

۱- فرض کنید از متغیر تصادفی x نمونه‌های مستقل $\{x_1, x_2, \dots, x_N\}$ را داریم و از روی آن تابع چگالی احتمال را با روش پنجره Parzen

$$\hat{p}(x) = \frac{1}{N} \sum_{k=1}^N \frac{1}{h} \Phi\left(\frac{x-x_k}{h}\right)$$
 تخمین زده‌ایم:

در درس بایاس تخمین گر را محاسبه کردیم: $E\{\hat{p}(x)\} = \int_{y'} \frac{1}{h} \Phi\left(\frac{x-\lambda}{h}\right) p(\lambda) d\lambda$ واریانس تخمین گر را محاسبه کنید.

۲- با استفاده از ۱۰ داده از یک بردار ویژگی دو مولفه‌ای به صورت زیر، تابع چگالی احتمال در نقطه $(1,1)$ را با روش 3NN تخمین بزنید.
 $(2,2), (2,3), (3,2), (3,3), (4,4), (5,5), (6,6), (7,7), (8,8), (9,9)$

۳- در یک مسئله طبقه‌بندی دو کلاسی با یک ویژگی، داده‌های آموزشی زیر بدست آمده است:

$$\omega_1 : 1, 1.3, 1.5, 1.6, 1.8 \quad \omega_2 : 1.7, 1.9, 2, 2.2, 2.5$$

می‌خواهیم داده‌های تست 1.4, 1.65, 1.75, 1.85, 2.1 را طبقه‌بندی کنیم.

الف) داده‌های تست را با فاصله اقلیدسی طبقه‌بندی کنید.

ب) دو تابع چگالی احتمال $P(X|\omega_i)$ را با روش Parzen و با تابع پالس با پارامتر $h=1$ تخمین زده و رسم کنید و سپس داده‌های تست را با طبقه‌بندی کننده بیز با فرض هم‌احتمال بودن دو کلاس طبقه‌بندی کنید.

پ) با تخمین دو تابع چگالی احتمال $P(X|\omega_i)$ با روش kNN ($k=3$)، داده‌های تست را با طبقه‌بندی کننده بیز با فرض هم‌احتمال بودن دو کلاس طبقه‌بندی کنید. در صورت لزوم از فاصله اقلیدسی استفاده کنید.

ت) داده‌های تست را با طبقه‌بندی کننده kNN در دو حالت $k=1$ و $k=3$ طبقه‌بندی کنید. در صورت لزوم از فاصله اقلیدسی استفاده کنید.

۴- می‌خواهیم پارامترهای یک توزیع را که مدل GMM با دو مولفه دارد با الگوریتم EM تخمین بزنیم. داده‌های x_1, x_2, \dots, x_N به طور مستقل از این توزیع تولید شده‌اند. پارامترهای مدل به صورت $\Theta = (\alpha, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$ است و داریم:

$$P(z|\Theta) \sim \text{Ber}(\alpha)$$

$$P(x|z=j, \Theta) \sim \mathcal{N}(\mu_j, \sigma_j^2) \quad j=1,2$$

الف) تابع چگالی $p(x|\Theta)$ را بدست آورید.

ب) با فرض معلوم بودن پارامترهای مدل، مقدار $P(z_n = j|x_n, \Theta)$ $n=1,2,\dots,N$ را بدست آورید.

پ) تابع $Q(\Theta) = \sum_z p(Z|X, \Theta) \ln p(X, Z|\Theta)$ را بدست آورید.

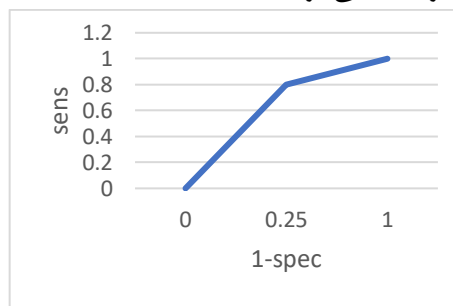
ث) با ماکزیمم کردن تابع مناسب، فرمولی برای پارامترهای مدل بدست آورید.

۵- در یک مسئله طبقه‌بندی دو کلاسی:

الف) اگر $Sens$ ۱۰۰٪ باشد، حدکثر و حداقل مقدار صحت چقدر است؟ مسئله را برای دو حالت دو کلاس هم‌احتمال و با احتمال متفاوت حل کنید.
 ب) اگر $Spec$ ۱۰۰٪ باشد، حدکثر و حداقل مقدار صحت چقدر است؟ (پ) اگر PP ۱۰۰٪ باشد، حدکثر و حداقل مقدار صحت چقدر است؟
 ت) آیا دو پارامتر از سه پارامتر فوق اگر ۱۰۰٪ باشد، مقدار صحت حتما ۱۰۰٪ است؟

۶- در یک مسئله طبقه بندی به دو کلاس بیمار (Positive) و سالم (Negative)، تقریبی از منحنی ROC به صورت زیر به دست آمده است. تعداد افراد در دو کلاس بیمار و سالم به ترتیب ۱۰۰ و ۱۲۰ نفر هستند. اگر ۱۰۰ نفر به تعداد افراد در کلاس سالم اضافه شود، منحنی ROC چه تغییری خواهد کرد؟ مقدار $AUC_{max} - AUC_{min}$ را حساب کنید.

راهنمایی: مانند شکل زیر منحنی ROC را در بهترین و بدترین حالت در سه نقطه حساب کرده و تقریب خطی بزنید.



۷- در یک مسئله ی طبقه بندی دو کلاسی، در مرحله آزمون داریم:

	P	N
P	37	8
N	3	52

الف) همه پارامترهای ارزیابی عملکرد طبقه بندی کننده را حساب کنید.
ب) احتمال پیشین کلاس ها را تخمین بزنید.

۸- در یک مساله طبقه بندی، یکبار با استفاده از ویژگی f1 و بار دیگر با استفاده از ویژگی f2 طبقه بند را آموزش داده ایم. ماتریس های در هم ریختگی متناظر با هر ویژگی در زیر نشان داده شده است. در صورت استفاده از هر دو ویژگی در آموزش مدل، در مورد ماتریس در هم ریختگی چه می توان گفت؟ مجموع TP+FN چقدر است؟ مجموع TN+FP چقدر است؟ صحت طبقه بندی در این حالت را به ساده ترین فرم بنویسید.

	P	N
P	70	1000
N	30	3900

f1

	P	N
P	65	900
N	35	4000

f2