subject : date :

سوال ۱:

الف)

$B = \min\{\|w\|_2 : \forall i \in m, y_i \langle w, x_i \rangle \geq 1\}$

$R = \max_i \|x_i\|_2$

$\hat{w}$ بردار

$\hat{w}^T w^{(k)} = \hat{w}^T (w^{(k-1)} + y_i x_i) = \hat{w}^T w^{(k-1)} + \hat{w}^T y_i x_i$

می خواهیم این ضرب داخلی به هر شیوه ممکن با یه عدد بزرگی میشود به وسیله ما

ضمن زدن می دانیم که δ عددی بزرگتر از یک است.

$\hat{w}^T w^{(k)} \geq \hat{w}^T w^{(k-1)} + \delta \geq k\delta$

دوباره این جمله را بازگشتی کنیم و این زنجیره ادامه می دهیم

$w^T w^{(k)} \geq k\delta + w^T w^{(0)} \geq k$ ①

که k جمله ایجاد می شود

$\|w^{(k)}\|^2 = \|w^{(k-1)} + y_i x_i\|^2 \leq \|w^{(k-1)}\|^2 + \|y_i x_i\|^2$

چون ابتدا زنجیره را تا جمله پیشین عقب بردیم پس

$y_i x_i^T w^{(k-1)} \leq 0$ در نتیجه سمت راست نابرابری کوچکتر مساوی است

ضمن گفته سوال می دانیم $\|x_i\| \leq R$

$\|w^{(k)}\|^2 \leq kR^2 + w^T w^{(0)} = 0$ پس

حاشیه قبل این عبارت را تا جمله اول ادامه می دهیم که برابر $kR^2 + 0$ می شود

② $k \leq \langle \hat{w}, w^k \rangle = \|\hat{w}\| \|w^k\| \cos\theta \leq B \|w^k\| \cos\theta \leq RB\sqrt{k} \cos\theta$

$\cos\theta = \frac{\langle \hat{w}, w^k \rangle}{\|\hat{w}\| \|w^k\|} \leq 1$ ①,② $\geq \frac{k}{RB\sqrt{k}} = \frac{\sqrt{k}}{RB}$

$\frac{\sqrt{k}}{RB} \leq 1 \rightarrow \sqrt{k} \leq RB \rightarrow k \leq (RB)^2$

فرض آن بزرگتر صورتش کوچکتر است.

ب)

update rule $w_{t+1} = w_t + \eta y_i x_i$

ضمن حالت قبل:

$\hat{w}^T w^k = \hat{w}^T (w^{(k-1)} + \eta y_i x_i) = \hat{w}^T w^{k-1} + \hat{w}^T \eta y_i x_i \geq \hat{w}^T w^{(k-1)} + \eta\delta$

دوباره می دهیم تا جمله اول

$\geq k\eta\delta + w^T w^{(0)} \geq k\eta$ ①

$\|w^k\|^2 = \|w^{k-1} + \eta y_i x_i\|^2 \leq \|w^{(k-1)}\|^2 + \|\eta y_i x_i\|^2 \leq \|w^{k-1}\|^2 + \eta^2 R^2$

به همان دلیل قسمت قبل

Scanned with CamScanner

subject :

date :

$\|w^{k+1}\|^2 \le k\eta^2 R^2$ ← با حد بالا می‌رسیم

$\cos\theta = \frac{w^T w^{k+1}}{\|w^T\|\|w^{k+1}\|} \le 1$ ← حد بالا $\frac{w^T w^{k+1}}{\|w^T\|\|w^{k+1}\|} \ge \frac{k\eta\gamma}{B\sqrt{k}R\eta} \ge \frac{\sqrt{k}\gamma}{BR}$

$\Rightarrow k \le R^2B^2$ ← حداکثر دفعه‌ای که می‌تواند اشتباه کند

سوال ۲:

الف) فرض می‌کنیم اتفاقا یک نمونه در مرحله t یک واحد اشتباه کرده باشد.

$$h_{t+1} \Rightarrow \epsilon_{t+1} = \sum D_{t+1} = \frac{1}{Z_t}\sum D_t \exp(-\alpha_t y_i h_t(x_i))$$

$y_i h_t(x_i) = -1$ ← misclassified بودن

$$= \frac{1}{Z_t}\sum D_t \exp(\alpha_t) = \frac{e^{\alpha_t}}{Z_t}\sum D_t = \frac{e^{\alpha_t}\times\epsilon_t}{Z_t}$$

$$e^{\frac{1}{2}\log\frac{1-\epsilon_t}{\epsilon_t}} = \frac{\epsilon_t}{\sqrt{\epsilon_t(1-\epsilon_t)}}$$

$$= \sqrt{\frac{1-\epsilon_t}{\epsilon_t^2(1-\epsilon_t)}}\,\frac{\epsilon_t}{2} = \frac{1}{2}$$

طبق فرض ϵ_t از $\frac{1}{2}$ کمتر است یعنی این حالت رخ نمی‌دهد.

پس فرض خلف باطل است.

ب) حاصل ضرب دو عبارت را می‌نویسیم:

$y_i h_t(x_i) \cdot D_{t+1} = 0$

$\rightarrow y_i h_t(x_i) \cdot \frac{D_t \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$ ← بر اساس اینکه misclassified است یا نه

misclassified: $-1 \cdot \frac{D_t \exp(\alpha_t)}{Z_t}$

classified: $+1 \cdot \frac{D_t \exp(-\alpha_t)}{Z_t}$ ← اگر درست دسته‌بندی کرده باشد

$$= -\sqrt{\frac{\epsilon_t}{(1-\epsilon_t)}}\sqrt{\epsilon_t(1-\epsilon_t)} + \sqrt{\frac{1-\epsilon_t}{\epsilon_t}}(1-\epsilon_t) = 0$$

ضرب داخلی آنها صفر است.

سوال ۳: الف)

$m(x) = p(Y=1|X=x)$

$\hat{R}(h)$ یا misclassification است درنتیجه باید ابتدا داده‌ای تولید کنیم و یک بردار داشته باشیم. برای $h(x)$ یکی از دو حالت گفته شده باید رخ دهد یعنی یا $p(Y=0|X=x)$ از $p(Y=1|X=x)$ بیشتر شود یا برعکس.

$$\begin{cases} m(x) > 1-m(x) & h^*(x)=1 \\ m(x) \le 1-m(x) & h^*(x)=0 \end{cases}$$

$$\rightsquigarrow \begin{cases} m(x) > 1/2 & h^*(x)=1 \\ m(x) \le 1/2 & h^*(x)=0 \end{cases} \qquad \rightsquigarrow \begin{cases} m(x) > 1/2 & h^*(x)=1 \\ 0.w & h^*(x)=0 \end{cases}$$

ب)

$$R(\hat{h}) - R^* \le 2\int |\hat{m}(x) - m^*(x)|\, p_x(x)\,dx$$

Scanned with CamScanner

subject : _____ date : _____

اگر $\hat{m}(x)$ و $m^*(x)$ خروجی بزرگتر یا کوچکتر از $\frac{1}{2}$ باشند $h(x)$ برای داده‌ی با تابع اصلی متفاوت عمل می‌کند بنابراین باید $R(h)$ برای حالتی حساب کنیم که $h(x)$ هایی که از m ها حاصل می‌شود متفاوت باشند.

$$R(h) = \frac{1}{n} \sum I(h(X_i) \neq Y_i)$$

حتی حالت مثل می‌دانیم برای هر داده احتمال اینکه برای کلاس ۱ باشد $m^*(x)$ است پس حاصل $1 - m^*$ حاصل می‌شود. حالتی که در نظر گرفتیم $h(x)$ یک است و خطای تخمین ماکزیمم برابر $m^*(x)$ است چون فرض کردیم $h(x)$ است و حتی برای کلاس سوم نتیجه خطای حالت برای این هم برابر این حالت است.

این خطاست است قدر مطلق می‌گیریم. ← فرض کردیم $\hat{m}(x)$ کوچکتر از $\frac{1}{2}$ است

$$|1 - m^*(x) - m^*(x)| = |1 - 2m^*(x)| = 2|\frac{1}{2} - m^*(x)| \geq 2|\hat{m}(x) - m^*(x)|$$

$$|m^*(x) - 1 + m^*(x)| = |2m^*(x) - 1| = 2|m^*(x) - \frac{1}{2}| \geq 2|\hat{m}(x) - m^*(x)|$$

$m'(x)$ احتمال کلاس یک به شرط x است حال برای این‌که x درست آموزش خطای حاصل از تابع زیر استفاده می‌کنیم

$$\int 2|\hat{m}(x) - m^*(x)|(px)dx \geq R(\hat{h}) - R^*$$

سوال ۴.

برای ایجاد انتروپی ریشه به شرط فیچرهای مختلف را است می‌آوریم داولین فیچر را انتخاب می‌کنیم:

$$H = -\frac{3}{11}\left[\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}\right] + \frac{5}{11}\left[-\frac{3}{5}\log\frac{3}{5} - \frac{2}{5}\log\frac{2}{5}\right] + \frac{3}{11}\left[-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}\right] = 0.942$$

- مدیریت: نارامی ۲/۳، راحتی ۱/۳ ← ست شد
- اسیایی: نارامی ۳/۵، راحتی ۲/۵
- ایرانی: نارامی ۲/۳، راحتی ۱/۳

$$H = \frac{5}{11}\left(-\frac{2}{5}\log\frac{2}{5}\right) + \frac{4}{11}\left(-\frac{2}{4}\log\frac{2}{4} - \frac{2}{4}\log\frac{2}{4}\right) + \frac{2}{11}\left(-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right) - \frac{3}{5}\log\frac{3}{5} = 0.855$$

- ازان: نارامی ۲/۴، راحتی ۲/۴
- متوسط: نارامی ۲/۵، راحتی ۳/۵
- گران: نارامی ۲/۲، راحتی ۰/۲

$$H = \frac{4}{11}\left(-\frac{2}{4}\log\frac{2}{4} - \frac{2}{4}\log\frac{2}{4}\right) + \frac{3}{11}\left(-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}\right) + \frac{4}{11}\left(-\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4}\right) = 0.978$$

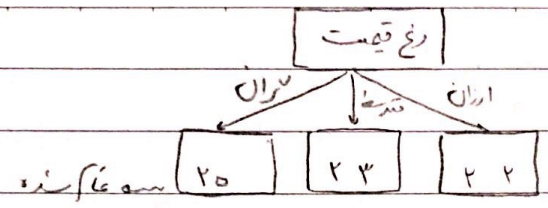
- شمال شهر: ۲/۴ ← ۲/۴
- غرب شهر: ۲/۳ ← ۱/۳
- شرق شهر: ۲/۴ ← ۲/۴

$$H = 0.978$$

- vegeterian: ۲/۴ ← ۲/۴
- gluten free: ۲/۴ ← ۱/۴
- None: ۲/۴ ← ۲/۴

کمترین انتروپی برای نوع مدیت است یعنی بیشترین IG پس براساس نوع مدیت تقسیم می‌کنیم.

Scanned with CamScanner



حالا کار را برای دسته‌ی ارزان ادامه می‌دهیم:

نوع غذا ← { فست فود: ناراضی ۱/۱ ، راضی ۰/۱ ؛ اسیایی: ۱/۲ ، ۱/۲ ؛ ایرانی: ۰/۱ ، ۱/۱ }

$$H = 0 + \frac{2}{4}\left(-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right) = 0.5$$

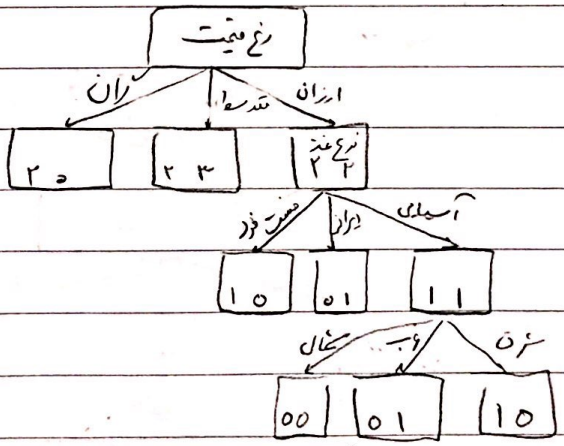
محل رستوران ← { شمال شهر: ۱/۱ ، ۰/۱ ؛ غرب شهر: ۱/۲ ، ۱/۲ ؛ شرق شهر: ۱/۱ ، ۰/۱ }

مثل حالت قبل ← $H = 0.5$

محدودیت‌ها ← { vegeterian: ۱/۲ ، ۱/۲ ؛ gluten free: ۱/۱ ، ۰ ؛ none: ۰ ، ۱/۱ }

مثل حالت قبل ← $H = 0.5$

در این حالت فرقی ندارد بر چه اساسی مرتب کنیم پس نوع غذا را انتخاب می‌کنیم پس از آن هم تفاوتی نمی‌کند پس محل را انتخاب می‌کنیم



حالا سراغ متوسط می‌رویم و آن را بررسی می‌کنیم

نوع غذا ← { فست فود: ۰/۱ ، ۱/۱ ؛ اسیایی: ۱/۲ ، ۱/۲ ؛ ایرانی: ۱/۲ ، ۱/۲ }

$$H = 0 + \frac{2}{5}\left(-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right) + \frac{2}{5}\left(-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right) = 0.8$$

محل رستوران ← { شمال: ۲/۳ ، ۱/۳ ؛ غرب: ۰ ، ۰ ؛ شرق: ۰ ، ۲/۲ }

$$H = \frac{3}{5}\left(-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}\right) + 0 + 0 = 0.55$$

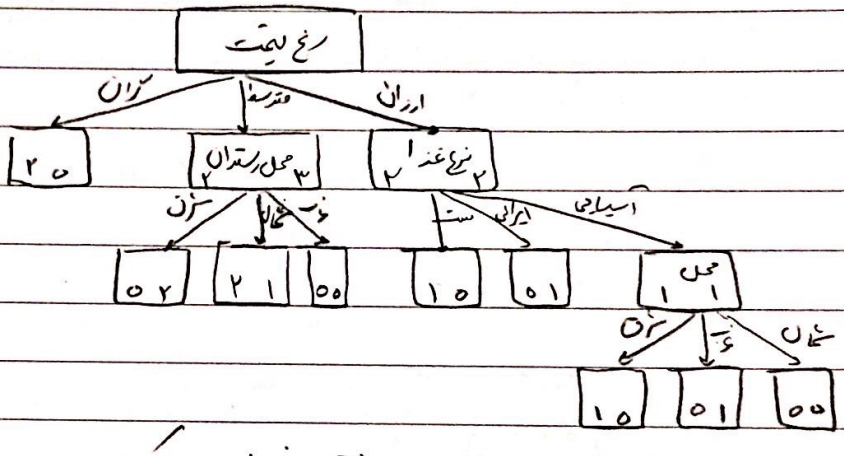
Scanned with CamScanner

subject : date :

محدودیت ها
- vegeterian → 0/1 , 1/1
- gluten free → 1/2 , 1/2
- none → 1/2 , 1/2

$\rightarrow H = 0 + \frac{1}{2}\left(-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right) + \frac{1}{2}\left(-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right) = 1$

کمترین انتروپی مربوط به محل رستوران است در نتیجه براساس آن ادامه می دهیم

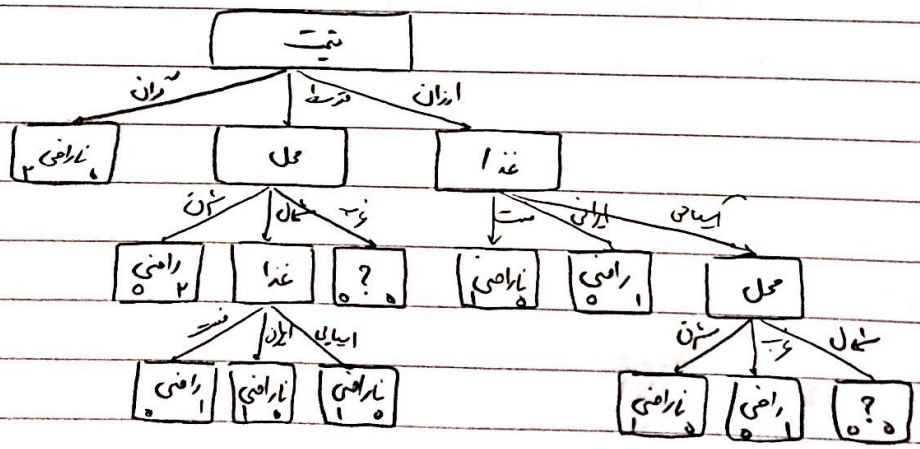


فقط برای متوسط و شمال مانده است در نتیجه دو نیچر را بررسی می کنیم:

نوع غذا
- ست → 0/1 , 1/1
- ایرانی → 1/1 , 0/1
- آسیایی → 1/1 , 0/1

$H = 0$

پس همین را انتخاب می کنیم. درخت نهایی به شکل زیر است



ب) با درخت پس می رویم:

12) ارزان ← ست بود ← ناراضی = 0

13) متوسط ← شرق ← راضی = 1

14) ارزان ← ایرانی ← راضی = 1

(5) ← ارزان ← آسیایی ← شرق ← ناراضی = 0

(6) ← ارزان ← ایرانی ← راضی = 1

ب)

$[0,1,0,1,0]$
$[0,1,1,0,1]$ ← $TP=1$, $TN=1$, $FP=2$, $FN=1$

$$f.score = \frac{TP}{TP+\frac{1}{2}(FP+FN)} = \frac{1}{1+\frac{1}{2}(3)} = \frac{2}{5} = \boxed{0.4}$$

سوال ۵:

الف) احتمال اینکه داده ها متعلق به کلاس ها باشند را به صورت $p(\phi_n,t_n)\times\pi_k$ می نویسیم

می دانیم داده ها مستقلند در نتیجه می توانیم احتمال ها را ضرب کنیم

$$\prod_{1}^{N}\prod_{1}^{k}\left(p(\phi_n,t_n)\pi_k\right)^{t_{nk}} \xrightarrow[\text{جمع log می گیریم}]{\text{برای تبدیل ضرب}} \sum_{1}^{N}\sum_{1}^{k} t_{nk}\log\left(p(\phi,t_n)\pi_k\right) =$$

$$\sum_{1}^{N}\sum_{1}^{k} t_{nk}\log\left(p(\phi,t_n)\right) + \sum_{1}^{N}\sum_{1}^{k} t_{nk}\log(\pi_k)$$

یک شرط داریم که $\sum_{1}^{k}\pi_k$ ها برابر یک است در نتیجه معادله لاگرانژ را می نویسیم و مشتق را حساب می کنیم

$$\log\left(p(\phi,t_n)\,|\,\pi_k\right) + \lambda\left(\sum_{1}^{k}\pi_k - 1\right) \rightarrow \begin{cases}\frac{\partial f}{\partial \pi} = \frac{N_i}{\pi_i}\\ \frac{\partial g}{\partial \pi} = 1\end{cases} \rightarrow \lambda = \frac{-N_i}{\pi_i}$$

اگر یک سیگمای k تایی حساب کنیم داریم که $\lambda = -N$ ← $-N = \frac{-N_i}{\pi_i} \rightarrow \boxed{\pi_i = \frac{N_i}{N}}$

ب)

$$p(\phi,c_k) = N(\phi\,|\,\mu_k,\Sigma)$$

می دانیم داده ها مستقل اند در نتیجه احتمال آنها در هم ضرب می شود

$$\prod_{1}^{N}(2\pi)^{-\frac{d}{2}}\det|\Sigma|^{-\frac{1}{2}}e^{-\frac{1}{2}(\phi_n-\mu)^T\Sigma^{-1}(\phi_n-\mu)}$$

μ_k و Σ را می خواهد سه برای کلاس k احتمال را حساب می کنیم

$$\xrightarrow[\text{جمع تبدیل شود}]{\text{log می گیریم تا ضرب}} \log\left(p(\phi,c_k)\right) = -\frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{1}^{N}(\phi_n-\mu_k)^T\Sigma^{-1}(\phi_n-\mu_k) + C$$

برحسب μ_k مشتق می گیریم ←

$$\frac{d}{d\mu} -\frac{1}{2}\sum_{1}^{N}\left(\phi_n^T\Sigma^{-1}\phi_n - \phi_n^T\Sigma^{-1}\mu_k - \mu_k^T\Sigma^{-1}\phi_n + \mu_k^T\Sigma^{-1}\mu_k\right) = 0 \rightarrow -\frac{1}{2}\sum_{1}^{N} -2\phi_n^T\Sigma^{-1} + 2\mu_k\Sigma^{-1} = 0$$

$$\rightarrow N\mu_k = \sum_{1}^{N}\phi_n^T \rightarrow \boxed{\mu_k = \frac{\sum_{1}^{N}\phi_n^T}{N}}$$

Scanned with CamScanner

subject : date :

$\frac{\partial}{\partial A} tr(AB) = \frac{\partial}{\partial A} tr(BA) = B^T$

$x^T A x = tr(A x x^T)$ ←

برای بدست آوردن Σ این ماتریس Σ مشتق می‌گیریم:

با استفاده از ویژگی‌های ماتریس‌ها می‌دانیم:

$$\frac{d}{d\Sigma}\left(\frac{-N}{2}\log|\Sigma| - \frac{1}{2} tr\left(\Sigma^{-1}\sum_{1}^{N}[(Q_n - \mu_k)(Q_n - \mu_k)^T]\right)\right)$$

$$\rightarrow \frac{+N}{2}\frac{1}{\Sigma} - \frac{1}{2}\sum_{1}^{N}(Q_n - \mu_k)(Q_n - \mu_k)^T = 0 \rightarrow \Sigma = \frac{1}{N}\sum_{n=1}^{N}(Q_n - \mu_k)(Q_n - \mu_k)^T$$

Scanned with CamScanner