93 درصد بیمار بودن به درستی تشخیص میدهد.

Precision score: 0.9361702127659575

یعنی در 93درصد مواقعی که طبقهبند ما تشخیص میدهد کیس بیمار است، درست گفته است.

Naive Bayes classification accuracy 0.9385964912280702

**یعنی در 92 درصد مواقع بیمار بودن یا نبودن به درستی تشخیص داده شده اند.**




**پ)**

(benign **برابر** 0 **و** malignant **برابر** 1 ).

Recall: 0.82758

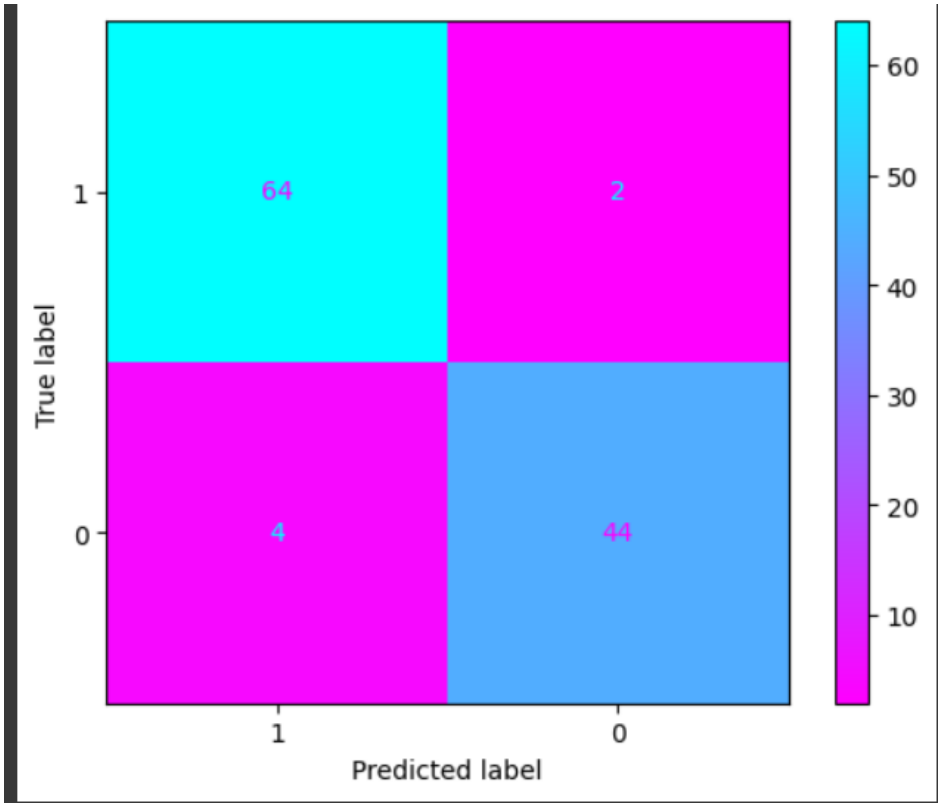
**یعنی در 82 درصد بیمار بودن به درستی تشخیص میدهد.**

Precision: 0.9795

**یعنی در 97 درصد مواقعی که طبقهبند ما تشخیص میدهد کیس بیمار است، درست گفته است.**

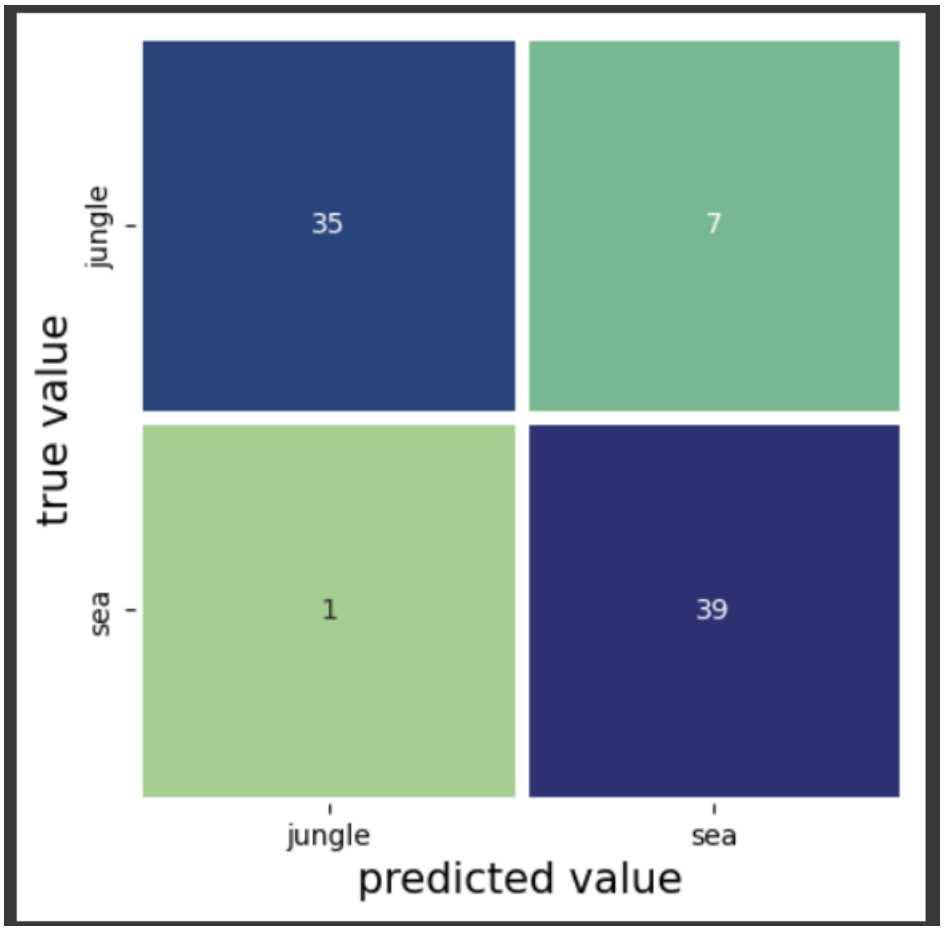
Accuracy: 0.92307

**یعنی در 92 درصد مواقع بیمار بودن یا نبودن به درستی تشخیص داده شده اند.**




همان‌طور که می‌انتظار داشتیم، نتایج به شدت به یکدیگر نزدیک هستند.(چه با استفاده از کتابخانه چه نوشتن از اول کد این طبقه بند ) این نکته نشان می‌دهد که پیاده‌سازی الگوریتم Naïve Bayes بدون استفاده از کتابخانه به درستی کار می‌کند. به لحاظ مهندسی نتیجه های ما در این طبقه بند بسیار خوب است؛ اما به لحاظ انسانی، این الگوریتم باید در تشخیص بیماری دقیق‌تر عمل کند. به عبارت دیگر، باید بیشتر به سمت تشخیص بیماری متمایل باشد تا این که بگوید سالم، و بعداً این تشخیص می‌تواند در آزمون‌های بعدی مثل نمونه برداری یا اسکن های پزشکی تأیید یا رد شود.(2 نمونه با این که باید سرطانی تشخیص داده میشدن و لیبل یک میگرفتن لیبل غیر سرطانی گرفته اند) بنابراین، دقت (Precision) در این مسئله باید ۱۰۰ درصد باشد.( 1 برابر malignant و 0 برابر benign)

## پاسخ 7. Implementation of binary classifier




Recall: 0.8333333333333334

**یعنی در 82 درصد بیمار بودن به درستی تشخیص میدهد.**

Precision: 0.9722222222222222

**یعنی در 97 درصد مواقعی که طبقهبند ما تشخیص میدهد کیس بیمار است، درست گفته است.**

Accuracy: 0.9024390243902439

**یعنی در 92 درصد مواقع بیمار بودن یا نبودن به درستی تشخیص داده شده اند.**

**پی نوشت : کد هر دو سوال اخر ضمیمه شده است.**