دانشکده‌ی مهندسی کامپیوتر

یادگیری ماشین		زمستان ۱۳۹۹
	تمرین سری اول	
مدرس: دکتر محمّدحسین رهبان		زمان تحویل: ۱۵ اسفند

سوال ۱ در یک لیگ حذفی فوتبال، ۶۴ تیم وجود دارد. این تیم‌ها دو به دو در دور اول با هم مسابقه داده و برندگان به دور دوم مسابقات می‌روند و مجدد دو به دو با هم مسابقه می‌دهند. به همین ترتیب این مسابقات تا جایی که تنها یک تیم باقی بماند، ادامه خواهند داشت. با توجه به این‌که ۶۴ تیم در لیگ هستند، بازی‌های این مسابقات ۶ دور به طول می‌انجامد.

شخصی در چالشی برای پیش‌بینی نتایج بازی‌های این لیگ شرکت کرده است. او باید در ابتدای فصل، پیش‌بینی خود از همه بازی‌ها را ارائه دهد و در انتهای فصل امتیازی براساس پیش‌بینی‌هایی که در ابتدای فصل انجام داده بود، می‌گیرد. قانون چالش به این شکل است که به ازای هر پیش‌بینی درست در دور اول، ۱ امتیاز بدست می‌آورد. در هر دور بعدی از مسابقات، امتیاز کسب شده دو برابر دور قبلی می‌شود (اگر مسابقه‌ای در دور دوم را درست پیش‌بینی کرده باشد ۲ امتیاز می‌گیرید و ...).

این شخص هیچ اطلاعاتی از این تیم‌ها ندارد، به همین خاطر تصمیم می‌گیرد که با پرتاب یک سکه سالم، تمام بازی‌ها را پیش‌بینی کند. امید ریاضی امتیاز نهایی وی در انتهای فصل را بدست آورید.

سوال ۲ سوزنی به طول L را روی کاغذی که دارای خط کشی‌های افقی با فاصله یکسان d است، پرتاب می‌کنیم. با این فرض که همه مکان‌های مرکز سوزن و همه جهت‌های قرار گرفتن آن هم احتمال هستند، با چه احتمالی این سوزن خطی روی صفحه را قطع می‌کند؟

سوال ۳

آ) با در نظر گرفتن متغییر تصادفی نامنفی X، نامساوی زیر را اثبات کنید (نامساوی مارکوف[1]).

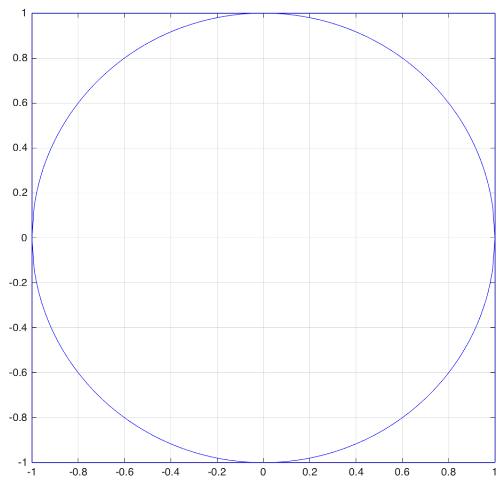
$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[X]}{\alpha}$$

ب) با در نظر گرفتن نتیجه بخش **آ**، نشان دهید برای متغیر تصادفی دلخواه Z با امید ریاضی μ و واریانس σ^2، نامساوی زیر برقرار است (نامساوی چبیشف[2]).

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

[1] Markov's Inequality
[2] Chebyshev's inequality

پ) می‌خواهیم مقدار عدد π را تخمین بزنیم. برای این کار روی صفحه مختصات دوبعدی، دایره‌ای به شعاع واحد و مربع محیطی آن را شبیه شکل زیر رسم می‌کنیم. مساحت این دایره π و مساحت مربع محیطی آن ۴ است. برای تخمین مقدار π تعدادی نقاط تصادفی داخل این مربع تولید کرده و نسبت تعداد نقاطی که داخل دایره قرار می‌گیرند را به تعداد کل به عنوان مقدار عدد π در نظر می‌گیریم.



۱. با استفاده از نامساوی چبیشف تعداد عددهای تصادفی‌ای که باید تولید کنیم تا با قطعیت ۹۵ درصد بدانیم که خطای تخمین از ۱ درصد کمتر است را مشخص کنید.

۲. با استفاده از نامساوی هافدینگ[1] به سوال قبل پاسخ دهید.

سوال ۴

آ) با مدل پرسپترون، توابع AND و OR با دو ورودی دودویی را مدل‌سازی کنید.

ب) نشان دهید که هیچ مدل پرسپترونی وجود ندارد که بتواند تابع XOR با دو ورودی دودویی را با خطای صفر مدل‌سازی کند.

پ) فرض کنید که بردار وزن $\mathbf{w}^\star$ با اندازه واحد وجود دارد که به ازای تمام نمونه‌های داخل مجموعه دادگان، نامساوی زیر برقرار است:

$$y^{(i)}\langle \mathbf{x}^{(i)}, \mathbf{w}^\star\rangle \geq \rho \qquad for\ i = 1, \ldots, n$$

همچنین فرض کنید تمام نمونه‌ها در کره‌ای به شعاع r قرار دارند، یعنی:

$$||\mathbf{x}^{(i)}|| \leq r \qquad for\ i = 1, \ldots, n$$

که $r, \rho > 0$ هستند. نشان دهید تعداد به روزرسانی‌های الگوریتم پرسپترون در این حالت از $\left(\frac{r}{\rho}\right)^2$ کمتر است (راهنمایی: یک کران بالا برای اندازه بردار وزن در مرحله kام و یک کران پایین برای ضرب داخلی $\langle \mathbf{w}^\star, \mathbf{w}_k\rangle$ پیدا کنید).

[1] Hoeffding's inequality

سوال ۵ مجموعه دادگان $\mathcal{D}$ شامل ۲۵ نمونه و یک تابع هدف نامشخص $f:\mathcal{X}\rightarrow\{-1,+1\}$ وجود دارند که در آن $\mathcal{X}=\mathbb{R}$ است. برای یادگیری تابع f یک مجموعه فرضیه ساده $\mathcal{H}=\{h_1,h_2\}$ داریم که h_1 یک تابع ثابت با مقدار $+1$ و h_2 یک تابع ثابت با مقدار -1 است. برای یادگیری f، دو الگوریتم A_1 و A_2 را در نظر داریم. A_1 فرضیه‌ای را انتخاب می‌کند که کمترین خطا را روی مجموعه دادگان داشته باشد و الگوریتم A_2 فرضیه دیگر را انتخاب می‌کند. فرض کنید که توزیع روی $\mathcal{X}$ به شکلی است که $\mathbb{P}(f(x)=+1)=p$.

آ) آیا A_1 می‌تواند فرضیه‌ای تولید کند که تضمین کند عملکرد بهتری از انتخاب تصادفی ($random$) برای نقاط خارج از $\mathcal{D}$ دارد؟

برای قسمت‌های بعدی سوال فرض کنید که مقدار تمام نمونه‌های داخل $\mathcal{D}$ برابر $+1$ است.

ب) آیا ممکن است که A_2 فرضیه‌ای بهتر از A_1 ایجاد کند؟ توضیح دهید.

پ) اگر $p=0.9$ باشد، احتمال اینکه A_1 فرضیه‌ای بهتر از A_2 ایجاد کند، چقدر است؟

ت) آیا مقداری از p وجود دارد که به احتمال زیاد A_2 بتواند فرضیه بهتری نسبت به A_1 تولید کند؟

پاینده باشید