پویا خانی ۹۹۲۱۰۲۸۳ درس یادگیری ماشین

تمرین سری اول

سوال ۱) همانطور که می‌دانیم، تعداد کل بازی ها در مرحله اول ۳۲ ، در مرحله دوم ۱۶، در مرحله سوم ۸، در مرحله چهارم ۴، در مرحله پنجم ۲ ، و در مرحله ششم یعنی مرحله آخر، ۱ است. یعنی مجموعا ۶۳ =۱ +۲ + ۴ + ۸ + ۱۶ + ۳۲ بازی انجام میشود تا قهرمان مسابقات مشخص شود. حال می‌دانیم رابطه امید ریاضی برای متغیر های تصادفی گسسته می‌شود:

$$E[x] = \sum_{i=1}^{N} x_i * P_X(x_i)$$

که x_i همان امتیاز کسب شده در مسابقه i در صورت پیش بینی صحیح و $P_X(x_i)$ تابع توزیع احتمال آن مسابقه را نشان میدهد(احتمال اینکه صحیح پیش بینی انجام شود). همچنین میدانیم که احتمال اینکه در مرحله اول قمارباز درست حدس بزند ½ است. در مرحله دوم، ۳ حالت میتواند به وجود بیاید:

۱- برنده دو بازی در مرحله اول را درست حدس زده باشد و الان در مرحله دوم بازی بین دو تیم مد نظر او برگزار شود و برنده این بازی را درست حدس بزند که احتمال آن ½ است و ۲ امتیاز کسب میکند.
۲- برنده دو بازی در مرحله اول را درست حدس زده باشد و الان در مرحله دوم بازی بین دو تیم مد نظر او برگزار شود و برنده این بازی را درست حدس نزند که احتمال آن ½ است و صفر امتیاز کسب میکند.
۳- برنده یکی از دو بازی مرحله اول را درست حدس زده باشد و الان در مرحله دوم، تیمی که در مرحله اول برنده شدن اش را درست حدس زده بوده را نیز برنده اعلام کند و واقعا هم برنده شود که ۲ امتیاز کسب میکند.
۴- برنده یکی از دو بازی مرحله اول را درست حدس زده باشد و الان در مرحله دوم، تیمی که در مرحله اول برنده شدن اش را درست حدس زده بوده را نیز برنده اعلام کند ولی این اتفاق نیفتد و تیم دیگر برنده شود که امتیاز صفر میگیرد.
۵- برنده هر دو بازی مرحله اول را اشتباه حدس بزند که در این صورت زیرشاخه آن همیشه اشتباه میشود و امتیاز صفر دریافت میکند.

پس اینطور میتوان اثبات کرد که در مرحله اول که همه چیز عادی است و ۳۲ بازی برگزار میشود که در هر بازی به صورت مستقل ½ احتمال حدس درست وجود دارد که در این حالت ۱ امتیاز دریافت میکند.

در مرحله دوم، در صورت حدس درست، امتیاز کسب شده دو برابر میشود اما حدس درست نسبت به مرحله اول، اول، احتمال نصف دارد یعنی ¼ چون در دو حالت این اتفاق می افتد: دو بازی مرحله اول را درست حدس زده و بازی مرحله دوم را هم

درست حدس بزند که احتمال آن میشود $½ * ½ * ½ = 1/8$. همچنین احتمال دیگر آن است که حالت ۳ اتفاق بیفتد که احتمال آن نیز $½ * ½ * ½ = 1/8$ است. مجموعا میشود $1/8 + 1/8 = ¼$.

به همین ترتیب میتوان اثبات کرد عدد درون پرانتز رابطه زیر همیشه $½$ میشود. یعنی امید ریاضی نهایی برابر است با:

$$E[x] = 32\left(1 * \frac{1}{2}\right) + 16\left(2 * \frac{1}{4}\right) + 8\left(4 * \frac{1}{8}\right) + 4\left(8 * \frac{1}{16}\right) + 2\left(16 * \frac{1}{32}\right) + 1(32 * \frac{1}{64}) = 31.5$$

سوال ۲) ابتدا یک متغیر که نسبت طول سوزن به فاصله بین خطوط کاغذ را نشان میدهد، تعریف میکنیم:

$$k \equiv \frac{l}{d}$$

باید دو حالت را بررسی کنیم. حالت اول وقتی است که طول سوزن از فاصله بین خطوط کوچک‌تر مساوی باشد. یعنی k کوچک‌تر مساوی ۱ باشد. حالت بعدی وقتی است که طول سوزن بزرگتر از فاصله بین خطوط باشد یعنی k بزرگتر از ۱ باشد.

<u>بررسی حالت اول:</u> در حالتی که سوزن کوچکی داریم، احتمال اینکه سوزن پس از پرتاب روی کاغذ، یکی از خطوط را قطع کند برابر است با:

$$P(k) = \int_0^{2\pi} \frac{l\,|\cos(\theta)|}{d}\,\frac{d\theta}{2\pi} = \frac{2l}{\pi d}\int_0^{\frac{\pi}{2}} \cos(\theta)\,d\theta = \frac{2l}{\pi d} = \frac{2k}{\pi}$$

به عنوان روش دوم برای حل این حالت، با فرض اینکه x برابر باشد با فاصله مرکز سوزن با نزدیک ترین خط و همچنین θ برابر باشد با زاویه بین سوزن و خط، داریم:

$$PDF(x) = \begin{cases} \frac{2}{d} & 0 \le x \le \frac{d}{2} \\ 0 & elsewere \end{cases}$$

$$PDF(\theta) = \begin{cases} \frac{2}{\pi} & 0 \le \theta \le \frac{\pi}{2} \\ 0 & elsewere \end{cases}$$

حال تابع چگالی احتمال مشترک این دو متغیر تصادفی مستقل میشود ضرب تابع های چگالی احتمال آن ها یعنی:

$$PDF(x,\theta) = \begin{cases} \frac{4}{d\pi} & 0 \le x \le \frac{d}{2}\ ,\ 0 \le \theta \le \frac{\pi}{2} \\ 0 & elsewere \end{cases}$$

لذا سوزن خط را قطع میکند فقط اگر $x \le \frac{l}{2}\sin(\theta)$.

حال احتمال قطع کردن خط توسط سوزن وقتی سوزن کوچکی داشته باشیم برابر است با:

$$P(k) = \int_0^{\frac{\pi}{2}} \int_0^{\frac{l}{2}\sin(\theta)} \frac{4}{d\pi}\, dx\, d\theta = \frac{2l}{d\pi} = \frac{2k}{\pi}$$

بررسی حالت دوم: در حالتی که سوزن بزرگی داریم، احتمال اینکه سوزن پس از پرتاب روی کاغذ،حداقل یکی از خطوط را قطع کند به طریق مشابه، برابر است با:

$$P(k) = \int_0^{\frac{\pi}{2}} \int_0^{m(\theta)} \frac{4}{d\pi}\, dx\, d\theta \quad \quad where \quad \quad m(\theta) = min\left\{\frac{l}{2}\sin(\theta), \frac{d}{2}\right\}$$

با تقسیم بازه انتگرال به دو بازه $d < l \sin(\theta)$ و $d > l \sin(\theta)$ داریم:

$$P(k) = \int_0^{sin^{-1}\left(\frac{d}{l}\right)} \int_0^{\frac{l}{2}\sin(\theta)} \frac{4}{d\pi}\, dx\, d\theta + \int_{sin^{-1}\left(\frac{d}{l}\right)}^{\frac{\pi}{2}} \int_0^{\frac{d}{2}} \frac{4}{d\pi}\, dx\, d\theta$$

$$= \int_0^{sin^{-1}\left(\frac{d}{l}\right)} \frac{2l}{d\pi} \sin(\theta)\, d\theta + \int_{sin^{-1}\left(\frac{d}{l}\right)}^{\frac{\pi}{2}} \int_0^{\frac{d}{2}} \frac{2}{\pi}\, d\theta$$

$$= \frac{2}{\pi}\left(\frac{l - \sqrt{l^2 - d^2}}{d} + cos^{-1}\left(\frac{d}{l}\right)\right)$$

سوال ۳) الف – با استفاده از تعریف امید ریاضی متغیر تصادفی X و همچنین نامنفی بودن آن، داریم:

$$E(x) = \int_{-\infty}^{+\infty} x\, f(x)\, dx = \int_0^{+\infty} x\, f(x)\, dx$$

$$\rightarrow E(x) = \int_0^{\alpha} x\, f(x)\, dx + \int_{\alpha}^{+\infty} x\, f(x)\, dx$$

$$\geq \int_{\alpha}^{+\infty} x\, f(x)\, dx \geq \int_{\alpha}^{+\infty} \alpha\, f(x)\, dx = \alpha \int_{\alpha}^{+\infty} f(x)\, dx = \alpha\, P(X \geq \alpha)$$

$$\rightarrow P(X \geq \alpha) \leq \frac{E(x)}{\alpha}$$

ب – اگر متغیر تصادفی X در قسمت الف را برابر با $(Z - \mu)^2 = \sigma^2$ و همچنین مقدار آلفا را برابر با ϵ^2 در نظر بگیریم، با توجه به نامساوی مارکوف داریم:

$$P(|Z - \mu| \geq \epsilon) = P((Z - \mu)^2 \geq \epsilon^2) \leq \frac{E((Z - \mu)^2)}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2}$$

پ–۱- برای تخمین مقدار π ، اگر نقاط تصادفی ای که انتخاب میکنیم، داخل دایره باشد آن را میپذیریم و اگر بیرون دایره بود، آنرا رد میکنیم. نسبت نقاطی که میپذیریم به نقاطی که نمیپذیریم میشود نسبت مساحت های آن ها یعنی $\frac{\pi}{4}$.

دو عدد تصادفی در بازه [-1,1] به نام های m و n انتخاب میکنیم. اگر $m^2 + n^2 \leq 1$ بود، یعنی نقطه درون دایره است و $X_i = 1$ میشود اما اگر $m^2 + n^2 > 1$ باشد، $X_i = 0$ میشود. اینکار را به تعدادn بار تکرار میکنیم. داریم:

$$P(X = 1) = P(m^2 + n^2 \leq 1) = \frac{\text{مساحت دایره}}{\text{مساحت مربع}} = \frac{\pi}{4}$$

$$and\ P(X = 0) = 1 - P(X = 1) = 1 - \frac{\pi}{4}$$

$$\rightarrow E[x] = \mu = \frac{\pi}{4} \quad and \quad Var(X) = \sigma^2 = \frac{\pi}{4}\left(1 - \frac{\pi}{4}\right)$$

با استفاده از نامساوی چبیشف داریم:

$$P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \frac{\pi}{4}\right| \geq \epsilon\right) \leq \frac{\frac{\pi}{4}\left(1 - \frac{\pi}{4}\right)}{n\epsilon^2}$$

وقتی میگوییم میخواهیم خطای تخمین از ۱ درصد کمتر باشد یعنی $\pi \pm .01 \rightarrow \frac{\pi}{4} \pm \frac{1}{400}$ لذا $\epsilon = \frac{1}{400}$.

هم چنین وقتی میخواهیم ۹۵ درصد قطعیت داشته باشیم یعنی داریم:

$$\frac{\frac{\pi}{4}\left(1 - \frac{\pi}{4}\right)}{n\epsilon^2} \leq \frac{5}{100}$$

نکته ای که هست این است که ما واریانس را نداریم. اما میدانیم که تابع *p(1-p)* وقتی *p=1/2* باشد ماکسیمم میشود و ماکسیمم آن ¼ است . پس میتوانیم واریانس را کوچک تر مساوی ¼ فرض کنیم. بنابرین نتیجه میشود که:

$$\frac{1/4}{n(1/400)^2} \leq \frac{5}{100} \quad \rightarrow \quad n \geq 800{,}000$$

توجه شود این قسمت با این فرض حل شد که مقدار دقیق عدد پی را نمیدانیم و یعنی واریانس را باید تقریب بزنیم.اگر مقدار پی را فرض کنیم میدانیم، واریانس مقدار دیگری بدست می آید و صورت نامساوی هایلایت شده ضربدر ۴ میشود و به جای

اپسیلون نیز یک صدم میگذاریم و عدد ۵۳۹۳۵۳ بدست می آید.چون در صورت سوال این ابهام ذکر نشده بود، بنده فرض کردم مقدار پی کاملا ناشناخته است و با فرض اول راه حل کامل را نوشتم.

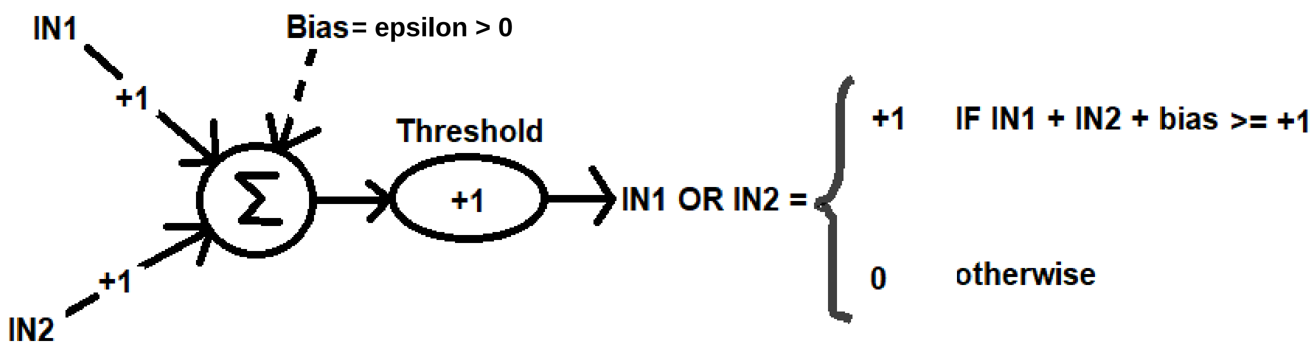
پ-۲- با استفاده از نامساوی هافدینگ باید احتمال اینکه اختلاف عدد پی و $\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$ بیشتر مساوی میزان خطا یعنی *0.01* باشد، کمتر از یک مقدار کوچک زیر باشد:

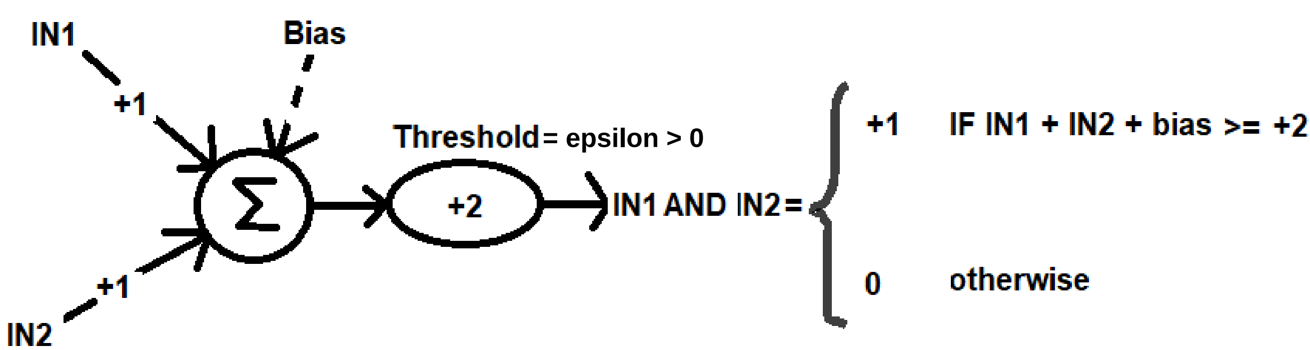
$$P\left(\left|\frac{\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)}{4} - \frac{\pi}{4}\right| \geq \frac{0.01}{4}\right) \leq 2e^{-2(4*10^{-6})n}$$

حال با توجه به اینکه در سوال گفته شده باید با اطمینان ۹۵ درصد این مقدار خطا را داشته باشیم پس:

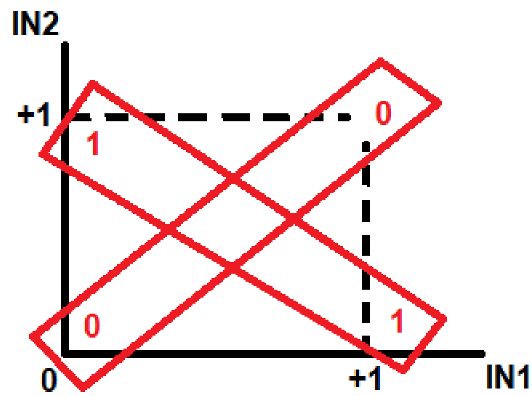
$$2e^{-2(4*10^{-6})n} \leq \frac{5}{100} \rightarrow n \geq 295111$$

سوال ۴ – الف –





سوال ۴ – ب – در پرسپترون تک لایه، خروجی ترکیب خطی از ورودی ها است. یعنی اینکه میتونه فضای ورودی-خروجی را classify بکند اگر و تنها اگر بتوانیم با یک خط مستقیم آن هارا در فضای مختصات تفکیک کنیم. در مسئله XOR این کار ممکن نیست زیرا نمیتوان خروجی صفر و یک را در فضای مختصات با یک خط جدا کرد:



برای اثبات اینکه به صورت خطی تفیک پذیر نیستند، نامعادله های زیر را میتوانیم داشته باشیم:

(I) Sign(w0) = 0 → w0 <= 0

(II) Sign(w0+w2) = 1 → w0+w2 > 0

(III) Sign(w0+w1) = 1 → w0+w1 > 0

(IV) Sign(w0+w1+w2) = 0 → w0+w1+w2 <= 0

→ W2 > 0, w1 > 0, w0+w1+w2 !<= 0 → it is proof for this question

سوال ۴ – ج – اگر تعداد بروزرسانی های الگوریتم پرسپترون را k در نظر بگیریم، فرض میکنیم که برای یک عدد ثابت N داریم $\|w_{(k+1)}\| \le Nk$. حال با توجه به فرض هایی که در صورت سوال گفته شده:

$$w_{(k+1)} \cdot w^* = \left(w_{(k)} + y^{(i)} x^{(i)} \right) w^* = w_{(k)} \cdot w^* + y^{(i)} \left(w^* \cdot x^{(i)} \right) \ge w_{(k)} \cdot w^* + \rho$$

$$\rightarrow w_{(k+1)} \cdot w^* \ge k\rho$$

حال میدانیم که اندازه بردار وزن داده شده در صورت سوال واحد است و همچنین:

$$\|w_{(k+1)}\| \times \|w^*\| \ge w_{(k+1)} \cdot w^*$$

لذا با توجه به نامساوی های بالا نتیجه میشود:

$$\left\|w_{(k+1)}\right\| \geq k\rho$$

حال فرض میکنیم برای یک عدد ثابت دیگر به نام M داریم $\left\|w_{(k+1)}\right\| \leq M\sqrt{k}$. حال با توجه به دانش قبلی و فرض سوال داریم:

$$\left\|w_{(k+1)}\right\|^2 = \left\|w_{(k)} + y^{(i)}x^{(i)}\right\|^2 = \left\|w_{(k)}\right\|^2 + (y^{(i)})^2\left\|x^{(i)}\right\|^2 + 2y^{(i)}\left(w_{(k)} \cdot x^{(i)}\right)$$
$$\leq \left\|w_{(k)}\right\|^2 + r^2$$

نامساوی بالا اتفاق افتاد چون که قسمت هایلایت شده در k امین بروزرسانی، عددی صفر یا منفی است. حال نتیجه میشود که :

$$\left\|w_{(k+1)}\right\|^2 \leq k\,r^2 \rightarrow \left\|w_{(k+1)}\right\| \leq \sqrt{k}\,r$$

حال با ترکیب دو نامساوی نتیجه شده داریم:

$$k\rho \leq \left\|w_{(k+1)}\right\|^2 \leq \sqrt{k}\,r \rightarrow k\rho \leq \sqrt{k}\,r \rightarrow k^2\rho^2 \leq kr^2 \rightarrow k \leq \left(\frac{r}{\rho}\right)^2$$

سوال ۵ – آ – خیر. زیرا فرضیه های موجود ممکن است بنا به شرایطی اصلا کل فضای نمونه را پوشش ندهند. برای مثال فرض میکنیم نمونه های گرفته شده(D) همگی برچسب ۱- داشته باشند. لذا الگوریتم A1 بخاطر کمینه کردن خطا، فرضیه h2 را انتخاب میکند. H2 داده های درون D را پوشش میدهد و به عبارتی fit میشود اما برای کل فضای نمونه که برای مثال همه داده ها با برچسب ۱+ هستند، اصلا مناسب نیست. از طرف دیگر تابع انتخاب تصادفی یعنی احتمال ½ برای ۱+ و احتمال ½ برای ۱- لذا در بدترین حالت که همان حالتی است که گفته شد، در نیمی از موارد درست عمل میکند پس تضمینی در کار نیست.

قسمت ب – بله .در مثال که برای قسمت آ گفته شد، گفتیم که تمام داده های درون D برچسب ۱- و داده های بیرون D برچسب ۱+ دارند. حال طبق فرض سوال، برعکس این حالت را در نظر میگیریم. یعنی درون ۱+ و بیرون ۱- . با توجه به این فرض ، الگوریتم A1 فرضیه h1 را انتخاب کرده و الگوریتم A2 فرضیه h2 را به اجبار انتخاب میکند. حال با بررسی کل فضای نمونه ، مشخص میشود که فرضیه h2 بسیار مناسب تر از h1 بوده است و دلیل شکست الگوریتم h1 که A1 را انتخاب کرد، در اصل محدود بودن دید آن به نمونه های درون D بود.

قسمت پ – در این مثال گفته شده، هم داده های درون D برچسب ۱+ دارند و هم ۹۰ درصد داده های کل فضای نمونه هم برچسب ۱+ دارند. یعنی عملا الگوریتم A1 بخاطر کمینه کردن خطا روی داده های نمونه گیری شده فرضیهh1 را انتخاب میکند و سپس با اعمال روی کل مجموعه داده ۹۰ درصد با داده ها fit میشود. در صورتی که الگوریتم A2 به اجبار h2 را انتخاب کرده و در ۱۰ درصد موافق با داده ها fit میشود. لذا همیشه الگوریتم A1 بهتر از A2 عمل میکند.

قسمت ت – طبق استدلال های گفته شده در قسمت پ ، اگر p کوچک تر از 0.5 باشد ، یعنی کمتر از نصف داده های درون کل فضای نمونه برچسب ۱+ داشته باشند اما کل داده های درون D برچسب ۱+ داشته باشند، طبیعتا الگوریتم A1 فرضیه h1 را انتخاب کرده و الگوریتم A2 هم h2 را انتخاب میکند. لذا موفقیت h1 کمتر از ۵۰ درصد و موفقیت h2 بیشتر از ۵۰ درصد میشود. لذا جواب سوال میشود $p < 0.5$