۱. یک مدل طبقه‌بندی مولد برای k طبقه با پیشین‌های $p(C_k) = \pi_k$ و توزیع‌های شرطی $p(\emptyset|C_k)$ را در نظر بگیرید. به‌طوری که $\emptyset$ بردار ویژگی ورودی است. فرض کنید مجموعه داده $\{\emptyset_n, t_n\}$ را داریم که $n = 1, \dots, N$ بطوری که $\emptyset_n$ یک متغیر تک بعدی و t_n متغیر پاسخ است به‌طوری که اگر داده $\emptyset_n$ مختص کلاس j باشد، اندیس jام بردار t_n برابر ۱ است و دیگر اندیس‌ها برابر ۰ هستند. فرض کنید داده‌ها بصورت مستقل از مدل $\emptyset_n|C_k \sim Ln\ N(\mu_n, \sigma^2)$ تولید شده‌اند برای هر $n = 1, \dots, N$.

الف) تابع درستنمایی $p(T, \emptyset|\pi_1, \dots, \pi_k, \mu_1, \mu_2, \dots, \mu_k, \sigma^2)$ را بیابید. بطوری که $\emptyset = (\emptyset_1, \dots, \emptyset_n)$ و $T = (T_1, \dots, T_n)$.

ب) برآورد بیشینه درستنمایی از پارامترهای $\pi_1, \dots, \pi_k, \mu_1, \mu_2, \dots, \mu_k, \sigma^2$ را بدست آورید.

ج) داده جدید $\emptyset^*$ را در نظر بگیرید. توزیع پسین $p(C_j|\emptyset^*)$ را بصورت پارامتری بیابید و تابعی بنویسید که کلاس داده $\emptyset^*$ را مشخص کند (برای نوشتن تابع می‌توانید از عملگرهایی مانند max استفاده کنید)

راهنما : اگر $X \sim LnN(\mu, \sigma^2)$ آن‌گاه تابع چگالی این متغیر تصادفی بصورت زیر است

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$$

۲. یک مدل طبقه‌بندی مولد [1] برای k طبقه با پیشین‌های $p(C_k) = \pi_k$ و توزیع‌های شرطی $p(\emptyset|C_k)$ را در نظر بگیرید. به‌طوری که $\emptyset$ بردار ویژگی ورودی است. فرض کنید مجموعه داده $\{\emptyset_n, t_n\}$ را داریم که $n = 1, \dots, N$ بطوری که $\emptyset_n = (\emptyset_{n1}, \dots, \emptyset_{nD})$ و t_n بردار پایه پاسخ است یعنی اگر داده $\emptyset_n$ مختص کلاس j باشد، اندیس jام بردار t_n برابر ۱ است و دیگر اندیس‌ها برابر ۰ هستند. فرض کنید داده‌ها بصورت مستقل از این مدل تولید شده‌اند.

الف) مدل بیز ساده لوحانه را در نظر بگیرید (ویژگی‌ها مستقل از یکدیگرند)، در این حالت فرض کنید $\emptyset_{nd}|C_k \sim \exp(\lambda_k)$ برای هر $d = 1, \dots, D$ و $n = 1, \dots, N$. در ابتدا تابع درستنمایی $p(\emptyset, T|\lambda, \pi)$ را بدست آورید به‌طوری که $\lambda = (\lambda_1, \dots, \lambda_k)$، $\emptyset = (\emptyset_1, \dots, \emptyset_n)$، $\mathrm{T} = (t_1, \dots, t_n)$ و $\pi = (\pi_1, \dots, \pi_k)$. (داده‌ها و ویژگی‌های هر داده مستقل هستند).

[1] Generative Classification Model

ب) برآورد بیشینه درست‌نمایی از π_jها و λ_jها برای $j = 1, \dots, k$ را بیابید.

ج) داده جدید $\emptyset^*$ را در نظر بگیرید. توزیع پسین $p(C_j|\emptyset^*)$ را بصورت پارامتری بیابید و تابعی بنویسید که کلاس داده $\emptyset^*$ را مشخص کند (برای نوشتن تابع می‌توانید از عملگرهایی مانند max استفاده کنید)

راهنما : برای سوال ۱ و ۲ فرض بر این است که

$$p(T, \emptyset|\theta, \pi) = \prod_{n=1}^{N} p(t_n, \emptyset_n|\theta, \pi),$$

$$p(t_n, \emptyset_n|\theta) = p(t_n|\pi).p(\emptyset_n|t_n, \theta)$$

۳. با توجه به سوال ۲، فرض کنید به‌طور خاص ۳ متغیر ورودی بصورت زیر داریم

$$\emptyset_1 = (0.1, \quad 1, \quad 2), \quad t_1 = (1, 0)$$
$$\emptyset_2 = (0, \quad 1, \quad 5), \quad t_1 = (1, 0)$$
$$\emptyset_3 = (5, \quad 6, \quad 7), \quad t_1 = (0, 1)$$

الف) برآوردهای بیشینه درست‌نمایی از پارامترها در بخش ب سوال قبل را برای این مجموعه داده اعمال کنید.

ب) داده $\emptyset^* = (4, 5, 6)$ را در نظر بگیرید. پیش‌بینی کنید که این داده به کدام کلاس اختصاص دارد.

۴. فرض کنید مجموعه‌ای داده‌ای را از یک جامعه جمع‌آوری کرده‌ایم. هر داده شامل ۲ ویژگی به شرح زیر است

X_1 : میزان ساعت کاری در روز (می‌تواند مقادیر بین ۰ الی ۱۵ ساعت را دریافت کند)

X_2 : تعداد روزهای تعطیل در هفته (می‌تواند مقادیر بین . تا ۷ را دریافت کند)

همچنین هر داده شامل یک متغیر پاسخ y است که می‌تواند مقادیر ۰، ۱، ۲ داشته باشد.

y: میزان سلامت روحی (مقدار ۰ به معنای سلامت روحی پایین، ۱ به معنای سلامت روحی متوسط، ۲ به معنای سلامت روحی بالا است).

فرض کنید ما یک مدل $softmax\ regression$ را برازش کرده ایم. برای کلاس ۰ و ۱ پارامترهای زیر برآورد شده اند

$$\widehat{w}_0 = (\widehat{w}_{00} = -9.5, \widehat{w}_{01} = 1.1, \widehat{w}_{02} = -0.2),$$
$$\widehat{w}_1 = (\widehat{w}_{10} = -10, \widehat{w}_{11} = 1, \widehat{w}_{12} = 0.2)$$

$\widehat{w}_0$ وزن‌های تخمین زده شده برای کلاس ۰ و $\widehat{w}_1$ وزن‌های تخمین زده شده برای کلاس ۱ هستند.

فرض کنید فردی ۱۰ ساعت در روز کار می‌کند و ۲ روز در هفته تعطیل است. احتمال اینکه فرد سلامت روحی متوسط داشته باشد را تخمین بزنید. همچنین احتمال سلامت روحی پایین را تخمین بزنید.

۵. قصد داریم با استفاده از رگرسیون خطی، طبقه‌بندی k-کلاسه را پیاده سازی کنیم. بدین معنا که به‌ازای هر داده ورودی $x_n = [x_{n1}, \dots, x_{nd}]^T$ و $n = 1,2,\dots,N$ یک بردار پایه t_n (متغیر پاسخ) داریم که اندیس jام آن برابر ۱ است اگر داده x_n مختص کلاس j باشد و دیگر اندیس‌های مخالف j در این بردار برابر صفر هستند و بنابراین اندیس $j = 1, \dots, k$ نشان‌دهنده شماره کلاس یک داده است. فرض کنید بردار $w_j = [w_{j1}, \dots, w_{jd}]$ نیز ضرایب مدل خطی برای کلاس j باشند که باید برآورد شوند. فرض کنید هدف ما استفاده از پایه‌های دلخواه برای ورودی‌هاست. بنابراین قرار می‌دهیم $\emptyset(x_n) = [\emptyset(x_{n1}), \dots, \emptyset(x_{nd})]^T$

قرار دهید $\emptyset(\tilde{x}_n)^T = [1, \emptyset(\tilde{x}_n)^T]$ و $\widetilde{w}_j^T = [1, w_j^T]$ برای هر $n = 1, \dots, N,\ \ j = 1, \dots, k$. ماتریس $\emptyset(\widetilde{X})$ را به‌گونه‌ای می‌سازیم که سطر nام آن، $\emptyset(\tilde{x}_n)^T$ باشد و ماتریس $\widetilde{W}$ را به‌گونه‌ای می‌سازیم که ستون jام آن برابر با $\widetilde{w}_j$ باشد. همچنین ماتریس T را به‌گونه‌ای می‌سازیم که ستون nام آن، t_n باشد. تابع خطا را نسبت به ورودی و خروجی بصورت زیر تعریف می‌کنیم.

$$E_D(\widetilde{W}) = 0.5\ tr\left\{\left(\emptyset(\widetilde{X}).\widetilde{W} - T\right)^T.\left(\emptyset(\widetilde{X}).\widetilde{W} - T\right)\right\}$$

به‌طوری که عملگر . ضرب ماتریسی است.

الف) با توجه به تعریف تابع خطا، بصورت دقیق بیان‌کنید که این تابع بیان‌گر چیست؟ چرا باید از تابع tr استفاده کنیم؟

ب) گرادیان تابع خطای بالا را نسبت به $\widetilde{W}$ محاسبه کرده و بهترین برآورد از $\widetilde{W}$ را بدست آورید.

ج) (ماشین بردار پشتیبان[2]) فرض کنید قصد داریم یک طبقه‌بند دو کلاسه ایجاد کنیم بطوری که ابر صفحه‌ای را بیابیم بطوری که فاصله بین کلاس‌ها ماکزیمم شود. بطور مشابه مجموعه داده، به‌صورت بالا تعریف می‌شود. برای سادگی فرض کنید $t_i \in \{-1, +1\}$ نشان دهنده این است که داده x_i، عضو یکی از این دو کلاس است و $i = 1, \dots, N$. می‌دانیم در این روش، اگر $g(x_i)$ معادله ابر صفحه باشد، مقدار آن برابر صفر است اگر x_i نقطه‌ای برروی ابر صفحه باشد $g(x_i) > 0$، بدین معناست که علامت $g(x_i)$ مثبت است و داده x_i مختص کلاس $+1$ است و اگر منفی باشد، داده x_i، مختص کلاس -1 است. حال اگر t_i را علامت $g(x_i)$ در نظر بگیریم و $r(x_i)$ را طول بردار عمودی داده x_i برروی ابرصفحه در نظر بگیریم، مقدار $k_i = t_i . r(x_i)$ معیاری برای محاسبه فاصله عمودی یک نقطه از ابر صفحه است. به‌عنوان مثال اگر $k_i = 1$ باشد، بدین معناست که فاصله عمودی داده مربوطه از ابر صفحه برابر با ۱ است.

ج_۱) مقدار $r(x_i)$ را برای داده دلخواه x_i محاسبه کنید.

ج_۲) هدف مسئله این است که ابرصفحه‌ای را بیابیم تا فاصله بین دو کلاس ماکزیمم شود. برای این‌کار در ابتدا فرض کنید $k^* = min_{x_i}(t_i . r(x_i))$ بیان‌گر حداقل فاصله نقاط از ابر صفحه باشد. یعنی وجود دارد x^* بطوری که $k^* = t^* . r(x^*)$ حداقل مقدار را دارد که به آن بردار پشتیبان[3] نیز می‌گویند. مجددا فرض کنید $g(x^*) = w^T x^* + w_0$ معادله ابر صفحه باشد. در ابتدا دو طرف معادله را در یک اسکالر مانند h ضرب کنید.

_ نشان دهید که ابر صفحه $hg(x)$ برای تمام نقاط روی صفحه مقدار صفر خواهد گرفت و در نتیجه همان ابرصفحه است.

- فرض کنید برای داده x^* تعریف شده در بخش الف، قصد داریم مقدار h را طوری انتخاب کنیم که قدر مطلق فاصله بردار پشتیبان از ابر صفحه برابر ۱ باشد یعنی $h. t^* . (w^T x^* + w_0) = 1$ باشد. در این حالت h را بدست آورید. با مقیاس‌بندی مناسب می‌توان نتیجه گرفت $t^* . (w^T x^* + w_0) = 1$. در این حالت مقدار k^* را بدست آورید و سپس تابع هدفی بیابید تا بهترین مقدار از بردار w را بیابد تا چنین خواسته‌ای برآورده شود. (نیازی به حل تابع هدف نیست)

[2] Support Vector Machine
[3] Support Vector

۶. مدل رگرسیون پوآسنی زیر را در نظر بگیرید. فرض کنید مجموعه‌ی داده زیر را داریم

$$\mathrm{D} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

بطوری که به‌ازای $i = 1,2, \ldots, n$، y_i یک اسکالر و x_i یک بردار به طول d است. داریم :

$$y_i | x_i, \theta \sim Poisson(e^{{x_i}^T\theta})$$

و

$$p(y_1, y_2, \ldots, y_n | x_1, x_2, \ldots, x_n, \theta) = \prod_{i=1}^{n} p(y_i | x_i, \theta)$$

الف) در ابتدا نشان دهید که مدل توزیع پوآسن، از خانواده‌ی توزیع نمایی است.

ب) مقدار پارامتر θ را با ماکزیمم سازی تابع توأم بالا، برآورد کنید. (توجه کنید که نیاز به حل دقیق نیست و بدست آوردن معادله کافی است. در نهایت با روش‌های بهینه‌سازی مبتنی بر گرادیان، می‌توان پارامتر بیشینه کننده تابع درست‌نمایی را بصورت تقریبی برآورد کرد)