این پاسخها نمونه‌هایی هستند که حاوی نکات اصلی می‌باشند و لزوما تنها پاسخ‌های درست نیستند.

۱- پاسخ به این سوال به دیدگاه شما به یادگیری ماشین وابسته است و از این رو این سوال پاسخ یکتایی ندارد.

۲- با زیادشدن پیچیدگی فضای فرضیه مقدار خطای generalization نیز زیادتر می‌گردد و برای مدل نمودن چنین فضای فرضیه‌ای نیازمند داده‌های آموزشی بیشتری هستیم. در صورتی که تعداد داده‌های آموزشی کافی نداشته باشیم قادر به مدل کردن پیچیدگی فضای فرضیه نخواهیم شد و از آنجا که تعمیم خوبی نیز برای تخمین داده های جدید نخواهیم داشت از این رو خطای generalization زیادخواهد شد و دچار overfitting خواهیم شد. درصورتی که پیچیدگی فضای فرضیه زیاد نباشد قادر خواهیم بود با تعداد داده‌های آموزشی کم‌تر و با خطای generalization کمتری آن را مدل نماییم.

۳-

الف)

$$\Pr(S\,|\,W) = \frac{\Pr(S,W)}{\Pr(W)} = \frac{\Pr(W\,|\,S)\Pr(S)}{\Pr(W\,|\,S)\Pr(S) + \Pr(W\,|\,H)\Pr(S)}$$

ب) با درنظر گرفتن این فرض مشکلات زیر ایجاد می‌شود:

۱- در مدل ترتیب کلمات درنظر گرفته نمی‌شود.

۲- رابطه کلمات مجاور با هم مدل نمی‌شود.

۳-قواعد صرفی در مدل درنظر گرفته نمی‌شود.

۴-در حالت کلی کلمات و علائمی وجود دارد که بصورت متوازن توزیع نمی‌شوند و با این فرض آنها به درستی مدل نمی‌شوند.

ج) برای عبور از فیلترها می‌توان از تصویر متن استفاده نمود و یا عبارات را بصورت رمزی و با ترکیب حروف با برخی علائم وارد نمود مانند وارد نمودن C@s!o بجای Casio.

۴- اگرمتغییر تصادفی T را زمان لازم برای خروج از غار درنظر بگیریم. ما باید امید ریاضی زمان لازم برای خروج را بیابیم. مطابق با آنچه در مساله آمده است معادله زیر را می‌نویسیم:

$$E(T) = \frac{1}{3}(3) + \frac{1}{3}(1 + E(T)) + \frac{1}{3}(2 + E(T))$$
$$E(T) = 6$$

۵- با استفاده از روابط صورت مساله داریم:

$$L(\theta \mid x) = \prod_{i=1}^{N} P(x \mid \theta) = \frac{1}{(\pi\theta^2)^n}$$

از طرفی می‌دانیم این احتمال زمانی وجود دارد که شرط $\| x \| \le \theta$ برقرار باشد و در غیر اینصورت مقدار Likelihood صفر خواهد شد؛ بنابراین اگر q_{MAX} را بزرگترین مقداری در نظر بگیریم که در میان داده های X_i درنظر گرفته شده است. مقدار MLE برابر q_{MAX} می‌شود.

۶- الف) تعداد ویژگی ها و تعداد داده ها :

```
print (X.shape[۱]) # number of features

print (X.shape[۰]*X.shape[۱]) # number of data points
```

ب) هیستوگرام هر ویژگی

```
# Histogram of features
plt.show(plt.hist(X[:,0]))
plt.show(plt.hist(X[:,1]))
plt.show(plt.hist(X[:,2]))
plt.show(plt.hist(X[:,3]))
```

ج) میانگین هر ویژگی

```
# mean of features
print (np.mean(X[:,0]))
print (np.mean(X[:,1]))
print (np.mean(X[:,2]))
print (np.mean(X[:,3]))
```

د) مقادیر واریانس و انحراف معیار هر ویژگی

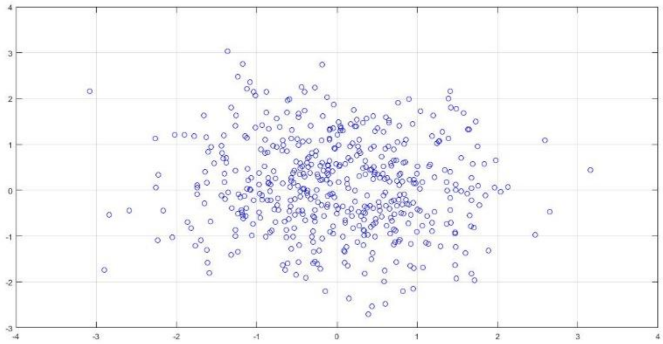
```
# variance of features
print (np.var(X[:,0]))
print (np.var(X[:,1]))
print (np.var(X[:,2]))
print (np.var(X[:,3]))
# std of features
print (np.std(X[:,0]))
print (np.std(X[:,1]))
print (np.std(X[:,2]))
print (np.std(X[:,3]))
```

ه) نرمال سازی داده ها

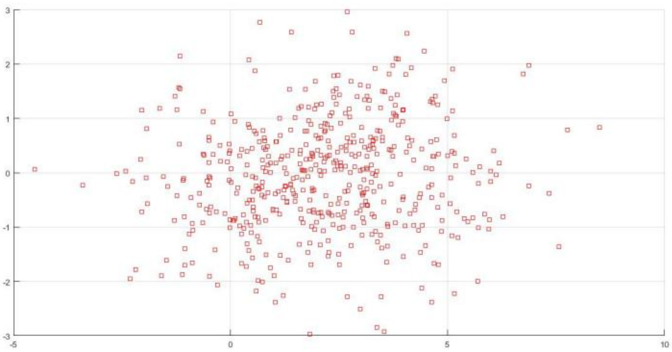
```
for feauture in range (0,4):
    X[:,feauture]= (X[:,feauture]-np.mean(X[:,feauture]))/np.std(X[:,feauture])
print(X)
```

-۷

الف)



شکل ۱ -پراکندگی کلاس اول



شکل 2 - پراکندگی داده های کلاس دوم

ب) این بخش را به صورت دستی انجام داده و سپس با متلب نشان می‌دهیم:

$$\text{Calss } C_1 : f_x(x|C_1) = \frac{1}{2\pi\sigma_1^{\,2}} \exp\left(-\frac{1}{2\sigma_1^{\,2}} \| x - \mu_1 \|^2\right)$$

$$\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \sigma_1^{\,2} = 1$$

$$\text{Calss } C_2 : f_x(x|C_2) = \frac{1}{2\pi\sigma_2^{\,2}} \exp\left(-\frac{1}{2\sigma_2^{\,2}} \| x - \mu_2 \|^2\right)$$

$$\mu_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \quad \sigma_2^{\,2} = 4$$

$$\text{threshold} = 1$$

$$\frac{f_x(x|C_2)}{f_x(x|C_1)} = \frac{\sigma_2^{\,2}}{\sigma_1^{\,2}} \exp\left(-\frac{1}{2\sigma_1^{\,2}} \| x - \mu_1 \|^2 + \frac{1}{2\sigma_2^{\,2}} \| x - \mu_2 \|^2\right) = \text{threshold} = 1$$

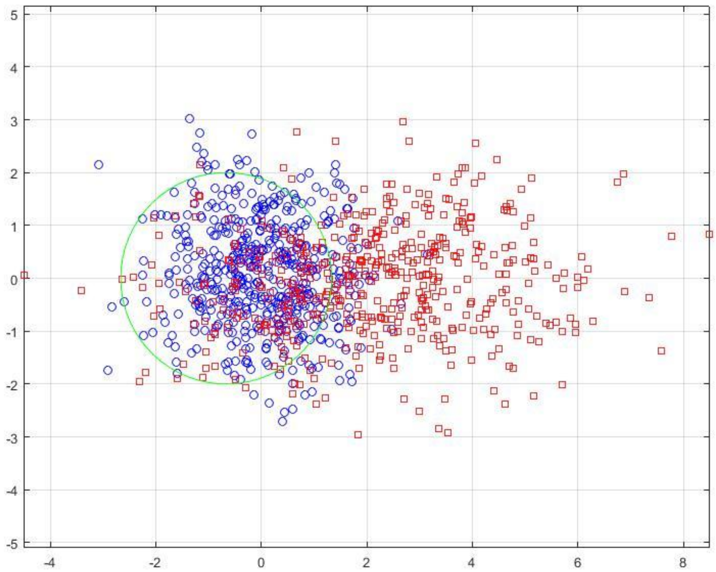
$$\xrightarrow{\log()} \frac{1}{\sigma_2^{\,2}} \| x - \mu_2 \|^2 - \frac{1}{\sigma_1^{\,2}} \| x - \mu_1 \|^2 = 4\log\left(\frac{\sigma_1}{\sigma_2}\right)$$

$$\xrightarrow{\text{straightforward manipulation}} \| x - x_c \| = r^2$$

$$x_c = \frac{\sigma_2^{\,2}\mu_1 - \sigma_1^{\,2}\mu_2}{\sigma_2^{\,2} - \sigma_1^{\,2}}$$

$$r^2 = \frac{\sigma_2^{\,2}\sigma_1^{\,2}}{\sigma_2^{\,2} - \sigma_1^{\,2}} \left[\frac{\| \mu_1 - \mu_2 \|^2}{\sigma_2^{\,2} - \sigma_1^{\,2}} + 4\log\left(\frac{\sigma_2}{\sigma_1}\right) \right]$$

$$x_c = \begin{bmatrix} \frac{-2}{3} \\ 0 \end{bmatrix} \quad r \approx 2.$$



شکل ۳- مرز جداساز دو کلاس

ج) اگر بصورت عددی داده هایی که به صورت درست دسته‌بندی شده‌اند را حساب نماییم. تقریبا ۸۱ درصد از داده ها درست دسته‌بندی شده‌اند.