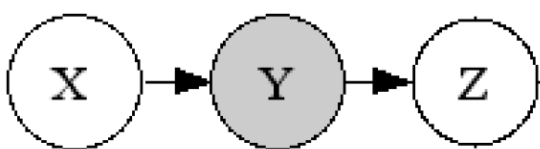
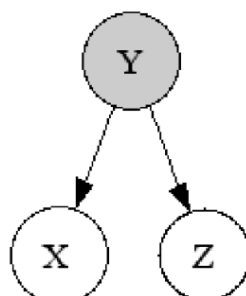


بسمه تعالی

بارم امتحان ۱۳۵ نمره است. با گرفتن ۱۰۰ نمره، نمره کامل امتحان را میتوانید کسب کنید. همچنین با جواب دادن به سوالات بیشتر حداکثر ۲۰ نمره امتیازی نیز می توانید کسب کنید.  
مدت زمان امتحان ۲ ساعت و ۴۵ دقیقه است.

ردیف	سوال
۱	<p>سوالات کوتاه پاسخ:</p> <p>الف) با ذکر یک مثال مفهوم bias-variance tradeoff را شرح دهید (۳ امتیاز)</p> <p>ب) رابطه بین پیچیدگی مسائل یادگیری ماشین و ابعاد ویژگی را شرح دهید (۳ امتیاز)</p> <p>ج) تفاوت مجموعه ارزیابی (Validation) و آموزش (Train) چیست. (۳ امتیاز)</p> <p>د) اگر بخواهید یک توصیف جامع از یادگیری ماشین داشته باشید، چگونه آن را توصیف می کنید. (۳ امتیاز)</p> <p>ه) خوب بودن یک ویژگی را به چه معیارهایی می توان سنجید (۳ امتیاز)</p> <p>و) آیا به کمک رگرسیون می توان ارتباطات غیر خطی را نیز مدل کرد؟ راهکارهای پیشنهادی خود را شرح دهید (۶ امتیاز)</p> <p>ز) مزیت ساختارهای بیزی جهت مدلسازی را نسبت به ساختارهای مبتنی بر درستنمایی بیشنه بیان کنید؟ (۳ امتیاز)</p> <p>ح) تفاوت تابع درستنمایی بیشنه و تابع توزیع احتمال چیست؟ (۳ امتیاز)</p> <p>ط) چگونه رویکرد Expectation Maximization چه مسائلی که داده کامل نیست (incomplete) می تواند مفید واقع شود؟ (۳ امتیاز)</p> <p>ی) کدام دسته از مدلسازی های یادگیری ماشین مستعد استفاده از رویکرد HMM است؟ (۳ امتیاز)</p> <p>ک) در مورد وابستگی و استقلال دو متغیر تصادفی <math>X</math> و <math>Z</math> در شبکه های بیزی در شکل الف و ب چه می توان گفت (۶ امتیاز)</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> <p>الف</p>  </div> <div style="text-align: center;"> <p>ب</p>  </div> </div> <p>ل) مولفه های ماتریس Confidence بیان کننده چه هستند؟ (۳ امتیاز)</p> <p>م) به صورت معمول پیچیدگی حافظه روش های پارامتریک بیشتر است یا ناپارامتریک؟ با ذکر مثال پاسخ خود را شرح دهید (۳ نمره)</p> <p>ن) آیا افزایش تعداد ویژگی ها منجر به افزایش عملکرد مدل های یادگیری ماشین خواهد شد ؟ با دلیل یا ذکر مثال توضیح دهید. (۵ نمره)</p>

۲	فرض کنید که نسترن و زهرا، امتحان یادگیری ماشین دارند. نسترن فقط مراجع ضمیمه اضافی را مطالعه می‌کند، و زهرا، تمام سوالات به همراه جواب آن‌ها را از کتاب و جزوه اصلی درس، حفظ می‌کند. به نظر شما کدام یک در امتحان موفق می‌شود؟ دلایل خود را برای موفقیت و یا عدم موفقیت هر کدام توضیح دهید. پیشنهاد خود را برای موفقیت یکی یا هر دو آنها ارائه دهید (۱۰ نمره).																		
۳	با در نظر گرفتن رابطه $y=\beta_0 x+\beta_1 x^2$ برای برازش نقطه‌های $(x_i,y_i), i=1,2,...,k+1$ ، مطلوب است. (۱۵) الف) با استفاده از روابط نویسی به صورت ماتریس، تخمین least square را از ضرایب رابطه به دست آورید. ب) یک رابطه برای ماتریس کواریانس حاصل از تخمین‌ها به دست آورید.																		
۴	اگر برای $x \in [0,3]$ توزیع‌های مربوط به دو کلاس و ماتریس $\lambda$ را به صورت زیر تعریف کرده باشیم، مطلوب است محاسبه بازه‌هایی که نواحی کلاس اول و دوم را توصیف می‌کنند. (۱۵ امتیاز) $p(x \omega_1) = x^2 - \frac{8}{3}$ $p(x \omega_2) = x - \frac{7}{6}$ $\lambda = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}$																		
۵	یک دسته بندی کننده دو کلاسه را در نظر بگیرید که برچسب صفر و یک ( $Y \in \{0,1\}$ ) را به نمونه‌هایی که متشکل از دو ویژگی باینری ( $X_1, X_2 \in \{0,1\}$ ) است، تخصیصی می‌دهد. چنانچه $X_1, X_2$ مشروط به $Y$ را مستقل فرض کنیم، با فرض یکسان بودن تابع پیشین ( $P(Y=0)=P(Y=1)=0.5$ ) احتمالات شرطی زیر را تعریف می‌کنیم. <table><tr><td><math>P(X_1 Y)</math></td><td><math>X_1 = 0</math></td><td><math>X_1 = 1</math></td></tr><tr><td><math>Y = 0</math></td><td>0.7</td><td>0.3</td></tr><tr><td><math>Y = 1</math></td><td>0.2</td><td>0.8</td></tr></table> <table><tr><td><math>P(X_2 Y)</math></td><td><math>X_2 = 0</math></td><td><math>X_2 = 1</math></td></tr><tr><td><math>Y = 0</math></td><td>0.9</td><td>0.1</td></tr><tr><td><math>Y = 1</math></td><td>0.5</td><td>0.5</td></tr></table> <p>متوسط نرخ خطا را به صورت زیر تعریف می‌کنیم (اگر <math>Y</math> برچسب حقیقی باشد، <math>\hat{Y}(X_1, X_2)</math> برچسب پیش بینی شده توسط طبق بندی کننده است )</p> $P_D \left( Y = 1 - \hat{Y}(X_1, X_2) \right) = \sum_{X_1=0}^1 \sum_{X_2=0}^1 P_D \left( X_1, X_2, Y = 1 - \hat{Y}(X_1, X_2) \right) .$ <p>توجه کنید که زیر نویس <math>D</math> بیان کننده آن است که احتمالات بر اساس توابع احتمال واقعی داده ها ( true distribution of data) محاسبه می شود.</p> <p>الف) پیش بینی دسته بندی کننده بیز (naïve Bayes classifier) را برای حالات زیر به دست آورید (۱۲ نمره)</p>	$P(X_1 Y)$	$X_1 = 0$	$X_1 = 1$	$Y = 0$	0.7	0.3	$Y = 1$	0.2	0.8	$P(X_2 Y)$	$X_2 = 0$	$X_2 = 1$	$Y = 0$	0.9	0.1	$Y = 1$	0.5	0.5
$P(X_1 Y)$	$X_1 = 0$	$X_1 = 1$																	
$Y = 0$	0.7	0.3																	
$Y = 1$	0.2	0.8																	
$P(X_2 Y)$	$X_2 = 0$	$X_2 = 1$																	
$Y = 0$	0.9	0.1																	
$Y = 1$	0.5	0.5																	

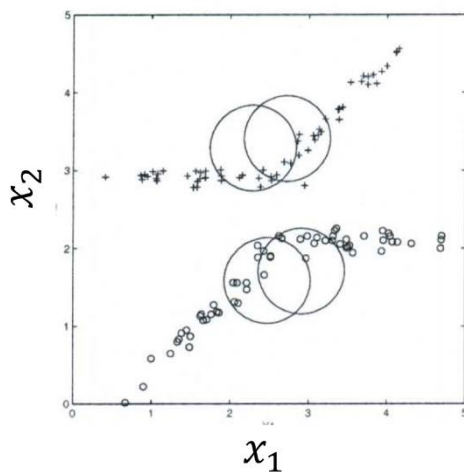
$X_1$	$X_2$	$P(X_1, X_2, Y = 0)$	$P(X_1, X_2, Y = 1)$	$\hat{Y}(X_1, X_2)$
0	0			
0	1			
1	0			
1	1			

ب) با فرض این که دسته بندی کننده بیز (naïve Bayes classifier) از بیشمار داده جهت یادگیری استفاده کرده است، متوسط خطای پیش بینی  $Y$  زمانی که هر دو ویژگی داده شده باشد ( مشروط بر هر دو ویژگی  $\{X_1, X_2\}$  ) را به دست آورید (۱۰ نمره)

ج) به نظر شما کدام یک از موارد زیر خطای کمتری دارد چرا (۸ نمره)

- پیش بینی دسته بندی کننده بیز (naïve Bayes classifier) برای  $Y$  زمانی که تنها  $X_1$  داده شده باشد
- پیش بینی دسته بندی کننده بیز (naïve Bayes classifier) برای  $Y$  زمان که تنها  $X_2$  داده شده باشد

۶ شکل زیر داده‌های مربوط به دو کلاس + و o را نشان می‌دهد. مدل Gaussian Mixture Model با دو تابع گوسی برای مدل کردن + ها و دو تابع گوسی برای مدل کردن o ها استفاده شده است. شرایط اولیه زیر را در نظر بگیرید.



خروجی دو گام اول الگوریتم  $EM$  را ترسیم کنید و در مورد همگرایی آن بحث کنید. (۱۵ نمره)

به راستی که دانش، مایه حیات دل‌ها، روشن کننده دیدگان کور و نیروبخش بدن‌های ناتوان است.  
امام علی (ع)