تمارین سری ششم درس SPML   حمیدرضا امیرزاده   ۴۰۱۲۰۶۹۹۹

۲) الف) برای داده i ام کلاس اول رابطه مقابل را داریم: $z_1^{(i)} = w_1^T h^{(i)} + b_1$

از آنجایی که این یک معادله خطی می‌باشد که $n+1$ مجهول دارد. بنابراین اگر $n+1$ بردار h مستقل خطی داشته باشیم، می‌توان معادله زیر را حل کرد:

$$\begin{bmatrix} z_1^{(1)} \\ \vdots \\ z_1^{(n+1)} \end{bmatrix} = \begin{bmatrix} h^{(1)} \\ \vdots \\ h^{(n+1)} \end{bmatrix} w_1 + \begin{bmatrix} b_1 \\ \vdots \\ b_1 \end{bmatrix}$$

بنابراین مطابق این روش می‌توان وزن‌های رگرسیون لجستیک را بازیابی کرد.

ب) از آنجایی که فضای مجهول ما $n+1$ بعدی است بنابراین اگر تعداد داده بیشتری در نظر گرفته شود، نتایج از داده $n+2$ به بعد را می‌توان به صورت ترکیب خطی $n+1$ داده اول به دست آورد. بنابراین اطلاعات جدیدی به دست نمی‌آید.

ج) برای هر بردار احتمال داریم

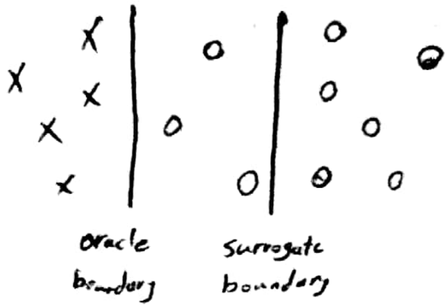
$$p_j = \frac{\exp(z_j)}{\sum_{k=1}^{L} \exp(z_k)}$$

حال می‌توان یکی از احتمال‌های خروجی را احتمال پایه در نظر گرفت و هر یک از سایر احتمال‌ها را بر آن تقسیم کرد و سپس لگاریتم بگیریم:

$$\log \frac{p_j}{p_i} = \log \frac{\exp(z_j)}{\exp(z_i)} = z_j - z_i = (w_j - w_i)h + b_j - b_i$$

در اینجا نیز $w_j - w_i$، n عنصر دارد و یک بایاس داریم. بنابراین اگر فرض کنیم خروجی‌های شبکه عصبی مستقل خطی باشند نیاز به n داده داریم که بتوان n معادله برای $n+1$ مجهول را تشکیل دهیم و مشابه حالت الف مسئله را حل کنیم. اگر از بیشتر از n داده استفاده کنیم فایده‌ای ندارد زیرا می‌توان یکی از داده‌ها را برحسب سایر داده‌ها نوشت. وزن‌هایی که از حل این معادلات به دست می‌آید برابر است با $w_j - w_i$ که w_i یک وزن ثابت است و j باید از ۱ تا K تغییر کند. برای بایاس نیز داریم $b_j - b_i$

③ الف) مثال زیر را در نظر بگیرید:



- اگر مطابق $S_{p+1} = \{x + \lambda_{p+1} \, sgn(J_F[\tilde{O}(x)]) : x \in S_p\} \cup S_p$ افزودن را انجام دهیم، داده های جدید به طریقی ساخته می‌شوند که از مرز oracle فاصله بیشتری بگیرند و در واقع پیش‌بینی oracle روی آن‌ها اطمینان بیشتری یابد. واضح است که اگر مدل جایگزین را روی چنین نمونه‌هایی آموزش دهیم، کمکی به نزدیک کردن مرز [redacted] جایگزین به مرز oracle نمی‌کند.

- اگر مطابق $S_{p+1} = \{x - \lambda_{p+1} \, sgn(J_F[\tilde{F}(x)]) : x \in S_p\} \cup S_p$ افزودن را انجام دهیم، داده های جدید به طریقی ساخته می‌شوند که به مرز مدل جایگزین نزدیک شوند. در واقع مثل این است که برای مدل جایگزین نمونه خصمانه تولید کرده‌ایم. [redacted] در این روش تضمینی وجود ندارد که با آموزش مدل روی چنین نمونه داده‌هایی، مرز مدل جایگزین به مرز oracle نزدیک شود، زیرا لزوماً به مرز oracle نزدیک نمی‌شویم.

- اگر مطابق $S_{p+1} = \{x - \lambda_{p+1} \, sgn(J_F[\tilde{O}(x)]) : x \in S_p\} \cup S_p$ افزودن را انجام دهیم، داده های جدید به طریقی ساخته می‌شوند که به مرز oracle نزدیک شوند. بنابراین با آموزش مدل جایگزین روی چنین نمونه داده‌های می‌توان مرز مدل جایگزین را به مرز oracle نزدیک کرد، که این هدف اصلی ما است.

۳) برای اینکه ε-DP تضمین شود باید تابع هزینه privacy در تمام نقاط تعریف شده باشد، ~~[illegible]~~ به بی‌نهایت میل نکند. اما وقتی ب در توزیع احتمال برای دو دیتاست x ، y که $\|x-y\|_1 \le 1$ گرفته شده، و در حالتی که داده‌های متفاوت در این دو دیتاست برای یکی مقدار ۱ و برای دیگری صفر باشد. به عنوان مثال

$x_i = 1$ ، $y_i = 0$ داریم: $f(x) = f(y)+1$

حال در عبارت loss برای ترم $\frac{p_x(z)}{p_y(z)}$ چون $f(x)=f(y)+1$ بنابراین میانگین توزیع $p_x(z)$ و $p_y(z)$ متفاوت است و در نقاطی خارج از بازه $[f(y)-\epsilon/4 , f(x)+\epsilon/4]$ که $p_y(z)$ دارای صفر است، $p_x(z)$ مقدار $\epsilon/4$ دارد و بنابراین $\frac{p_x(z)}{p_y(z)}$ در نقاطی برابر بی‌نهایت می‌شود. بنابراین عملاً مرزی در آن نقاط برای ε-DP بودن نداریم. بنابراین $\tilde{f}$ به صورت $O(\epsilon)$-DP نیست.

در بازه‌ای که هیچ‌کدام از $p_x(z)$ ، $p_y(z)$ دارای صفر نیستند ~~[illegible]~~ داریم:

$$\frac{p_x(z)}{p_y(z)} = \frac{\frac{1}{f(x)-\epsilon/4 - f(x)+\epsilon/4}}{\frac{1}{f(y)+\epsilon/4 - f(x)+\epsilon/4}} = \frac{1/\epsilon/4}{1/\epsilon/4} = 1 = \exp(0)$$

بنابراین در این بازه 0-DP است اما δ به صورت $O\left(\frac{1}{\epsilon/4}\right)$ می‌باشد. مطلوب این است که δ کوچک‌تر شود بنابراین باید ϵ را کوچک‌تر در نظر گرفت که این منجر به بزرگ‌تر شدن بازه $[-\epsilon/4 , \epsilon/4]$ می‌شود که خود باعث کاهش accuracy می‌شود.

Scanned with CamScanner

۵

$$Lip_{1,d_{ham}}(\hat{P}_n) = \sup\{\|\hat{P}_n(x_1^n) - \hat{P}_n(y_1^n)\|_1 \mid d_{ham}(x_1^n, y_1^n) \le 1\}$$

$$x_1^n = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{bmatrix}_{n\times 1} \qquad \hat{p}(x_1^n) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} 1\{x_i = 1\} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} 1\{x_i = K\} \end{bmatrix}_{K\times 1}$$

$$y_1^n = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_j \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} \qquad \hat{p}(y_1^n) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} 1\{y_i = 1\} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} 1\{y_i = K\} \end{bmatrix}_{K\times 1}$$

از آنجایی که $d_{ham}(x_1^n, y_1^n) \le 1$ بنابراین تمامی درایه‌های آن‌ها به جز حداکثر یکی با هم برابرند،

فرض کنیم درایه j ام آنها متفاوت باشد: $x_j \neq y_j$ ، $\forall i \neq j : x_i = y_i$

فرض کنید $\hat{p}(x_1^n)$ را به صورت زیر محاسبه کرده باشیم:

$$\hat{p}(x_1^n) = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{x_j} \\ \vdots \\ p_K \end{bmatrix}_{K\times 1}$$

از آنجایی که تفاوت x_1^n و y_1^n در درایه j ام آنهاست بنابراین، $\hat{p}(y_1^n)$ دقیقاً مشابه $\hat{p}(x_1^n)$ است با این تفاوت که در $\hat{p}(y_1^n)$ از احتمال دسته x_j، $\frac{1}{n}$ کاسته و به احتمال دسته y_j، $\frac{1}{n}$ افزوده می‌شود:

$$\hat{p}(y_1^n) = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{x_j} - \frac{1}{n} \\ \vdots \\ p_{y_j} + \frac{1}{n} \\ \vdots \\ p_K \end{bmatrix}_{K\times 1} \qquad \Rightarrow \sup(\|\hat{P}_n(x_1^n) - \hat{p}(y_1^n)\|_1) = 2/n$$

$$E[\|z-p\|_2^2] = E\Big[\sum_{i=1}^{k} (z_i - p_i)^2\Big]$$

$$= E\Big[\sum_i (\hat{p}_{n,i} + w_i - p_i)^2\Big] = \sum_i E\Big[\big((\hat{p}_{n,i} - p_i) + w_i\big)^2\Big]$$

$$= \sum_i E\Big[w_i^2 + (\hat{p}_{n,i} - p_i)^2 + 2 w_i (\hat{p}_{n,i} - p_i)\Big]$$

$$= \sum \Big[E[w_i^2] + E[(\hat{p}_{n,i} - p_i)^2] + 2E[w_i] + \big(\underbrace{\cancel{E[\hat{p}_{n,i}]}}_{p_i} - p_i\big)\Big]$$

$$= \sum \Big[E[w_i^2] + E[\hat{p}_{n,i}^2 + p_i^2 - 2\hat{p}_{n,i} p_i]\Big]$$

$$= \sum \Big[E[w_i^2] + E[\hat{p}_{n,i}^2] + p_i^2 - 2p_i E[\hat{p}_{n,i}]\Big]$$

$$= \sum \Big[E[w_i^2] + E[\hat{p}_{n,i}^2] - p_i^2\Big] = \sum \Big[E[w_i^2] + Var[\hat{p}_{n,i}]\Big]$$

$Var(K_i) = n p_i (1-p_i)$ $\qquad \hat{p}_{n,i} = \frac{K_i}{n}$ $\qquad \Rightarrow Var[\hat{p}_{n,i}] = \frac{1}{n^2} Var[K_i] = \frac{1}{n} p_i (1-p_i)$

$$Var[w_i] = 2\left(\frac{2}{n\varepsilon}\right)^2 = \frac{8}{n^2\varepsilon^2} = E[w_i^2] - E^2\underbrace{\cancel{[w_i]}}_{0}$$

$$\longrightarrow E[\|z-p\|_2^2] = \sum_{i=1}^{k} \Big[\frac{8}{n^2\varepsilon^2} + \frac{1}{n} p_i (1-p_i)\Big]$$

$$= \frac{8k}{n^2\varepsilon^2} + \frac{1}{n} \sum_{i=1}^{k} p_i (1-p_i)$$

$$\sum_{i=1}^{k} p_i (1-p_i) = \sum_{i=1}^{k} p_i - \sum_{i=1}^{k} p_i^2 = 1 - \sum_{i=1}^{k} p_i^2 \leq 1$$

$$\Rightarrow E[\|z-p\|_2^2] \leq \frac{8k}{n^2\varepsilon^2} + \frac{1}{n}$$

۶) الف) اگر فرض کنیم که

$$X_i = \begin{cases} 1 & : p \\ 0 & : 1-p \end{cases} \qquad E[X] \underset{\text{برنولی}}{=} P(X_i = 1) = p$$

آنگاه داریم:

$$Y_i = \begin{cases} X_i & ; \frac{1}{2}+\alpha \\ 1-X_i & ; \frac{1}{2}-\alpha \end{cases} \qquad E[Y] \underset{\text{برنولی}}{=} P(Y_i = 1)$$

$$P(Y_i=1) = P(Y_i=1 \mid X_i=1)\,P(X_i=1) + P(Y_i=1 \mid X_i=0)\,P(X_i=0)$$

$$= \left(\frac{1}{2}+\alpha\right)p + \left(\frac{1}{2}-\alpha\right)(1-p) = 2\alpha p + \frac{1}{2} - \alpha$$

$$\Rightarrow p = \frac{E[Y] + \alpha - 1/2}{2\alpha} \xrightarrow{\text{استفاده از تخمینگر نااریب برای میانگین}} \hat{p} = \frac{\sum_{i=1}^{n} \frac{Y_i}{n} + \alpha - 1/2}{2\alpha}$$

$$E[\hat{p}] = \frac{\frac{1}{n}E\left(\sum_{i=1}^{n} Y_i\right) + \alpha - 1/2}{2\alpha} = \frac{E[Y] + \alpha - 1/2}{2\alpha} = p$$

بنابراین $\hat{p}$ یک تخمینگر نااریب برای p است.

ب)

$$\text{Privacy loss} = \lg\left(\frac{P(Y_i=1 \mid X_i=1)}{P(Y_i=1 \mid X_i=0)}\right) = \lg\left(\frac{1/2+\alpha}{1/2-\alpha}\right)$$

$\alpha = 0$: $P.loss = \lg 1 = 0 \Rightarrow$ ODP است و هیچ اطلاعاتی نشت پیدا نمی‌کند

$\alpha = 1/2$: $P.loss = \lg \infty = \infty \Rightarrow$ هیچ حریم خصوصی وجود ندارد و همه اطلاعات نشت پیدا می‌کند

برای بدست آوردن دقت تخمینگر، از آنجایی که bias نداریم، فقط به محاسبه واریانس آن می‌پردازیم:

$$Var(\hat{p}) = var\left(\frac{\sum_{i=1}^{n}\frac{Y_i}{n} + \alpha - 1/2}{2\alpha}\right) = \frac{1}{4\alpha^2} Var\left(\sum_{i=1}^{n} \frac{Y_i}{n}\right) = \frac{1}{4\alpha^2 n} Var(Y_i)$$

$$var(Y_i) = E[Y_i^2] - E^2[Y_i] \quad , \quad E[Y_i^2] = E[Y_i]$$

$$\Rightarrow var(Y_i) = E[Y_i] - E^2[Y_i] = 2\alpha p + \frac{1}{2} - \alpha - \left(2\alpha p + \frac{1}{2} - \alpha\right)^2$$

$$= \frac{1}{4} - 4\alpha^2 p^2 + 4\alpha^2 p - \alpha^2$$

$$\underset{\text{جایگذاری}}{\Rightarrow} var(\hat{p}) = \frac{1}{n}\left(\frac{1}{16\alpha^2} - p^2 + p - \frac{1}{4}\right)$$

$\alpha = 0 \quad : \quad var(\hat{p}) = \infty$

$\alpha = 1/2 \quad : \quad var(\hat{p}) = \frac{p(1-p)}{n}$

توجه شود که $p(1-p)$ برابر $Var(X_i)$ می‌باشد. بنابراین در حالت $\alpha = 1/2$: $Var(Y_i) = var(X_i)$

ج) با نوشتن نامساوی چبیشف برای $\hat{p}$ داریم:

$$Pr[\,|\hat{p} - E[\hat{p}]| \geq \gamma\,] \leq \frac{var(\hat{p})}{\gamma^2}$$

$$\Rightarrow Pr[\,|\hat{p} - p| \geq \gamma\,] \leq \frac{1}{n\gamma^2}\left[\frac{1}{16\alpha^2} - \frac{1}{4} + p - p^2\right]$$

برای آنکه خطای تخمین از γ کوچکتر باشد یعنی $|\hat{p} - p| < \gamma$ باید عبارت $1 - \frac{1}{n\gamma^2}\left[\frac{1}{16\alpha^2} - \frac{1}{4} + p - p^2\right]$ به یک نزدیک شود که با افزایش n به اندازه کافی به این حد نزدیک و نزدیک‌تر می‌شویم.