Subject:

Date:

سوال ۴

آ) می‌خواهیم از maximum likelihood استفاده کنیم: دست کشیدی برای سادگی فضای داده‌ها را یک بردار با k بعد می‌بریم:

$$\hat{X}_i = \begin{cases} x_i = j \Rightarrow \hat{x}_{ij} = 1 \\ x_i \neq j \Rightarrow \hat{x}_{ij} = 0 \end{cases}$$

حال داریم:

$$ML \rightarrow \prod_{i=1}^{n} \prod_{l=1}^{k} (P_l)^{1[x_i = l]}$$

→ به ازای label داده نشده ⇐ برابر خودشان در به ازای نشان داده برابر است ($P_j^0 = 1$)

$$\Rightarrow ML = \prod_{i=1}^{n} \prod_{l=1}^{k} (P_l)^{\hat{x}_{il}} \Rightarrow \lg ML = \sum_{i=1}^{n} \sum_{l=1}^{k} \hat{x}_{il} \lg(P_l)$$

langrage multiplier

$$\Rightarrow \begin{cases} \sum_{i=1}^{n} \sum_{l=1}^{k} \hat{x}_{il} \lg P_l = \lg ML \\ \sum_{k=1}^{k} P_k = 1 \end{cases} \Rightarrow \frac{\partial}{\partial P_j} \lg ML - \frac{\partial}{\partial P_j} \left(\sum_{k=1}^{k} P_k - 1 \right) \times \lambda = 0$$

$$\Rightarrow \sum_{i=1}^{n} \hat{x}_{ij} \times \frac{1}{P_j} - \lambda = 0 \Rightarrow \frac{\sum_{i=1}^{n} \hat{x}_{ij}}{P_j} = \lambda \Rightarrow P_j = \frac{\sum_{i=1}^{n} \hat{x}_{ij}}{\lambda}$$

$$\Rightarrow \sum P_k = 1 \Rightarrow \sum_{k=1}^{k} \sum_{i=1}^{n} \frac{\hat{x}_{ik}}{\lambda} = 1 \Rightarrow$$

$$\lambda = \sum_{k=1}^{K} \sum_{i=1}^{n} \hat{x}_{ik}$$

$$\Rightarrow P_j = \frac{\sum_{i=1}^{n} \hat{x}_{ij}}{\sum_{k=1}^{k} \sum_{i=1}^{n} \hat{x}_{ik}} \Rightarrow \boxed{P_j = \frac{N_j}{N}}$$ که همان تخمین تجربی است.

در بالا Notation همان P ← $\hat{P}$ لازم تخمینی است.

Subject:
Date:

دارایه
ب

$$R = E_P[TV(\hat{P}, P)] = E_P[\frac{1}{2}\sum_k |P_i - \hat{P}_i|]$$

$$= \frac{1}{2} E_P[\sum_k |P_i - \hat{P}_i|]$$

$$\sum_k |P_i - \hat{P}_i| = \langle \overset{a}{1, \dots, 1} \rangle . \langle \overset{b}{|P_1 - \hat{P}_1|, \dots, |P_k - \hat{P}_k|} \rangle$$ کوشی شوارتز

$$a.b \le |a||b| \Rightarrow \sqrt{k} \times \sqrt{\sum |P_i - \hat{P}_i|^2}$$

$$\Rightarrow \frac{1}{2} E_P[\sum_k |P_i - \hat{P}_i|] \le \frac{1}{2} E_P[\sqrt{k}\sqrt{\sum_k |P_i - \hat{P}_i|^2}]$$

$$\Rightarrow \frac{\sqrt{k}}{2} E_P(\sqrt{\sum_k [P_i - \hat{P}_i]^2})$$ applying Jensen → concave $\sqrt{}$ $\Rightarrow \le \frac{\sqrt{k}}{2} \sqrt{E_P \sum |P_i - \hat{P}_i|^2}$

applying ML estimator

$\hat{P} = [P_i]$

$\hat{P} = \frac{N_i}{N}$

$$\Rightarrow R \le \frac{\sqrt{k}}{2} \sqrt{\sum_{P \in k} \underbrace{E_P[(P_i - \hat{P}_i)^2]}_{Var}} = \text{(1)} \Rightarrow R \le \frac{\sqrt{k}}{2} \sum_i \sqrt{E_P[(P_i - \hat{P}_i)^2]} \text{ (2)}$$

$$\sqrt{a+b} \le \sqrt{a} + \sqrt{b} \Leftrightarrow a+b \le a+b+2\sqrt{ab} \Leftarrow 0 \le 2\sqrt{ab}$$

$a, b \ge 0$ $\quad$ $a, b \ge 0$

$$\text{(2)} \rightarrow \frac{\sqrt{k}}{2} \sum_i \sqrt{var(\hat{P}_i)} = \frac{\sqrt{k}}{2} \sum \sqrt{\frac{P(1-P)}{n}} = \frac{1}{2} \times \frac{\sqrt{k}}{\sqrt{n}} \sum \sqrt{P(1-P)}$$

applying ML $\quad var(\hat{P}_i) = \frac{x_1 + x_2 + \dots x_n}{n} = \frac{1}{n} var(x_i)$

Subject: Date:

$$①,② \Rightarrow R \leq \frac{1}{2}\sqrt{k}\sqrt{\sum_i E_P (P_i-\hat{P}_i)^2} \xrightarrow{\text{applying ML}} \frac{1}{2}\sqrt{k}\sqrt{\sum_i var(\hat{P}_i)}$$

$$var(\hat{P}_i) = \sum \frac{x_i = i}{n} = \frac{1}{n} var(x_i) = \frac{P_i(1-P_i)}{n} \quad ③$$

$$③ \Rightarrow \frac{1}{2}\sqrt{k}\sqrt{\sum \frac{P_i(1-P_i)}{n}} = \frac{1}{2}\sqrt{\frac{k}{n}}\sqrt{\sum P_i(1-P_i)} \leq \frac{1}{2}\sqrt{\frac{k}{n}}\sqrt{\cancelto{1}{\sum_i P_i}}$$

$$\Longrightarrow R \leq \frac{1}{2}\sqrt{\frac{k}{n}} \leq \sqrt{\frac{k}{n}}$$

از طرفی می دانیم که مقدار TV ماکزیمم برابر ۱ است

$$\Longrightarrow R \leq \min\left(\sqrt{\frac{k}{n}}, 1\right)$$

سوال ۴ ب)

$$D(P,\hat{P}) = \sum P_i \lg \frac{P_i}{N_i/N}$$

اگر $N_i = 0$ باشد $\Leftarrow D = \infty \Leftarrow$ یعنی در بدترین حالت از یکی از کلاسها هیچ سمپلی نبینیم.

$\Leftarrow$ فاصله $P, \hat{P}$ طبق D_{KL} بی نهایت می افتد.

Subject: Date:

سوال ۲.۴ داریم:

$$\text{minimax} \geq \frac{\varepsilon^2}{2} d \left(1 - \max_{d_H(v,v')=1} TV(P_v^n, P_{v'}^n)\right)$$

$\varepsilon \ll \frac{1}{k}$

$d = k \rightsquigarrow (\theta_v)_i$: if $v_i = 1 \rightsquigarrow \frac{1}{k} + \alpha$; if $v_i = 0 \to \varepsilon$ $\quad \varepsilon \ll \frac{1}{k}$

$$\alpha = \frac{1 - \sum \frac{1}{k} \mathbb{1}\{v_i = 1\} - \sum \mathbb{1}\{v_i = 0\}\varepsilon}{\sum \mathbb{1}\{v_i = 1\}}$$

به ۰ ها ε می‌دهیم و بیشتر ۱ ها را اساساً $\frac{1}{k}$، به مقدار باقی مانده افزایش می‌دهیم.

$$\Rightarrow \max TV(P_v^n, P_{v'}^n) \leq \sqrt{\frac{1}{2} D(P_v^n, P_{v'}^n)}$$

باید D

$D(\theta \| \theta') = \sum \theta_i \log \frac{\theta_i}{\theta'_i}$ حاسبه D $\Leftarrow$ برای کدام ها متفاوت است

$$\begin{bmatrix} \frac{1}{k}+\alpha \\ \varepsilon \\ \varepsilon \\ \varepsilon \\ \frac{1}{k}+\alpha \end{bmatrix} \Leftrightarrow \begin{bmatrix} \frac{1}{k}+\alpha' \\ \frac{1}{k}+\alpha' \\ \varepsilon \\ \varepsilon \\ \frac{1}{k}+\alpha' \end{bmatrix}$$

$$\sum_{\theta_i \neq 0 \text{ or } \theta'_i \neq 0} \varepsilon \log \frac{\varepsilon}{\frac{1}{k}+\alpha'}$$

$$D = \sum_{v_i = 1 \wedge v'_i = 1} \left(\frac{1}{k}+\alpha\right) \log \frac{\frac{1}{k}+\alpha}{\frac{1}{k}+\alpha'} + \varepsilon \log \frac{\varepsilon}{\frac{1}{k}+\alpha'}$$

$$D \leq -k \times \varepsilon \log \varepsilon \qquad \varepsilon \ll \frac{1}{k}$$

$$2 TV^2 \leq nD \Rightarrow TV \leq \sqrt{-\frac{1}{2} kn \varepsilon \log \varepsilon}$$

$$\Rightarrow 1 - \sqrt{-\frac{1}{2} k^n \varepsilon \log \varepsilon} \geq \frac{1}{2} \Rightarrow -\frac{1}{2} k \varepsilon^n \log \varepsilon \geq \frac{1}{2} \Rightarrow k \varepsilon \log \varepsilon^n \leq 1$$

$$n \varepsilon \log \varepsilon \leq \frac{1}{k} \Rightarrow \Rightarrow \varepsilon \log \varepsilon \approx \frac{1}{kn} \text{ !!} \quad \xrightarrow{\text{taylor}}$$

$$\varepsilon \left((\varepsilon - 1) - \frac{1}{2}(\varepsilon-1)^2 + \frac{1}{3}(\varepsilon-1)^3 - \frac{1}{4}(\varepsilon-1)^4 + \dots \right) \approx \frac{1}{kn} \Rightarrow$$

$$\varepsilon^2 = \frac{1}{kn} \rightsquigarrow \varepsilon = \sqrt{\frac{1}{kn}}$$

• بنابراین باند $\frac{1}{\sqrt{kn}}$ بدست می‌آید.

فکر می‌کنم

* باید $\sqrt{\frac{k}{n}}$ بدست می‌آمد که عبارت $\min(1, \sqrt{\frac{k}{n}}) \approx$ ثابت شد. اما ما با Assouad نتوانستیم به

باند خواسته شده برسیم.

اگر می‌خواستیم به $\sqrt{\frac{k}{n}}$ ← $\epsilon = \frac{\sqrt{k}}{\sqrt{n}}$

باید D را با ϵ و $\epsilon . 1$ $D \leq \frac{-1}{k}$ در می‌آمد ← چون سمت حقی عبارت D را آپلند کرده‌ام

به این باند برسیم.

اگر تخمین‌گر ساده‌تری ست به حال بگیریم ممکن بود حساب ساده‌تر باشد.

اکنون باند D_{KL} می‌تواند خیلی tight نیست به همین دلیل نتیجه دلخواه را نمی‌دهد.

تخمین‌گر من به داده‌های 1 اهمیت زیاد و صفرها را با ϵ اکلس می‌کند.