Subject:
Year. Month. Date.

احمدرضا فروردین
61030121

تمرین سری اول یادگیری ماشین

سؤال ۱:

الف)

Type: supervised learning , regression task

ب)

Type: supervised learning , classification

ج)

Type: supervised learning, Probabilistic classification

د)

Type: Unsupervised learning

Subject:

Year. Month. Date.

سؤال 2 :

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & i=j \text{ (correct ans.)} \\ \lambda_r & i=c+1 \text{ (rejection)} \\ \lambda_s & \text{o.w (misclassification)} \end{cases}$$

می‌دانیم که برای داده‌ی x، تابع ریسک $R(\alpha_i|x)$ به صورت زیر خواهد بود:

$$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|x)$$

حال فرض کنید x را به کلاس ω_i منتسب داده‌ایم؛ با توجه به اینکه برای همه‌ی $j \neq i$، $\lambda(\alpha_i|\omega_j)=\lambda_s$ و $\lambda(\alpha_i|\omega_i)=0$، $R(\alpha_i|x)$ را به صورت زیر می‌توان نوشت:

$$R(\alpha_i|x) = \sum_{j\neq i} \lambda_s P(\omega_j|x) = \lambda_s(1-P(\omega_i|x))$$

توجه شود که تساوی بالا از $\left(\sum_{j\neq i} P(\omega_j|x)\right) + P(\omega_i|x) = 1$ بدست می‌آید.

همچنین برای ریسک انتخاب α_{c+1} داریم:

$$R(\alpha_{c+1}|x) = \lambda_r$$

حال داده‌ی x را متعلق به کلاس ω_i می‌دانیم هرگاه:

① $P(\omega_i|x) \geq P(\omega_j|x) \quad \forall j \in C$

② $\lambda_s(1-P(\omega_i|x)) \leq \lambda_r$

که در ② منظور، کمتر بودن ریسک طبقه‌بندی به کلاس مورد نظر است نسبت به ریسک رد کردن.

Subject:

Year. Month. Date.

که با جابه‌جایی در دو طرف نامساوی به عبارت زیر می‌رسیم:

$$P(\omega_i|x) \geq 1 - \frac{\lambda_r}{\lambda_s}$$

و در غیر این صورت x را رد می‌کنیم اگر:

$$P(\omega_i|x) < 1 - \frac{\lambda_r}{\lambda_s}$$

بنابراین نتیجه‌ی نهایی برای تصمیم‌گیری برابر خواهد بود با:

$$\text{انتخاب } \alpha_i \text{ اگر این شرایط برقرار باشد} \begin{cases} P(\omega_i|x) \geq P(\omega_j|x) \quad \forall j \,, \\ P(\omega_i|x) \geq 1 - \frac{\lambda_r}{\lambda_s} \end{cases}$$

در غیر این صورت x را رد می‌کنیم.

سوال 3:

Classes: $y \in \{0,1\}$

Features: $X = [x_1, x_2]^T$, x_i i.i.d

Equal priors: $P(y=0) = P(y=1) = \frac{1}{2}$

در Bayse classifier، قاعده‌ی تصمیم به صورت زیر است:

$$P(y=0|X) \underset{1}{\overset{0}{\gtrless}} P(y=1|X) \Rightarrow \frac{P(y=0|X)}{P(y=1|X)} \underset{1}{\overset{0}{\gtrless}} 1$$

$$\Rightarrow \frac{P(X|y=0)\,P(y=0)}{P(X|y=1)\,P(y=1)} \underset{1}{\overset{0}{\gtrless}} 1 \xrightarrow{\quad} x_i \text{ i.i.d}:$$

$$\frac{P(x_1|y=0)\,P(x_2|y=0)\,P(y=0)}{P(x_1|y_1)\,P(x_2|y=1)\,P(y=1)} \underset{1}{\overset{0}{\gtrless}} 1 \xrightarrow{P(y=0)=P(y=1)}$$

$$2e^{-2x_1}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x_2-1)^2} \underset{1}{\overset{0}{\gtrless}} e^{-x_1}\frac{1}{\sqrt{\frac{\pi}{2}}}e^{-2(x_2-\frac{1}{2})^2}$$

از دو طرف ln می‌گیریم و ساده می‌کنیم

$$\Rightarrow -x_1 - \frac{1}{2}(x_2-1)^2 + 2(x_2-\frac{1}{2})^2 \underset{1}{\overset{0}{\gtrless}} 0$$

با باز کردن توان 2

$$\Rightarrow -x_2 + \frac{3}{2}x_2^2 - x_1 \underset{1}{\overset{0}{\gtrless}} 0$$

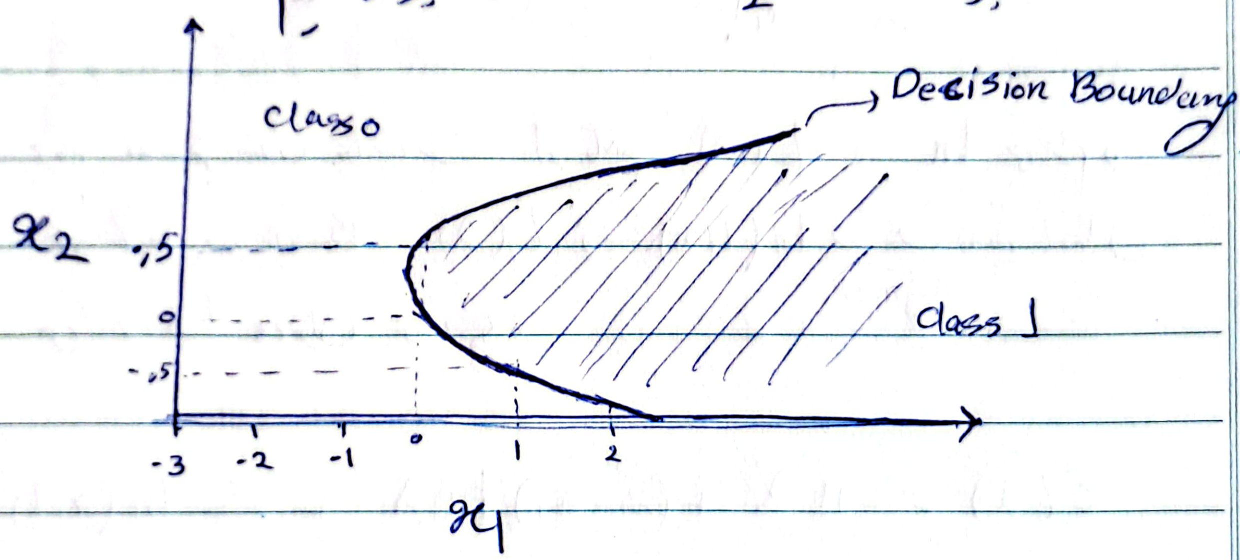
بنابراین $f(x_1,x_2)=\frac{3}{2}x_2^2-x_2-x_1$ تابع $f(x_1,x_2)>0$ آنگاه

$y=0$ (labeling) و در غیر این صورت برابر با $y=1$ (labeling).

بنابراین اگر $x_1>\frac{3}{2}x_2^2-x_2$ آنگاه $y=0$ و در غیر این صورت $y=1$ خواهد بود.

از آنجا که $f(x_1,x_2)$ به فرم معادله درجه 2 است لذا شکل نمودار آن هم سهمی خواهد بود.

برای $x_1=\frac{3}{2}x_2^2-x_1$ نمودار زیر را داریم:



ناحیه هاشور خورده، ناحیه class 1 و خارج از آن class 0 است.

مسئله ۴:

$$\begin{cases} x_1, x_2, \dots, x_n \sim \text{Poisson}(\lambda) \text{ , i.i.d.} \\ P(x_i \mid \lambda) = \dfrac{e^{-\lambda}\lambda^{x_i}}{x_i!} \end{cases}$$

$$\hat{\lambda}_{ML} = \arg\max_{\lambda} P(D \mid \lambda) = \arg\max_{\lambda} P(x_1, \dots, x_n \mid \lambda) = \arg\max_{\lambda} \prod_{i=1}^{n} P(x_i \mid \lambda)$$

الف) تابع likelihood را تشکیل می‌دهیم

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = e^{-n\lambda}\lambda^{\sum x_i}\left(\prod x_i!\right)^{-1}$$

برای ساده‌تر شدن محاسبات، از تابع $L(\lambda)$، ln می‌گیریم و از آنجا که بیشینه‌ی $L(\lambda)$ و $\ln(L(\lambda))$ در یک نقطه (argmax) اتفاق می‌افتد، لذا می‌توان نوشت:

$$\ell(\lambda) = -n\lambda + \left(\sum x_i\right)\ln\lambda - \sum \ln(x_i!)$$

برای ماکسیمم کردن تابع فوق، نسبت به λ مشتق می‌گیریم و برابر با صفر قرار می‌دهیم:

$$\frac{d\ell}{d\lambda} = -n + \frac{\sum x_i}{\lambda} = 0 \implies \hat{\lambda}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}$$

برای آنکه نشان دهیم $\hat{\lambda}_{MLE}$ نااریب است، باید نشان دهیم که

$$E[\hat{\lambda}_{MLE}] = \lambda$$

Subject:

Year. Month. Date.

داریم:

$$E[\hat{\lambda}_{MLE}] = E\left[\frac{1}{n}\sum x_i\right] = \frac{1}{n}\sum E[x_i]$$

$$= \frac{1}{n}(n\lambda) = \lambda$$

ب)

$$P(\lambda|D) \propto P(D|\lambda)\,P(\lambda) = e^{-n\lambda}\,\lambda^{\sum x_i}\,\lambda^{\alpha-1}\,e^{-\beta\lambda}$$

$$\Rightarrow P(\lambda|D) \propto \lambda^{\alpha+\sum x_i-1}\,e^{-(\beta+n)\lambda} \Rightarrow Gamma(\alpha+\sum x_i, \beta+n)$$

بنابراین توزیع Posterior، $Gamma(\alpha+\sum x_i, \beta+n)$ است.

ج)

$$\hat{\lambda}_{MAP} = \arg\max \underbrace{\ln P(\lambda|D)}_{L(\lambda)} \xrightarrow[\text{بالا}]{\text{با توجه به عبارت}}$$

$$\hat{\lambda}_{MAP} = \arg\max\,(\alpha+\sum x_i-1)\ln\lambda - (\beta+n)\lambda$$

با پیدا کردن argmax عبارت بالا یعنی نقطه $\frac{\alpha+\sum x_i-1}{\beta+n}$

برابر با $\hat{\lambda}_{MAP}$ خواهد بود:

$$\frac{dL}{d\lambda} = \frac{\alpha+\sum x_i-1}{\lambda} - (\beta+\lambda) = 0$$

$$\Rightarrow \hat{\lambda}_{MAP} = \frac{\alpha+\sum x_i-1}{\beta+n}$$

Subject:

Year. Month. Date.

د)

MLE : زمانی که تعداد داده‌ها زیاد است و اطلاعات ما برای Prior (توزیع اولیه برای پارامتر)، اطلاعات دقیق و موثقی نداشته یا مبهم باشد)

MAP : زمانی که تعداد داده‌ها کم است یا اطلاعات اولیه‌ی ما راجع به Prior، قابل اعتنا است (اطلاعات قوی راجع به Prior)

Subject:

Year. Month. Date.

سؤال 5 : $D_c = \{x_1, \dots, x_c\}$

$x' \in D_c$: nearest prototype to a test point x.

طبق قاعده‌ی Bayse. $P(\omega_m|x) = \max_i P(\omega_i|x)$ که ω_m برچسب انتخابی برای نمونه‌ی داده شده‌ی x است.

از طرفی برای محاسبه‌ی $P(e)$ که منظور احتمال خطا است، می‌توان نوشت:

$$P(e) = \int P(e|x)\, p(x)\, dx$$

قاعده‌ی Bayse برای تصمیم‌گیری، $P(e)$ را minimize می‌کند. یعنی در واقع طبق انتگرال بالا، $P(e|x)$ را به ازای هر x، به کمترین مقدار ممکن می‌رساند. فرض کنید $P^*(e|x)$، کمترین مقدار ممکن برای $P(e|x)$ و P^* کمترین مقدار ممکن برای $P(e)$ باشد. آنگاه:

$$P^*(e|x) = 1 - P(\omega_m|x)$$

$$P^* = \int P^*(e|x)\, p(x)\, dx$$

حال به برآورد کردن میانگین احتمال خطا برای روش nearest_neighbor می‌پردازیم.

به طور مشخص اگر $P_n(e)$ ، احتمال خطای نمونه با n مقدار باشد و

اگر $P = \lim_{n \to \infty} P_n(e)$ می‌خواهیم نشان دهیم که

$$P^* \leq P \leq P^* \left(2 - \frac{c}{c-1} P^*\right)$$

برای این کار در ابتدا به محاسبه‌ی $P(e|x)$ برای قاعده‌ی NN

می‌پردازیم؛ فرض کنید x' ، نزدیک‌ترین نقطه به عنصر داده شده‌ی x باشد:

$$P(e|x) = \int P(e, x'|x)\, dx' = \int P(e|x', x)\, P(x'|x)\, dx'$$

حال به بررسی $P(e|x', x)$ می‌پردازیم:

$$P(e|x, x') = P(\overbrace{\underset{\text{برچسب } x}{\omega} \neq \underset{\text{برچسب } x'}{\omega'}}^{e} \mid x, x')$$

$$= 1 - P(\omega = \omega' \mid x, x') \xlongequal{\omega \in \{1, \dots, c\}} 1 - \sum_{i=1}^{c} \left(P(\omega = i, \omega' = i \mid x, x')\right)$$

$$\xlongequal{i.i.d} 1 - \sum_{i=1}^{c} P(\omega_i|x)\, P(\omega'_i|x')$$

Subject:
Year. Month. Date.

از طرفی وقتی $n\to\infty$ به تابع دلتا میل می‌کند بنابراین:

$$\lim_{n\to\infty} P_n(e|x) = \int\left[1-\sum_{i=1}^{c} P(\omega_i|x)P(\omega_i|x')\right]\cdot\delta(x'_n - x)\,dx'_n$$

$$= 1-\sum_{i=1}^{c} P^2(\omega_i|x)$$

بنابراین با توجه به اینکه $P=\lim_{n\to\infty} P_n(e)$، داریم:

$$P=\lim_{n\to\infty} P_n(e)$$

$$=\lim_{n\to\infty}\int P_n(e|x)\,P(x)\,dx$$

$$=\int\left[1-\sum_{i=1}^{c} P^2(\omega_i|x)\right]P(x)\,dx \quad *$$

عبارت * نتیجه‌ی دقیق برای $P(e)$ در حالت استفاده از روش نزدیک‌ترین

همسایه است. می‌دانیم که در روش Bayse، کمترین خطا را داریم بنابراین

P^* که خطای Bayse است، یک حد پایین برای عبارت * خواهد بود.

حال به بررسی کران بالا برای * می‌پردازیم.

Ehsan

فرض کنید در طبقه‌بندی Bayse ، ω_i با برچسب با $i=m$ بیشترین احتمال باشد،

در واقع $P(\omega_m|x) > P(\omega_i|x)\ \forall i, i\neq m$

حال فرض کنید $P(\omega_m|x) \simeq 1$ و $i\neq m$ $0 \simeq P(\omega_i|x)$ ، در این

صورت $P(e|x) \simeq 1 - P^2(\omega_m|x)$ و از آنجا

که $P(e|x) = 1 - P^2(\omega_m|x) \le 2(\underbrace{1-P(\omega_m|x)}_{P^*(e|x)\ (\text{خطای بیز})})$

بنابراین می‌توان نتیجه گرفت که $P^* \le P \le 2P^*$.

برای حالت‌های کران دقیق‌تر داریم:

$$\sum_{i=1}^{c} P^2(\omega_i|x) = P^2(\omega_m|x) + \sum_{i\neq m} P^2(\omega_i|x)$$

و با توجه به اینکه:

$$P(\omega_i|x) \ge 0$$

$$\sum_{i\neq m} P(\omega_i|x) = 1 - P(\omega_m|x) = P^*(e|x)$$

می‌توان برای $P(\omega_i|x)$ مقادیر زیر را داشته باشیم:

$$P(\omega_i|x) = \begin{cases} \dfrac{P^*(e|x)}{c-1} & i\neq m \\ 1 - P^*(e|x) & i=m \end{cases}$$

Subject:

Year. Month. Date.

بنابراین نامساوی زیر حاصل می‌شود:

$$\sum_{i=1}^{c} P^2(\omega_i|x) \geq (1-P^*(e|x))^2 + (c-1)\frac{P^*(e|x)}{(c-1)^2}$$

که در نتیجه خواهیم داشت:

$$1-\sum_{i=1}^{c} P^2(\omega_i|x) \leq 2P^*(e|x) - \frac{c}{c-1}P^{*2}(e|x)$$

بنابراین با انتگرال گیری خواهیم داشت:

$$P^* \leq P \leq \int \left[2P^*(e|x) - \frac{c}{c-1}P^{*2}(e|x)\right] p(x)\, dx$$

$$\Rightarrow P^* \leq P \leq 2P^* - \frac{c}{c-1}P^{*2} = P^*\left(2-\frac{c}{c-1}P^*\right)$$

$$\Rightarrow P^* \leq P \leq P^*\left(2-\frac{c}{c-1}P^*\right)$$

Subject:

Year. Month. Date.

سوال 6:

* $p(x) = N(x,1)$

* $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$ (standard Gaussian kernel)

* KDE estimator based on N iid samples:

$$p_n(x) = \frac{1}{n} \sum_{i=1}^{N} \varphi\left(\frac{x - x_i}{h_n}\right)$$

با توجه به توضیحات و اطلاعات بالا، فضای واریانس و بایاس را برای

$p_n(x)$ محاسبه کرده و MISE را بدست می‌آوریم.

KDE توزیع درست $p(x)$ را با مشاهده‌ی $\{x_i\}$ با استفاده از کرنل φ

تقریب میزند:

$$p_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_n} \varphi\left(\frac{x - x_i}{h_n}\right)$$

که h_n پارامتری برای هموارسازی $p_n(x)$ است.

از طرفی برای MISE داریم:

$$MISE = E\left[\int \left(p_n(x) - p(x)\right)^2 dx\right]$$

که با ساده سازی و بسط دادن انتگرال به عبارت زیر میرسیم:

$$MISE = \int Bias^2\left(P_n(x_0)\right) dx + \int Var\left(P_n(x_0)\right) dx$$

حال به حساب Bias می‌پردازیم:

$$E\left(P_n(x_0) - P(x_0)\right) = E\left(\frac{1}{nh}\sum_{i=1}^{n} \ell\left(\frac{x_i - x_0}{h}\right)\right) - P(x_0)$$

$$= \frac{1}{h} E\left(\ell\left(\frac{x_i - x_0}{h}\right)\right) - P(x_0)$$

$$= \frac{1}{h}\int \ell\left(\frac{x - x_0}{h}\right) P(x)\, dx - P(x_0)$$

حال با تغییر متغیر $y = \frac{x - x_0}{h}$ و درنتیجه $dy = dx/h$ داریم:

$$E\left(P_n(x_0)\right) - P(x_0) = \int \ell\left(\frac{x - x_0}{h}\right) P(x) \frac{dx}{h} - P(x_0)$$

$$= \int \ell(y) P(x_0 + hy)\, dy - P(x_0)$$

(با توجه به اینکه $x = x_0 + hy$).

حال با استفاده از بسط تیلور برای h های کوچک داریم:

$$P(x_0 + hy) = P(x_0) - hy P'(x_0) + \frac{1}{2} h^2 y^2 P''(x_0) + O(h^2)$$

Subject:

Year. Month. Date.

بنابراین با جایگذاری داریم:

$$E\left(P_n(x_0) - P(x_0)\right) = \int \ell(y)\, p(x_0 - hy)\, dy - p(x_0)$$

$$= \int \ell(y) \left[p(x_0) + hy\,p'(x_0) + \frac{1}{2} h^2 y^2 p''(x_0) + o(h^2) \right] dy - p(x_0)$$

$$= \int \ell(y)\, p(x_0)\, dy + \int \ell(y)\, hy\, p'(x_0)\, dy + \int \ell(y)\, \frac{1}{2} h^2 y^2 p''(x_0)\, dy + o(h^2) - p(x_0)$$

$$= p(x_0) \underbrace{\int \ell(y)\, dy}_{=1} + h p'(x_0) \underbrace{\int y\, \ell(y)\, dy}_{=0} + \frac{1}{2} h^2 p''(x_0) \int y^2 \ell(y)\, dy + o(h^2) - p(x_0)$$

$$= p(x_0) + \frac{1}{2} h^2 p''(x_0) \int y^2 \ell(y)\, dy - p(x_0) + o(h^2)$$

$$= \frac{1}{2} h^2 p''(x_0) \int y^2 \ell(y)\, dy + o(h^2)$$

$$= \frac{1}{2} h^2 p''(x_0)\, \mu_K + o(h^2)$$

که $\mu_K = \int y^2 \ell(y)\, dy$. بنابراین بایاس برای KDE برابر است با:

$$\text{bias}\left(P_n(x_0)\right) = \frac{1}{2} h^2 p''(x_0)\, \mu_K + o(h^2)$$

Subject:

Year. Month. Date.

که برای $P_n(x) = N(x, \cdot)$ داریم:

$$\text{bias}(P_n(x_0)) = \frac{1}{2}h^2 p''(x_0) + o(h^2) \simeq \frac{1}{2}h^2 p''(x_0)$$

حال به محاسبه ی واریانس می پردازیم:

$$var(P_n(x_0)) = var\left(\frac{1}{nh}\sum_{i=1}^{n} \ell\left(\frac{x_i-x_0}{h}\right)\right)$$

$$= \frac{1}{nh^2} var\left(\ell\left(\frac{x_i-x_0}{h}\right)\right)$$

$$\leq \frac{1}{nh^2} E\left[\ell^2\left(\frac{x_i-x_0}{h}\right)\right]$$

$$= \frac{1}{nh^2}\int \ell^2\left(\frac{x-x_0}{h}\right) p(x)\,dx$$

$$= \frac{1}{nh}\int \ell^2(y)\, p(x_0+hy)\, dy \quad \left(\text{using } y = \frac{x-x_0}{h},\ dy = \frac{dx}{h}\right)$$

$$= \frac{1}{nh}\int \ell^2(y)\left[p(x_0) + hy\,p'(x_0) + o(h)\right] dy$$

$$= \frac{1}{nh}\left(p(x_0)\int \ell^2(y)\, dy + o(h)\right)$$

$$= \frac{1}{nh}\, p(x_0)\int \ell^2(y)\, dy + o\left(\frac{1}{nh}\right)$$

Ehsan

$$= \frac{1}{nh}\, p(x_0)\, \sigma_\ell^2 + o\left(\frac{1}{nh}\right)$$

که در آن $\sigma^2 = \int e^2(y)\, dy$ ، با توجه به اینکه برای

توزیع گوسی $\int e^2(y)\, dy = \frac{1}{2\sqrt{\pi}}$ داریم

$$Var(P_n(x_0)) = \frac{1}{nh} P(x_0) \frac{1}{2\sqrt{\pi}} + o\left(\frac{1}{nh}\right) \approx \frac{P(x_0)}{2\sqrt{\pi}\, nh}$$

حال برای محاسبه ی MISE داریم:

$$MISE = \int bias^2(P_n(x))\, dx + \int var(P_n(x))\, dx$$

$$\Rightarrow \begin{cases} \int bias^2(P_n(x))\, dx \approx \frac{h^4}{4} \int (P''(x))^2\, dx \\ \int var(P_n(x))\, dx \approx \frac{1}{nh} \frac{1}{2\sqrt{\pi}} \int P(x)\, dx = \frac{1}{2\sqrt{\pi}\, nh} \end{cases}$$

$$\Rightarrow MISE \approx \underbrace{\frac{h^4}{4} \int (P''(x))^2\, dx}_{bias^2} + \underbrace{\frac{1}{2\sqrt{\pi}\, nh}}_{Variance}$$

Subject:

Year. Month. Date.

حال برای بدست آوردن h_n بهینه داریم:

$$\frac{d}{dh}\left(c_1 h_n^4 + \frac{c_2}{n h_n}\right) = 0$$

که نتیجه برابر خواهد بود با

$$h_n^* \propto n^{-\frac{1}{5}}$$

سوال 7 :

1) با افزایش مقدار K در KNN :

* بایاس افزایش می‌یابد : مدل بیشتر به نمونه‌های اطراف خود (نمونه‌های در همسایگی خود) وابسته خواهد بود (که ممکن است مشابهت و تطابق کمتری با نقطه‌ی تست داشته باشند)، بنابراین مدل به الگوهای کمتری حساس خواهد بود، علی‌الخصوص در نزدیکی مرزبندی‌های کلاس. بنابراین امکان از دست دادن الگوهای واقعی و درست در data افزایش می‌یابد، در نتیجه مدل با در نظر گرفتن اکثریت در همسایگی، بایاس را افزایش می‌دهد.

* واریانس کاهش می‌یابد: با میانگین‌گیری روی تعداد همسایه‌های بیشتر، مرز کلاس‌ها و مدل، هموارتر خواهد بود، در نتیجه حساسیت به داده‌های با تعداد کمتر، کاهش یافته و نوسانات کوچک در مدل، کاهش می‌یابند، در نتیجه واریانس کاهش می‌یابد.

در نتیجه : $K\uparrow$ ← $bias\uparrow$ و $variance\downarrow$

2)

* مقیاس‌پذیری: فاصله‌ی Mahalanobis با نرمال‌سازی داده‌های با واحدهای مختلف (از طریق ماتریس کوواریانس داده‌ها) نسبت به فاصله‌ی اقلیدسی که باید داده‌ها با واحدهای مختلف را به صورت دستی به یک مقیاس برد، کارایی بهتری دارد و این کار را به صورت خودکار انجام می‌دهد.

Subject:

Year. Month. Date.

* در نظر گرفتن همبستگی بین ویژگی ها: فاصله‌ی Mahalanobis ، همبستگی بین داده‌ها را در محاسبه‌ی فاصله در نظر می‌گیرد و این باعث موثر بودن روابط بین ویژگی ها در محاسبه‌ی فاصله می‌شود؛ این ویژگی به خصوص زمانی که داده‌ها به هم ربط دارند مهم است، این در حالی است که فاصله‌ی اقلیدسی با همه‌ی داده‌ها به یک صورت برخورد می‌کند و همبستگی بین آنها را در نظر نمی‌گیرد.

3) استفاده از تابع کرنل تطبیق پذیر به جای تابع کرنل ثابت در موارد زیر ارزشمند است:

* Non-Uniform Data

Density : در مناطق با داده‌های پراکنده و متراکم، استفاده از کرنل های ثابت به هموارسازی بیش از حد منجر می‌شود و ممکن است جزئیات مهم را از بین ببرد. این در حالی است که در مناطق متراکم و با داده‌های کافی ممکن است نویزی بشود. کرنل های تطبیق پذیر این تنظیم کردن پهنا و هموارسازی را به صورت محلی (locally) انجام می‌دهند.

* وجود داده‌های پرت (outliers) : کرنل های تطبیق پذیر با افزایش محل Bandwidth داده‌های پرت مانع از تخریب تخمین چگالی می‌شود (برخلاف کرنل ثابت که ممکن است باعث این مشکل شود)

Subject:

Year. Month. Date.

(4

* نحوه‌ی کارکرد LSH که هدف آن کاهش پیچیدگی محاسباتی برای یافتن همسایه‌های نزدیک از طریق تبدیل و هش کردن نقاط مشابه در bucket های یکسان است به صورت مختصر چنین است:

نقاط مجموعه‌ی داده با hash function هایی که به گونه‌ای طراحی شده‌اند که نقاط مشابه را با احتمال بالا در یک hash bucket قرار می‌دهند، دسته‌بندی می‌شوند.

بنابراین هنگام جستجوی همسایه‌های نزدیک، تنها نقاط موجود در سطل‌های مشابه بررسی می‌شوند نه تمام داده‌ها.

* افزایش سرعت kNN با LSH: به جای مقایسه‌ی نقطه‌ی تست با تمام داده‌های آموزشی فقط با نقاط سطل خود مقایسه می‌شود.

در نتیجه پیچیدگی زمانی از $O(n)$ به $O(1)$ کاهش می‌یابد (در حالت ایده‌آل)

* مزیت LSH نسبت به K-d tree:

1) مقیاس‌پذیری در بعدهای بالا: K-d tree برای داده‌های با ابعاد بالا ($d>10$) کارایی خود را از دست می‌دهد در حالی که LSH همچنان مؤثر می‌باشد.

2) مدل کردن داده‌های در حال افزایش: K-d tree برای داده‌های پویا نیاز به تعادل سازی مجدد دارد در حالی که LSH انعطاف‌پذیر است.

Subject:

Year. Month. Date.

ج) جستجوی هوشمندانه‌تر: با استفاده از LSH hash codes در جستجوی داده‌ها نسبت به kd-tree بهتر عمل کرده (علی‌الخصوص در فضاهای با بعد بالا)