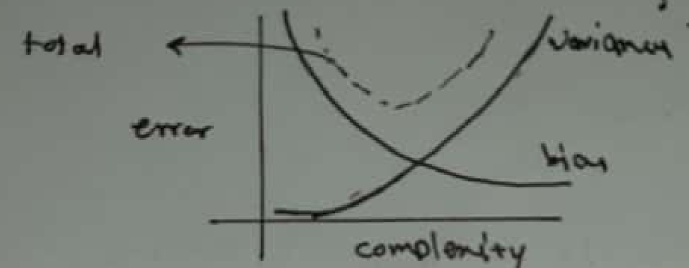
۱/۴ HW4 ۸۱۰۱۹۷۸۳۹ امیرحسین عباسکوهی

(۱

a) در روش KNN برای اینکه بتوان تعادل را بین bias و variance برقرار کرد و از طرفی هم به کمترین error برسیم (چه در train و چه در test). برای مثال در نمودار زیر:



(here we are talking about k)

در اینجا می‌بینیم که در حالتی که k عدد بزرگی باشد bias زیاد می‌شود و variance کم هست و در حالتی که k کوچک باشد واریانس زیاد و bias کم می‌باشد. پس باید k طوری مقداردهی شود که total را در نقطه minimum به دست آوریم.

در روش پنجره پارزن هم. اگر h خیلی بزرگ باشد bias خیلی زیاد است و واریانس خیلی کم و در h های کوچک برعکس این موضوع را داریم. باز هم h باید مقداری میانه انتخاب شود تا کمترین حالت total error خود شود.

b) در روش پارامتریک با استفاده از یک مدل (model based) به کمک فرضیات اولیه یک مدل مناسب برای تخمین مجموعه‌ای از پارامترها انتخاب می‌کنیم. در اینجا وابستگی به فرضیات اولیه زیاد است و در اکثر موارد فرضیات درست هستند پس خیلی دقت ما پایین می‌آید. در این روش‌ها سرعت بیشتر و نیاز به داده کمتر است. البته این روش‌ها flexible هم نیستند.

روش‌های non parametric فرض اولیه ندارند و سعی می‌کنند که دقت بیشتری حاصل شود. بسیار به تعداد داده وابسته هستند. flexible هستند اما interpretable نه.

2

c) در روش های kernel based به تعداد زیادی داده نیاز است. برای تخمین نیاز است

تمام sample ها ذخیره شوند همیشه آنها را داشته باشیم.

d) در KNN ، پارامتر V_n متغیر است، همان حجم می باشد:

$$P_n(x) = \frac{k_n/n}{V_n}$$

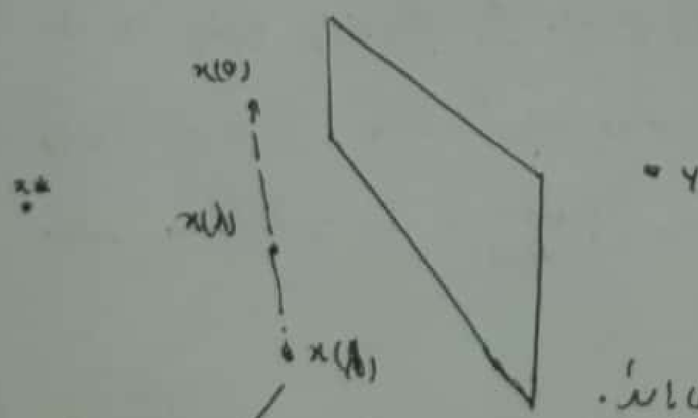
(در روش پارزن V_n حجم R_n ناحیه استفاده شده با n داده است)

$$P_n(x) = \frac{k_n/n}{V_n}$$

با اینکه V_n در هر دو رابطه هست اما مفهوم متفاوت است.

---

2)



فرض کنید x^* نمونه labled

و voronoi cell x^* باشد

و y هر نمونه labled دیگری باشد.

یک ابرصفحه unique فضا را به دو بخش که به y یا x^* نزدیک هستند طبقه بندی می کند.

دو نقطه دلخواه $x(0)$ و $x(1)$ را در voronoi cell مربوط به x^* در نظر بگیرید.

این دو نقطه قطعا در سمت فضای x^* هستند. حال خط واصل این دو نقطه را

در نظر بگیرید که می تواند براساس پارامتر $x(\lambda) = (1-\lambda)x(0) + \lambda x(1)$ ، $0 \le \lambda \le 1$ آن را

نمایش داد. از آنجایی که نیم فضای که توسط ابرصفحه تولید شده convex

است، تمام نقاط $x(\lambda)$ نزدیک به x^* خواهند بود (نه y)

3

این برای هر جفت نقطه $x(0)$, $x(l)$ صحیح است (در voronoi cell افراز)
همچنین نتیجه برای هر نمونه y_i ای درست است. پس $x(l)$ به x^* نزدیک‌تر
می‌ماند نسبت به هر نقطه labeled دیگری) $\longleftarrow$ براساس تعریف محدب
بودن پس voronoi cell محدب است.

(۳

a

$$\bar{P}_n(x) = \mathcal{E}[p_n(x)] = \frac{1}{nh_n}\sum_{i=1}^{n} \mathcal{E}\left[\varphi\left(\frac{x-x_i}{h_n}\right)\right]$$

$$= \frac{1}{h_n}\int_{-\infty}^{\infty} \varphi\left(\frac{x-v}{h_n}\right) p(v)\, dv$$

$$= \frac{1}{h_n}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-v}{h_n}\right)^2\right) \frac{1}{\sqrt{2\pi}\,\delta} \exp\left(-\frac{1}{2}\left(\frac{v-\mu}{\delta}\right)^2\right) dv$$

$$= \frac{1}{2\pi h_n \delta} \exp\left(-\frac{1}{2}\left(\frac{x^2}{h_n^2}+\frac{\mu^2}{\delta^2}\right)\right) \int_{\infty}^{\infty} \exp\left(-\frac{1}{2}v^2\left(\frac{1}{h_n^2}+\frac{1}{\delta^2}\right) - 2v\left(\frac{x}{h_n^2}+\frac{\mu}{\delta^2}\right)\right) dv$$

$$= \frac{1}{2\pi h_n \delta} \exp\left(-\frac{1}{2}\left(\frac{x^2}{h_n^2}+\frac{\mu^2}{\delta^2}\right)\right) + \frac{1}{2}\frac{\alpha^2}{\theta^2} \int_{\infty}^{\infty} \exp\left(-\frac{1}{2}\frac{(v-\alpha)^2}{\theta}\right) dv$$

$$\rightarrow \theta^2 = \frac{1}{1/h_n^2 + 1/\delta^2} = \frac{h_n^2\delta^2}{h_n^2+\delta^2}, \quad \alpha = \theta^2\left(\frac{x}{h_n^2}+\frac{\mu}{\delta^2}\right)$$

$$\xrightarrow{\text{integration}} \bar{p}_n(x) = \frac{\sqrt{2\pi}\cdot\theta}{2\pi h_n \delta} \exp\left(-\frac{1}{2}\left(\frac{x^2}{h_n^2}+\frac{\mu^2}{\delta^2}\right) + \frac{1}{2}\frac{\alpha^2}{\theta^2}\right)$$

$$= \frac{1}{\sqrt{2\pi}\, h_n \delta} \frac{h_n \delta}{\sqrt{h_n^2+\delta^2}} \exp\left(-\frac{1}{2}\left(\frac{x^2}{h_n^2}+\frac{\mu^2}{\delta^2} - \frac{\alpha^2}{\theta^2}\right)\right)$$

$\underline{a}$

$$\rightarrow \frac{x^2}{h_n^2} + \frac{\mu^2}{\delta^2} - \frac{\alpha^2}{\theta^2} = \frac{x^2}{h_n^2} + \frac{\mu^2}{\delta^2} - \frac{\theta^4}{\theta^2}\left(\frac{x}{h_n^2} + \frac{\mu}{\delta^2}\right)^2$$

$$= \frac{x^2 h_n^2}{(h_n^2+\delta^2)h_n^2} + \frac{\mu^2\delta^2}{(h_n^2+\delta^2)\delta^2} + \frac{2x\mu}{h_n^2+\delta^2}$$

$$= \frac{(x-\mu)^2}{h_n^2+\delta^2}$$

$$\longrightarrow \bar{p}_n(x) = \frac{1}{\sqrt{2\pi}\sqrt{h_n^2+\delta^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\delta^2}\right)$$

$$\longrightarrow \bar{p}_n(x) \sim N(\mu, h_n^2+\delta^2)$$

$\underline{b}$

$$p(x) - \bar{p}_n(x) = \frac{1}{\sqrt{2\pi}\,\delta}\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\delta^2}\right) - \frac{1}{\sqrt{2\pi}\sqrt{h_n^2+\delta^2}}\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\delta^2}\right)$$

$$= \frac{1}{\sqrt{2\pi}\,\delta}\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\delta^2}\right)\left\{1 - \frac{\delta}{\sqrt{h_n^2+\delta^2}}\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\delta^2} + \frac{1}{2}\frac{(x-\mu)^2}{\delta^2}\right)\right\}$$

$$= p(x)\left\{1 - \frac{1}{\sqrt{1+(h_n/\delta)^2}}\exp\left(-\frac{(x-\mu)^2}{2}\left\{\frac{1}{h_n^2+\delta^2} - \frac{1}{\delta^2}\right\}\right)\right\}$$

$$= p(x)\left\{1 - \frac{1}{\sqrt{1+(h_n/\delta)^2}}\exp\left(\frac{1}{2}\frac{h_n^2(x-\mu^2)}{h_n^2+\delta^2}\right)\right\}$$

$$\xrightarrow[\text{expand second order}]{\text{small } h_n} \frac{1}{\sqrt{1+(h_n/\delta)^2}} \simeq 1 - \frac{1}{2}(h_n/\delta)^2, \quad \exp\left(\frac{h_n^2}{2\delta^2}\frac{(x-\mu)^2}{h_n^2+\delta^2}\right) = 1 + \frac{h_n^2}{2\delta^2}\left(\frac{(x-\mu)^2}{h_n^2+\delta^2}\right)$$

$$\xrightarrow[\text{than } h_n^2]{\text{ignore higher}} p(x) - \bar{p}_n(x) = p(x)\left\{1 - \left(1-\frac{1}{2}\left(\frac{h_n}{\delta}\right)^2\right)\left(1+\frac{h_n^2}{2\delta^2}\frac{(x-\mu)^2}{h_n^2+\delta^2}\right)\right\}$$

$$= \frac{1}{2}\left(\frac{h_n}{\delta}\right)^2\left(1 - \frac{(x-\mu)^2}{h_n^2+\delta^2}\right)p(x) = \frac{1}{2}\left(\frac{h_n^2}{\delta}\right)\left[1 - \frac{(x-\mu)^2}{\delta}\right]p(x)$$

Scanned with CamScanner

5

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

$$= \frac{1}{n^2h_n^2}\sum_{i=1}^{n} Var\left[\varphi\left(\frac{x-x_i}{h_n}\right)\right]$$

$$= \frac{1}{nh_n^2} Var\left[\varphi\left(\frac{x-v}{h_n}\right)\right]$$

$$= \frac{1}{nh_n^2}\left\{ \mathcal{E}\left[\varphi^2\left(\frac{x-v}{h_n}\right)\right] - \left(\mathcal{E}\left[\varphi\left(\frac{x-v}{h_n}\right)\right]\right)^2 \right\}$$

$$\mathcal{E}\left[\varphi^2\left(\frac{x-v}{h_n}\right)\right] = \int \varphi^2\left(\frac{x-v}{h_n}\right) p(v)\, dv$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left[-\left(\frac{x-v}{h_n}\right)^2\right] \exp\left[\frac{-1}{2}\left(\frac{v-\mu}{\delta}\right)^2\right]\frac{1}{\sqrt{2\pi}\,\delta}\, dv$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-1}{2}\left(\frac{x-v}{h_n/\sqrt{2}}\right)^2\right] \exp\left[\frac{-1}{2}\left(\frac{v-\mu}{\delta}\right)^2\right]\frac{1}{\sqrt{2\pi}\,\delta}\, dv$$

$$= \frac{h_n/\sqrt{2}}{2\pi\sqrt{h_n^2/2+\delta^2}} \exp\left[\frac{-1}{2}\frac{(x-\mu)^2}{h_n^2/2+\delta^2}\right]$$

$$\frac{1}{nh_n^2}\mathcal{E}\left[\varphi^2\left(\frac{x-v}{h_n}\right)\right] = \frac{1}{nh_n 2\sqrt{\pi}}\frac{1}{\sqrt{2\pi}\,\delta} \exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\delta}\right)^2\right]$$

$$= \frac{1}{2n h_n\sqrt{\pi}} p(x)$$

$$\frac{1}{nh_n^2}\left\{\mathcal{E}\left[\varphi\left(\frac{x-v}{h_n}\right)\right]^2\right\} = \frac{1}{nh_n^2} h_n^2 \frac{1}{\sqrt{2\pi}\sqrt{h_n^2+\delta^2}} \exp\left[\frac{-1}{2}\frac{(x-\mu)^2}{h_n^2+\delta^2}\right]$$

$$\approx \frac{h_n}{nh_n}\frac{1}{\sqrt{2\pi}\,\delta}\exp\left[\frac{-1}{2}\left(\frac{(x-\mu)^2}{\delta^2}\right)\right] \approx 0$$

$$\Rightarrow Var[p_n(x)] \approx \frac{p(x)}{2nh_n\sqrt{\pi}}$$

۴) خواسته سوال رخ می دهد اگر وتنها اگر $D(a,b)$ یک metric باشد.

برای همین اثبات می کنیم آیا D ویژگی های متریک را دارد یا نه.

(1) $D(a,b) \geq 0$ ← چون sqrt داریم و در خود sqrt هم جمع مربع ها

(2) reflexivity ← $D(a,b) = 0$ if and only if $a = b$

به طور واضح $D = 0$ است اگر و تنها

اگر عضو Σ صفر شود پس ← $a = b$

همچنین اگر $a = b$ ← $D = 0$

(3) symmetry ← $D(a,b) = D(b,a)$ ← در بخش $(a_k - b_k)^2$ مى توانیم

$(b_k - a_k)^2$ بنویسیم تغییر در نتیجه قرار دهیم که ← symmetric داریم

(4) $D(a,b) + D(b,c) \geq D(a,c)$

$D(a,b) = \| A(a-b) \|_2$

$A = dig[\alpha_1 \dots \alpha_d]$

let, $x = a-c$, $y = c-b$ $\rightarrow \| Ax \|_2 + \| Ay \|_2 \geq \| A(x+y) \|_2$

$x' = Ax$, $y' = Ay$ $\rightarrow \| x' \|_2 + \| y' \|_2 \geq \| x' + y' \|_2$

$\xrightarrow[\text{both sides}]{\text{square}} (\| x' \|_2 + \| y' \|_2)^2 = \| x' \|_2^2 + \| y' \|_2^2 + 2 \| x' \|_2 \| y' \|_2 \geq \| x' \|_2^2 + \| y' \|_2^2 + 2 \sum_{i=1}^{d} x'_i y'_i$

$= \| x' + y' \|_2$

$\rightarrow \| x' \|_2 \| y' \|_2 \geq \sum_{i=1}^{d} x'_i y'_i$

7

$$\|x'\|_2 \|y'\|_2 \geq \left| \sum_{i=1}^{d} x_i y_i \right|$$

اگر با metric spaces در nearest neighbor کار کنیم، می‌توانیم اطمینان حاصل کنیم که

بهترین تخمین در این فضا وجود دارد. در نقاط دلخواه یک prototype به نام p

از p کوتاه‌ترین فواصل از یک input pattern به نام x^* داریم.

$$\delta = \min_{x} D(p, x)$$

که x عضوی از فضای metric است که compact است. حال دنباله‌ای

از اعداد $\{x_1, \ldots, x_n\}$ داریم که

$$D(p, x_n) \rightarrow \delta \quad \text{as} \quad n \rightarrow \infty$$

که $x_n \rightarrow x^*$ با توجه به compact بودن δ. براساس نابرابری مثلث:

$$\underbrace{D(p, x_n) + D(x_n, x^*)}_{n \to \infty:\ \delta} \geq D(p, x^*)$$

$$\Rightarrow \delta \geq D(p, x^*)$$

از طرفی هم $D(p, x^*) \geq \delta$ $\Leftarrow$ $D(p, x^*) = \delta$

یعنی با prototype های کافی می‌توان به p رسید که شبیه x^* است.

$$p \approx x^* \rightarrow p(\omega_i / p) \approx p(\omega_i / x^*)$$

Scanned with CamScanner

8

ه)

a

$P_Q(e)$ میانگین احتمال error براساس Q داده باشد

$P_Q(e) = \Pr[$ true category is ω_1 and ω_2 is most frequently labled $]$

$+ 2\Pr[$ true category is ω_2 and ω_1 is most frequently labled $]$

$= 2\Pr[$ true category is ω_1 and ω_2 is most frequently labled $]$

$= 2P(\omega_1)\Pr[$ label of ω_1 for fewer than $(k-1)/2$ points and rest $\omega_2]$

$= 2\frac{1}{2}\sum_{j=0}^{(k-1)/2} \Pr[$ j of Q chosen points are labled ω_1 rest $\omega_2]$

$= \sum_{j=0}^{(k-1)/2} \binom{Q}{j} \frac{1}{2^j} \frac{1}{2^{Q-j}} = \frac{1}{2^n} \sum_{j=0}^{(k-1)/2} \binom{Q}{j}$

b

$P_Q(e) = P_Q(e, k)$

$\rightarrow P_Q(e:1) = \frac{1}{2^n} < P_Q(e;k)$

$= \frac{1}{2^n} \sum_{j=0}^{(k-1)/2} \binom{Q}{j}$

> 0 for $k > 1$

Scanned with CamScanner

$$P_Q(e) = \frac{1}{2^n} \sum_{j=0}^{(k-1)/2} \binom{Q}{j} = Pr[\, B(Q, 1/2) \leq (k-1)/2 \,]$$

$$= Pr[\, Y_1 + \dots + Y_Q \leq \frac{k-1}{2} \,]$$

$Y_1 \dots Y_n$: independent

B : bionomial dist

$Pr[Y_i = 1] = Pr[Y_i = 0] = 1/2$

$k < \frac{Q}{\sqrt{Q}}$ اگر k با Q افزایش پیدا کند تا

$$P_Q(e) \leq Pr\left(Y_1 + \dots + Y_Q \leq \frac{Q/\sqrt{Q} - 1}{2} \right)$$

$$= Pr(Y_1 + \dots + Y_Q \leq 0)$$

for Q $\xrightarrow[\text{Very large}]{}$ 0

$\rightarrow \lim_{Q \to \infty} P_Q(e) = 0$