23/10/20

שאלה 1

מטריצה A מעל $\mathbb{R}$ תקרא מטריצה אי-שלילית אם עבור כל וקטור v, מתקיים $v^TAv \geq 0$.

הראה כי מטריצה סימטרית A היא אי שלילית אמ"מ ניתן לרשום אותה כ $A=XX^T$.

כיוון 1 - תהא A מטריצה סימטרית מעל $\mathbb{R}$, כך שעבור כל וקטור v, מתקיים $v^TAv \geq 0$.

A סימטרית $\Rightarrow$ היא ניתנת ללכסון אוניטרי, כלומר, קיימת מטריצה אוניטרית Q כך ש

$$D = Q^{-1}AQ \overset{Q^* \text{ מעל } \mathbb{R} \text{ הם } Q^t}{=} Q^tAQ \quad , \quad D = diag\{\lambda_1,\dots,\lambda_n\}$$

כמו כן, מכך ש A אי שלילית, ניתן להסיק כי כל הע"ע שלה הם ממשיים ואי שליליים.

A ו-D הן מטריצות דומות ולכן $\lambda_1,\dots,\lambda_n$ הם מספרים ממשיים אי-שליליים.

נקבל מכך ש $D=(\sqrt{D})^2$, ויתר על כן, משום שהיא אלכסונית, $\sqrt{D}=\sqrt{D}^t$.

מתקיים:

$$A = QDQ^t = Q\sqrt{D}\sqrt{D}^tQ^t = (Q\sqrt{D})(Q\sqrt{D})^t$$

נסמן $X=Q\sqrt{D}$ ונקבל כי $\boxed{A=XX^t}$. ■

כיוון 2: נניח כי $A=XX^t$. יהי v וקטור.

$$(*) \quad v^TAv = v^TXX^Tv = (v^TX)(v^TX)^T$$

נסמן $y=X^Tv$. גם y הוא וקטור במ"פ הסטנדרטי ומתקיים

$$\|y\| \geq 0 \Rightarrow \sqrt{\langle y,y\rangle} \geq 0 \overset{\text{מעל } \mathbb{R}}{\Rightarrow} \langle y,y\rangle \geq 0 \overset{\text{מ"פ סטנדרטית}}{\Rightarrow} y^TIy \geq 0 \Rightarrow$$

$$\Rightarrow (X^Tv)^T(X^Tv) \geq 0 \Rightarrow (v^TX)(v^TX)^T \geq 0 \overset{(*)}{\Rightarrow} \boxed{v^TAv \geq 0} \quad ■$$

23/10/20

שאלה 2

הראה כי קבוצת כל המטריצות הסימטריות האי שליליות היא קמורה.

כלומר שעבור כל שתי מטריצות סימטריות אי שליליות A,B ולסקלר $0\le\theta\le1$

המטריצה $\theta A+(1-\theta)B$ היא גם מטריצה סימטרית אי-שלילית.

יהיו A,B שתי מטריצות סימטריות אי-שליליות ויהי $0\le\theta\le1$ סקלר.

נגדיר $C=\theta A+(1-\theta)B$, נוכיח כי C מטריצה סימטרית אי-שלילית.

$$[C]_{i,j}=\theta[A]_{i,j}+(1-\theta)[B]_{i,j}=\theta[A]_{j,i}+(1-\theta)[B]_{j,i}=[C]_{j,i}\Rightarrow$$ C מט' סימטרית

יהי v וקטור.

$$v^TCv=v^T(\theta A+(1-\theta)B)v=\left[(v^T\theta A)+v^T(1-\theta)B\right]v$$

$v\alpha X=\alpha vX$

$$=\left[\theta v^TA+(1-\theta)v^TB\right]v=\theta v^TAv+(1-\theta)v^TBv\geq0$$

$$\left.\begin{array}{l}\theta\geq0,\ v^TAv\geq0\Rightarrow\theta v^TAv\geq0\\ \theta\leq1\Rightarrow(1-\theta)\geq0,\ v^TBv\geq0\Rightarrow(1-\theta)v^TBv\geq0\end{array}\right\}$$

ולכן C היא מטריצה סימטרית אי שלילית.

23/10/20

שאלה 1

חשבון אינפיניטסימלי של מטריצות הוא הרחבה של אינפי למטריצות ווקטורים.

נגדיר את הנגזרת של סקלר y ביחס לוקטור x כוקטור ה-עמודה אשר מקיים

$$\left(\frac{dy}{dx}\right)_i = \frac{dy}{dx_i} \qquad \forall i=1,\dots,n.$$

תהא $A \in M_{n\times n}$, הוכח כי $\frac{d(x^TAx)}{dx} = (A+A^T)x$

$$A = \begin{bmatrix} a_{11},\dots, & a_{1n} \\ \vdots & \vdots \\ a_{n1},\dots, & a_{nn} \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$x^TAx = x^T \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j \\ \vdots \\ \sum_{j=1}^n a_{nj}x_j \end{bmatrix} = \sum_{i=1}^n x_i \sum_{j=1}^n a_{ij}x_j =$$

$$= \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_ix_j$$

$$\boxed{\left(\frac{d(x^TAx)}{dx}\right)_q} = \frac{d\sum_{i=1}^n \sum_{j=1}^n a_{ij}x_ix_j}{dx_q} =$$

$$= \frac{d\, a_{qq}x_q^2}{dx_q} + \sum_{j\in[1,n]/\{q\}} \frac{d\, a_{qj}x_qx_j}{dx_q} + \sum_{i\in[1,n]/\{q\}} \frac{d\, a_{iq}x_ix_q}{dx_q} =$$

$$= 2a_{qq}x_q + \sum_{j\in[1,n]/\{q\}} a_{qj}x_j + \sum_{i\in[1,n]/\{q\}} a_{iq}x_i =$$

$$= \sum_{j=1}^n a_{qj}x_j + \sum_{i=1}^n a_{iq}x_i = \boxed{\sum_{j=1}^n (a_{qj}+a_{jq})x_j}$$

$$(A+A^T) = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} + \begin{bmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{nn} \end{bmatrix} \qquad [A+A^T]_{i,j} = a_{ij}+a_{ji}$$

$$\boxed{((A+A^T)x)_q} = \sum_{j=1}^n [A+A^T]_{q,j}\cdot x_j = \boxed{\sum_{j=1}^n (a_{qj}+a_{jq})x_j} \quad \blacksquare$$

23/10/20

שאלה 2

תהא $p=(p_1,\dots,p_n)$ התפלגות דיסקרטית כאשר $\sum_{i=1}^{n} p_i=1$ ו- $p_i\geq 0$ $\forall i=1,\dots,n$

האנטרופיה, אשר מודדת את אי הודאות של ההתפלגות מוגדרת ע"י

$$H(p)=-\sum_{i=1}^{n} p_i\log p_i$$

כאשר מגדירים $0\log 0=0$. השתמשו בכופלי לגראנז' על מנת להראות שההתפלגות האחידה היא בעלת האנטרופיה הגדולה ביותר.

נגדיר פונ' $g:\mathbb{R}^n\to\mathbb{R}$ ע"י $g(p)=\sum_{i=1}^{n} p_i$

נרצה למצוא p שעבורו $H(p)=-\sum_{i=1}^{n} p_i\log p_i$ הוא מקסימלי תחת האילוץ $g(p)=1$.

יהא $\hat{p}$ וקטור ההתפלגות אותו אני מחפשים. נשתמש בכופלי לגראנז':

$$\mathcal{L}(p,\lambda)=H(p)-\lambda(g(p)-1)$$

ונרצה למצוא λ, p עבורם $\nabla\mathcal{L}=0$.

$$\forall q=1,\dots,n\quad (\nabla\mathcal{L})_q=\frac{d\mathcal{L}}{dp_q}=\frac{dH(p)}{dp_q}-\lambda\frac{dg(p)}{dp_q}=$$

$$=\frac{d\left(-\sum_{i=1}^{n} p_i\log p_i\right)}{dp_q}-\lambda\frac{d\left(\sum_{i=1}^{n} p_i\right)}{dp_q}=$$

$$=\frac{d(-p_q\log p_q)}{dp_q}-\lambda\frac{dp_q}{dp_q}=-\left(\log p_q+p_q\cdot\frac{1}{p_q}\right)-\lambda=-\log(p_q)-1-\lambda$$

$$(\nabla\mathcal{L})_{n+1}=\frac{d\mathcal{L}}{d\lambda}=\frac{d(H(p))}{d\lambda}-\frac{\lambda\, dg(p)}{d\lambda}=-\left(\sum_{i=1}^{n} p_i-1\right)$$

$$\begin{cases}-\log(p_i)-1-\lambda=0 & \forall i=1,\dots,n\\ \sum_{i=1}^{n} p_i=1\end{cases}$$

23/10/20

$$\boxed{\begin{cases} -\log_2(p_i) - 1 - \lambda = 0 \\ \sum_{i=1}^{n} p_i = 1 \end{cases}} \qquad \forall i = 1, \dots, n \Rightarrow p_1 = p_2 = \dots = p_n = \frac{1}{n}$$

$$\lambda = -\log_2\left(\frac{1}{n}\right) - 1 = \log_2(n) - 1 \quad , \quad \sum_{i=1}^{n} \frac{1}{n} = 1 \checkmark$$

מכאן נובע שהאנטרופיה מקסימלית כאשר ההתפלגות אחידה.

כמו כן, הפתרון מקיים את 4 התנאים על ההסתברויות, אך הוא מקיים את הקבוצה שהיא הצגנו,

(כאשר $\frac{1}{n} \geq 0$). הקבוצה מקיימת להגדרת האנטרופיה כמדד לאי ודאות של ניסוי

מקרי, **כאשר ההתפלגות היא אחידה, אי הודאות שלנו לגבי הקבוצה היא המקסימלית.**

מכאן זו ההתפלגות. ■

23/10/20

שאלה 3

יהיו $X_0,\dots,X_{n-1}$ n משתנים מקריים בלתי תלויים, חיוביים ($\forall i, X_i \in [0,\infty)$)

בעלי פונקציית התפלגות רציפה f_x. נוכיח טענה כי מתקיים

$$\mathbb{P}\left[X_0 \geq \max\{X_1,\dots,X_n\}\right] = \frac{1}{n}$$

נתחיל בהוכחת הלמה הבאה:

$$\mathbb{P}\left[X_0 \geq \max\{X_1,\dots,X_2\}\right] = \int_0^\infty \left(F_{X_0}(a)\right)^{n-1} \cdot f_{X_0}(a)\, da$$

כאשר $F_x(a)$ היא פונ' ההתפלגות המצטברת של X_0 בנקודה a

ו $f_x(a)$ היא פונ' הצפיפות של X_0 בנקודה a.

$$\mathbb{P}\left[X_0 \geq \max\{x_1,\dots,x_n\}\right] \overset{\text{הסתברות שלמה}}{=}$$

$$= \int_{a=0}^\infty f_{X_0}(a) \cdot \mathbb{P}(X_1 \leq a \mid X_0 = a) \cdot \mathbb{P}(X_2 \leq a \mid X_0 = a \wedge X_1 \leq a) \cdots$$

$$\overset{\text{ב"ת ושווי התפלגות}}{=} \int_{a=0}^\infty \left(\mathbb{P}(X_0 \leq a)\right)^{n-1} \cdot f_{X_0}(a)\, da \overset{\text{ע"פ הגדרה}}{=} \boxed{\int_0^\infty \left(F_{X_0}(a)\right)^{n-1} \cdot f_{X_0}(a)\, da}$$ ∎

כעת נשתמש בעובדה ש פונקציית ההתפלגות המצטברת היא האינטגרל של פונ' הצפיפות.

$$F_x(a) = \int_{-\infty}^{a} f_x(a)\, da \quad , \quad f_x(a) = F_x'(a)$$

$$\mathbb{P}\left[X_0 \geq \max\{x_1,\dots,x_n\}\right] = \int_{a=0}^\infty \left(F_{X_0}(a)\right)^{n-1} \cdot f_{X_0}(a)\, da =$$

$$= \int_{a=0}^\infty \left(F_{X_0}(a)\right)^{n-1} \cdot F_{X_0}'(a)\, da = \quad F_{X_0}(a) \overset{\text{נסמן}}{=} u \quad \left(\int_a^b f(x)dx = \int_\alpha^\beta f(\varphi(t)) \cdot \varphi'(t)\, dt\right)$$

$$= \int_{a=0}^\infty u^{n-1}\, du = \frac{u^n}{n}\Big|_0^\infty = \frac{1}{n} \cdot \left(F_{X_0}(a)\Big|_0^\infty\right) = \frac{1}{n}(1-0) = \boxed{\frac{1}{n}}$$ ∎

23/10/20

שאלה 1

יהיו X,Y משתנים מקריים כאשר Y מקבל את הערכים $\{1,\dots,L\}$

יהיה l_{0-1} פונ' הפסד המוגדרת ע"י $l_{0-1}(\hat{Y},Y)=\begin{cases}0 & \hat{Y}=Y\\ 1 & o.w\end{cases}$

הראה כי $h=\arg\min_{f:X\to\mathbb{R}} \mathbb{E}[l_{0-1}(f(X),Y)]$

נתונה ע"י $h(x)=\arg\max_{i\in\{1,\dots,L\}} \mathbb{P}(Y=i\mid X=x)$

$$L(h)=\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mathbb{P}(X=x,Y=y)\cdot l(x,h(x))$$

המסווג האופטימלי הוא $\arg\min_h L(h)$.

$$l(y,\hat{y})=\begin{cases}0 & y=\hat{y}\\ 1 & y\neq\hat{y}\end{cases} \Rightarrow L(h)=\mathbb{E}[l_{0-1}(Y,h(X))]=\mathbb{P}(Y\neq h(X))$$

לכן נרצה למצוא מסווג אופטימלי שבהינתן X ימציא הסתברות גבוהה ביותר ל $\hat{Y}=Y$.

עם בייס:

$$\arg\max_{y\in Y} \mathbb{P}(Y=y\mid X=x)=\arg\max_{y\in Y}\frac{\mathbb{P}(X=x\mid Y=y)\cdot\mathbb{P}(Y=y)}{\mathbb{P}(X=x)}$$

$\mathbb{P}(X=x)\geq 0$ לכן נוכל להתעלם מהמכנה כאשר אנו מחפשים את המסווג.

$$\arg\max_{y\in Y}\mathbb{P}(Y=y\mid X=x)\sim\arg\max_{y\in Y}\mathbb{P}(X=x\mid Y=y)\cdot\mathbb{P}(Y=y)$$

עבור $x_0\in\mathcal{X}$ רוצים להקטין את הביטוי

$$\sum_{y=1}^{L}\mathbb{P}(X=x_0,Y=y)\cdot l_{0-1}(y,h(x_0))=$$

$$=\mathbb{P}(X=x_0)\left[\sum_{y=1}^{L}\mathbb{P}(Y=y\mid X=x_0)\cdot l_{0-1}(y,h(x_0))\right)$$

לכן h^* האופטימאלי במקרה זה הוא זה שמצאנו

$$\boxed{h^*(x_0)=\arg\max_{y\in\{1,\dots,L\}}\mathbb{P}(Y=y\mid X=x_0)}$$

23/10/20

שאלה 2

יהא $X=(X_1,\dots,X_n)^T$ וקטור של משתנים מקריים.

נאמר כי X מתפלג גאוסיאנית במספר משתנים עם תוחלת $\mu\in\mathbb{R}^n$ ומטריצת שונות $n\times n$ חיובית למחצה Σ אם פונ' הצפיפות של ההתפלגות שלו נתונה ע"י

$$f(x:\mu,\Sigma)=\frac{1}{(2\pi)^{n/2}\,|\Sigma|^{1/2}}\cdot e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

כאשר $\mathbb{E}[X_i]=\mu_i$ ו- $\mathrm{cov}(X_i,X_j)=\Sigma_{i,j}$ לכל $i,j=1,\dots,n$. $X\sim N(\mu,\Sigma)$

בשאלה נרחיב את כלל ההחלטה שראינו בתרגול לגאוס ממימד אחד.

נניח כי הנתונים שלנו הם זוגות $\langle X,y\rangle$ כאשר $X\in\mathbb{R}^d$ ו- $y\in\{0,1\}$.

נשים לב כי מטריצת השונות בשתי ההתפלגויות $f_0(x)=f(X|Y=0)\sim N(\mu_0,\Sigma)$

זהה, אך וקטורי התוחלות שונים. $f_1(x)=f(X|Y=1)\sim N(\mu_1,\Sigma)$

בנוסף נתון כי $\mathbb{P}(Y=1)=p$

בהינתן נקודה X, אנו רוצים לסווגה כ $Y=0$ או $Y=1$, אמ"ם

$\mathbb{P}[Y=1|x]>\mathbb{P}[Y=0|X]$. נמצא תנאי פשוט יותר.

$$\mathbb{P}[Y=1|x]>\mathbb{P}[Y=0|X] \iff$$

$$\frac{f_X(x|Y=1)\cdot\mathbb{P}(Y=1)}{f_X(x)}>\frac{f_X(x|Y=0)\cdot\mathbb{P}(Y=0)}{f_X(x)} \iff$$

$$\frac{f_X(x|Y=1)}{f_X(x|Y=0)}>\frac{\mathbb{P}(Y=0)}{\mathbb{P}(Y=1)}$$

23/10/20

$$\frac{f_X(x|Y=1)}{f_X(x|Y=0)} > \frac{\mathbb{P}(Y=0)}{\mathbb{P}(Y=1)} \iff \frac{f(x,\mu_1,\Sigma)}{f(x,\mu_0,\Sigma)} > \frac{1-p}{p} \iff$$

$$\frac{e^{-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)}}{\cancel{(2\pi)^{n/2}|\Sigma|^{1/2}}} \cdot \frac{\cancel{(2\pi)^{n/2}|\Sigma|^{1/2}}}{e^{-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)}} > \frac{1-p}{p} \iff$$

$$e^{-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)+\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)} > \frac{1-p}{p} \iff$$

$$-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)+\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0) > \log\left(\frac{1-p}{p}\right) \iff$$

$$(x-\mu_0)^T\Sigma^{-1}(x-\mu_0) - (x-\mu_1)^T\Sigma^{-1}(x-\mu_1) > 2\log\left(\frac{1-p}{p}\right) \iff$$

$$\cancel{x^T\Sigma^{-1}x} - x^T\Sigma^{-1}\mu_0 - \mu_0^T\Sigma^{-1}x + \mu_0^T\Sigma^{-1}\mu_0$$

$$-\left[\cancel{x^T\Sigma^{-1}x} - x^T\Sigma^{-1}\mu_1 - \mu_1^T\Sigma^{-1}x + \mu_1^T\Sigma^{-1}\mu_1\right] > 2\log\left(\frac{1-p}{p}\right) \iff$$

$$-x^T\Sigma^{-1}\mu_0 - \mu_0^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu_1 + \mu_1^T\Sigma^{-1}x > 2\log\left(\frac{1-p}{p}\right) - \mu_0^T\Sigma^{-1}\mu_0 + \mu_1^T\Sigma^{-1}\mu_1 \iff$$

$$x^T\left(\Sigma^{-1}(\mu_1-\mu_0)\right) + \left((\mu_1^T-\mu_0^T)\Sigma^{-1}\right)x > 2\log\left(\frac{1-p}{p}\right) - \mu_0^T\Sigma^{-1}\mu_0 + \mu_1^T\Sigma^{-1}\mu_1 \iff$$

סקלר - גם של 1x1, לכן $M^T = M$

$$x^T\left(\Sigma^{-1}(\mu_1-\mu_0)\right) + \overbrace{\left(x^T\left(\Sigma^{-1^T}(\mu_1-\mu_0)\right)\right)}^{}{}^T > 2\log\left(\frac{1-p}{p}\right) - \mu_0^T\Sigma^{-1}\mu_0 + \mu_1^T\Sigma^{-1}\mu_1 \iff$$

Cov גם
כי סימטרית

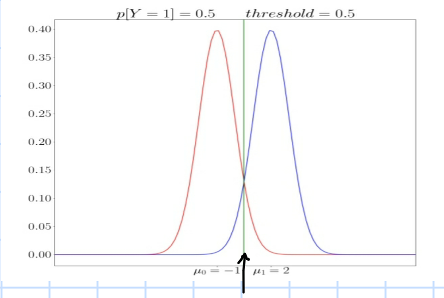
$$\boxed{2x^T\left(\Sigma^{-1}(\mu_1-\mu_0)\right) > 2\log\left(\frac{1-p}{p}\right) - \mu_0^T\Sigma^{-1}\mu_0 + \mu_1^T\Sigma^{-1}\mu_1}$$

24/10/20

$$\mathbb{P}[Y=1|X]=\mathbb{P}[Y=0|X]$$

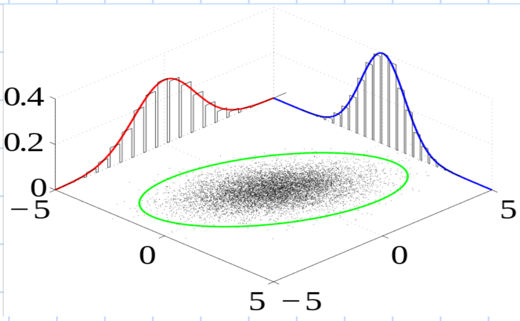
חסם ההחלטה לבייס זה מוגדר ע"י קבוצת נקודות שבהן

מה הצורה הכללית של החסם הכללי (כאשר $d>1$) ?



בהתחלה ראינו את המקרה עבור $d=1$.

במקרה זה וקטור החסם הוא וקטור ממימד 0 (סקלר) שהוא נקודת המפגש של ההתפלגויות במימד זה.



עבור המקרה בו $d=2$ ההתפלגות של Y כאשר המתפלגים נורמליים נראית כך :

כלומר בכל אחד מן המימדים ישנה התפלגות נורמלית עם שונויות שוות ותוחלות ממוצעות המקבילה למימד.

תכן בהגענו לקבל חסם החלטה אנו במימד 2 נשווה בין 2 "כיפות" שכאלו. ונמצא חסם ממימד 1.

עבור מימד d נרצה למצוא את כל ה V אשר מקיימים

$$V=\{v\in\mathbb{R}^n \mid \mathbb{P}(Y=1|v)-\mathbb{P}(Y=0|v)=0\}$$

מתרגיל 4 א' :

$$V=\left\{v\in\mathbb{R}^n \;\middle|\; 2v^T\underbrace{\left(\Sigma^{-1}(\mu_1-\mu_0)\right)}_{u\in\mathbb{R}^n}=\underbrace{2\log_2\left(\frac{1-p}{p}\right)-\mu_0^T\Sigma^{-1}\mu_0+\mu_1^T\Sigma^{-1}\mu_1}_{c\in\mathbb{R}}\right\}$$

נסמן

$$V=\left\{v\in\mathbb{R}^n \;\middle|\; v^Tu=\frac{c}{2}\right\}$$

V הוא תת מרחב ממימד $d-1$, כלומר על מרחב, אשר מפריד בין שני תתי-מרחבים.

24/10/20

שאלה 3

יהיו $X_1, X_2, \ldots, X_n$ משתנים מקריים בלתי תלויים בעלי התפלגות אחידה בקטע $[-3, 5]$.

נגדיר $S = \sum_{i=1}^{n} X_i$

השתמשו באי-שוויון הופדינג על מנת למצוא $N \in \mathbb{N}$ כך שעבור כל $n > N$

$$P(S > n^2 + 0.2n) < 0.1$$

נתונים n מ"מ בת"ל כך ש $P[-3 \le X_i \le 5] = 1$.

עבור התפלגות אחידה מתקיים

$$\bar{X} = \sum_{i=1}^{n} X_i / n = \frac{S}{n} \Rightarrow S = n\bar{X}$$

$$E[\bar{X}] = E\left[\frac{S}{n}\right] = \frac{1}{n} E[S] = \frac{n}{n} = 1$$

עפ"י א' שוויון הופדינג מתקיים

$$P(|\bar{X} - E[\bar{X}]| \ge \varepsilon) \le 2e^{-\frac{2n\varepsilon^2}{(b-a)^2}}$$

$$P(S > n^2 + 0.2n) = P(n\bar{X} > n^2 + 0.2n) \overset{n>0}{=} P(\bar{X} > n + 0.2) = P(\bar{X} - 0.2 > n)$$

$$P(\bar{X} - 1 > n - 0.8) = P(\bar{X} - E[\bar{X}] > n - 0.8) \le P(|X - E[\bar{X}]| > n - 0.8) \overset{\text{רוצים}}{<} 0.1 \Rightarrow$$

מספיק N עבורו

$$2e^{-\frac{2N(N-0.8)^2}{(5-(-3))^2}} \le 0.1$$

$$e^{-\frac{N(N-0.8)^2}{64}} \le 0.05$$

$$-N(N-0.8)^2 \le 64 \log(0.05)$$

$$N^3 - 1.6N^2 + 0.64N \ge -64 \log(0.05) \ge 64 \cdot 1.5 = 96$$

עבור $N=5$ מתקיים $125 - 40 + 3.2 = 88.2 < 96$

עבור $N=6$ מתקיים $216 - 57.6 + 3.84 > 156 > 96$

סוף $\boxed{N=6}$ חסם המקיים את המבוקש.

24/10/20

שאלה 4

נניח כי אנו מעוניינים לבנות מי איזון אשר מנפיק n משימות ל m שרתים.

נניח כי למשימה ה j נדרשים $0 \le L_j \le 1$ יחידות זמן.

המטרה היא לשייך את n המשימות ל m השרתים כך שהן יהיו מאוזנות באופן מקסימלי.

כל שרת עובד בנפרד על משימותיו וזמן הסיום שלו הוא סכום הזמנים של משימותיו.

נגדיר $L = \sum_{j=1}^{n} L_j$

עם איזון מושלם זמן הסיום של כל שרת הוא $\frac{L}{m}$. ישנם אלגוריתמים טובים לקירוב לכך, אך היום רוצים לנתח את הביצועים של מתן המשימות באופן אקראי.

נניח כי אנו נותנים משימות באופן אקראי לכל משימה, עם חזרות.

נגדיר:
$$R_{i,j} = \begin{cases} L_j & \text{המשימה ה } j \text{ ניתנה לשרת ה } i \\ 0 & \text{אחרת} \end{cases}$$

$$R_i = \sum_{j=1}^{n} R_{i,j} \quad (\text{סך העומס על השרת ה } i)$$

מהי התוחלת של R_i ?

תוחלת של קבוע

$$\sum_{i=1}^{m} R_i = L \Rightarrow$$

לינאריות תוחלת קיימת גם עבור מ״מ תלויים

$$\mathbb{E}\left[\sum_{i=1}^{m} R_i\right] = \mathbb{E}[L] = L \Rightarrow$$

סימטריה בין השרתים

$$\sum_{i=1}^{m} \mathbb{E}[R_i] = L \Rightarrow$$

$$\boxed{\mathbb{E}[R_i] = \frac{L}{m}}$$

24/10/20

$$\mathbb{E}[R_i] = \frac{L}{m}$$

אנו רוצים לחסום את ההסתברות שהעומס על השרת ה i גדול ב $\delta = 10\%$ מאשר

תוחלת העומס. השתמש בצורת המכפלה של אי-שוויון צ'רנוף B ממנו לחסום אותו

$$\mathbb{P}\left(R_i \geq (1+\delta)\mathbb{E}[R_i]\right)$$

$$R_i = \sum_{j=1}^{n} R_{ij} \qquad R_{i,j} : i \times j \to [0,1]$$

$$\mathbb{E}[R_i] = \frac{L}{m}$$

מתקיימים התנאים המקדימים לאי שוויון

נקבל:

$$\mathbb{P}\left(R_i > 1.1\,\mathbb{E}[R_i]\right) \leq \left(\frac{e^{0.1}}{1.1^{1.1}}\right)^{\frac{L}{m}} < \boxed{0.9925^{L/m}}$$

ג. כעת אנו רוצים לחסום את ההסתברות שקיים שרת כלשהו אשר גדול עומס הגדול

ביותר מ $\delta = 10\%$ מהתוחלת. נתן חסם

$$\mathbb{P}\left(R_1 > (1+\delta)\mathbb{E}[R_1] \vee \dots \vee R_m > (1+\delta)\mathbb{E}[R_m]\right)$$

א"ש + סימטריה

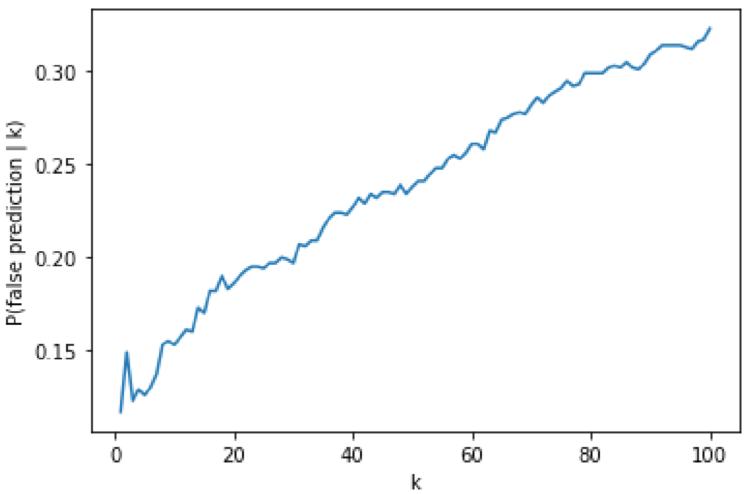
$$\mathbb{P}\left(\bigcup_{i=1}^{m} R_i > (1+\delta)\mathbb{E}[R_i]\right) \leq \sum_{i=1}^{m} \mathbb{P}\left(R_i > (1+\delta)\mathbb{E}[R_i]\right) =$$

$$m \cdot \left(\frac{e^{0.1}}{1.1^{1.1}}\right)^{\frac{L}{m}} < \boxed{m(0.9925)^{L/m}}$$

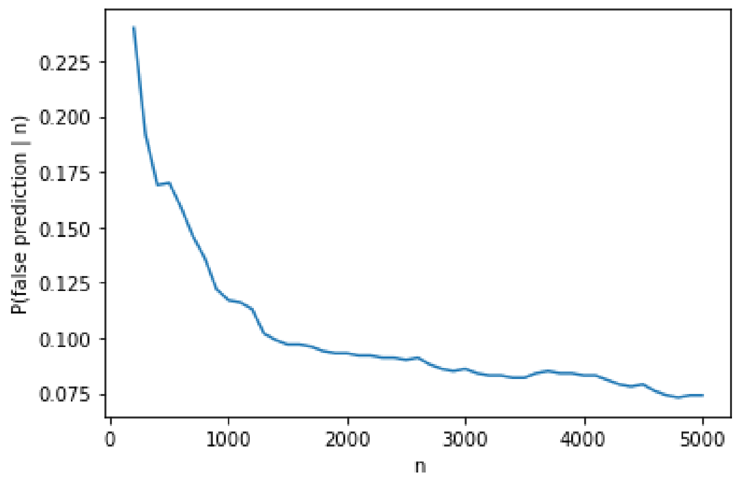
29/10/20

תשובות מילוליות לחלק התכנות

ב. הדיוק של המסווג בהרצת כל תמונות הבדיקה על 1000 תמונות האימון הראשונות הוא 0.847. ממסווג אקראי הייתי מצפה לדיוק של 0.1, שכן בהינתן תמונה, היא בעלת label ℓ, ובאופן אקראי מבין 10 הלייבלים האפשריים, ההסתברות לסיווג נכון הוא $\frac{1}{10}$.



ג. לפניכן גרף המתאר את אחוז השגיאה בהינתן K, כאשר ערכי K מתוארים בציר ה-X. ה-K הטוב ביותר הוא 1, ובאופן כללי, ככל שמגדילים את K כך דיוק המסווג פוחת. הסבר לכך לדעתי היא שבשיטה זו לא נמדדים קשרים בין פיקסלים שונים באותה התמונה אלא מחפשים תמונה בה סכום המרחק של כל פיקסל מהפיקסל המתאים לו היא מינימלית.



ד. לפניכן גרף המתאר את ההסתברות לשגיאה כפונקציה של h - גודל קבוצת הניסוי, ביחס ל $k=1$. ניתן לראות שהגרף יורד בשיפוע חזק אך מתייצב במגמת ירידה מתונה, ונובע משום שעבור h מסוים חסר דוגמאות $p\in P$, אין תמונה מספיק קרובה ב S הנותנת בוודאות גבוהה את הלייבל הנכון. ככל ש h גדל יש אפשרות למצוא תמונה בעלת מאפיינים יותר קרובים ובכך לקבל דיוק גבוה יותר.