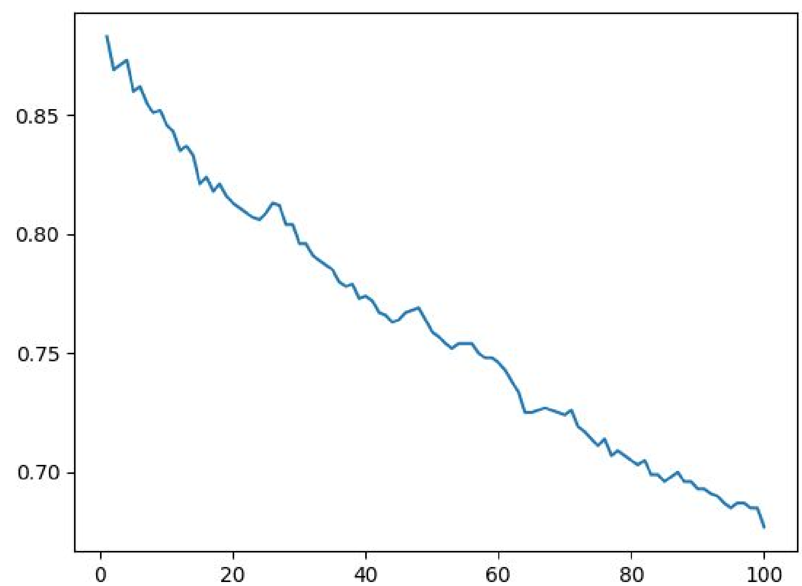
Programming assignment

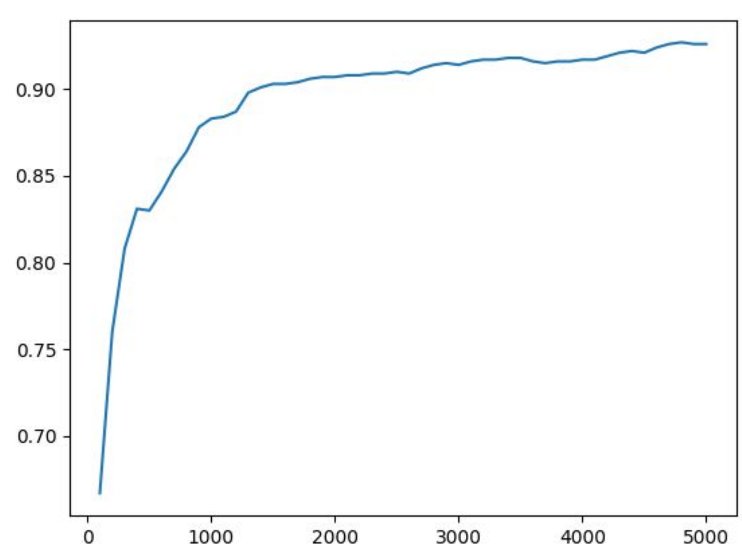
a. Submitted separately

b. The accuracy is 0.846. In a random classifier I'd expect to get an accuracy of 0.1, the chance of choosing the right label out of 10 possible ones under a uniform distribution.

c. As you can see in the following chart, the optimal k is 1 (x-axis: k, y-axis: accuracy):



d. As you can see in the following chart the more training images the more accurate the model is. This correlation is somewhat logarithmic:



מבוא ללמידה חישובית - תרגיל בית 1

Decision Rules and Concentration Bounds

(1 נניח שלא $h\in D\to\{1,\dots,L\}$ כאשר D הוא תחום
סופי כך ש- $P(X\in D)=1$.
נניח בשלילה שקיים $\hat{x}$ עבורו $h(\hat{x})\notin\{1,\dots,L\}$ ונגדיר פונקציה
$h'(x)$: $h'(x)=\begin{cases}\arg\max_{j\in\{1,\dots,L\}} P[Y=j|X=\hat{x}], & x=\hat{x}\\ h(x), & \text{else}\end{cases}$ ברור שלא "הפסקנו"

לכן, נראה $E(l_{0-1}(h(X),Y))\geq E(l_{0-1}(h'(X),Y))$

לכן נקבל ש- $h(x)\in\{1,\dots,L\}$ לכל x. מכאן נתחיל:

$$E(l_{0-1}(f(X),Y))=\sum_{\hat{x}\in D}\sum_{j=1}^{L}P[X=\hat{x},Y=j]\cdot l_{0-1}(f(\hat{x}),j)=$$

$$=\sum_{\hat{x}\in D}\sum_{j=1}^{L}P(X=\hat{x})\cdot P(Y=j|X=\hat{x})\cdot l_{0-1}(f(\hat{x}),j)=$$

$$=\sum_{\hat{x}\in D}P(X=\hat{x})\cdot\overbrace{\sum_{j=1}^{L}P(Y=j|X=\hat{x})\cdot l_{0-1}(f(\hat{x}),j)}^{\circledast}$$

לכן מספיק למזער את $\circledast$ עבור $\hat{x}$ כלשהו (קבוע), נניח ש- $f(\hat{x})\in\{1,\dots,L\}$

$$P(X=\hat{x})\cdot\sum_{j=1}^{L}P(Y=j|X=\hat{x})\cdot l_{0-1}(f(\hat{x}),j)=$$

$$=P(X=\hat{x})\cdot\sum_{\substack{j=1\\ j\neq k}}^{L}P(Y=j|X=\hat{x}),\quad f(\hat{x})=k\in\{1,\dots,L\}$$

כלומר נרצה למזער את הסכום ש- $P(Y=k|X=\hat{x})$ יהיה
מקסימלי כאשר $f(\hat{x})=k$, לכן: $h(\hat{x})=\arg\max_{j\in\{1,\dots,L\}}P(Y=j|X=\hat{x})$
כנדרש.

(3

$$P(S>n^2+0.2n)=P(\bar{X}>n+0.2)=P(\bar{X}-1>n-0.8)\leq$$

$$\leq P(|\bar{X}-1|>n-0.8)\leq$$

נשים לב ש- $X_i\in[-3,5]$ בהתפלגות אחידה - $E(X_i)=1$ לכן גם
$E(\bar{X})=1$. אז נוכל להשתמש באי"ש הופדינג:

$$\leq 2e^{-\frac{2n(n-0.8)^2}{64}}=2e^{-\frac{1}{32}(n^3-1.6n^2+0.64n)}$$

נדרוש שהביטוי הנ"ל יהיה קטן מ-0.1 ונקבל: $(\ln(0.05)\approx -3)$

$$-\frac{1}{32}[n^3-1.6n^2+0.64n]<-3\Rightarrow n^3-1.6n^2+0.64n-96>0$$

בדקתי עבור $n=5$ הא"ש לא מתקיים ועבור $n=6$ כן, לכן $N=6$

2)

א. לפי ב"ס מסווג להראות:

$\frac{f_X(\hat{x}|y=1)\cdot P(y=1)}{f_X(\hat{x})} > \frac{f_X(\hat{x}|y=0)\cdot P(y=0)}{f_X(\hat{x})}$

כלומר: $f_X(\hat{x}|y=1)\cdot p > f_X(\hat{x}|y=0)(1-p)$

נציב את f_X:

$$\frac{\exp\left(-\frac{1}{2}(\hat{x}-\mu_1)^t\Sigma^{-1}(x-\mu_1)\right)\cdot p}{2\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} > \frac{\exp\left(-\frac{1}{2}(\hat{x}-\mu_0)^t\Sigma^{-1}(x-\mu_0)\right)}{2\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}(1-p)$$

$$\Downarrow$$

$$-\frac{1}{2}\left[(\hat{x}-\mu_1)^t\Sigma^{-1}(\hat{x}-\mu_1)-(\hat{x}-\mu_0)^t\Sigma^{-1}(x-\mu_0)\right] > \ln\left(\frac{1-p}{p}\right)$$

$$\Downarrow$$

$$\hat{x}^t\Sigma^{-1}\hat{x}-\underline{\hat{x}^t\Sigma^{-1}\mu_1}-\underline{\underline{\mu_1\Sigma^{-1}\hat{x}}}+\mu_1\Sigma^{-1}\mu_1-\hat{x}^t\Sigma^{-1}\hat{x}+\underline{\hat{x}^t\Sigma^{-1}\mu_0}+\underline{\underline{\mu_0\Sigma^{-1}\hat{x}}}-\mu_0\Sigma^{-1}\mu_0 < 2\ln\left(\frac{1-p}{p}\right)$$

$$\Downarrow$$

$$\hat{x}^t\Sigma^{-1}(\mu_0-\mu_1)+\Sigma^{-1}\hat{x}(\mu_0-\mu_1)+\mu_1\Sigma^{-1}\mu_1-\mu_0\Sigma^{-1}\mu_0 < 2\ln\left(\frac{1-p}{p}\right)$$

נשים לב ש- $\hat{x}^t\Sigma^{-1}=\Sigma^{-1}\hat{x}$ כי Σ סימטרית ולכן גם Σ^{-1}.

לכן סה"כ נקבל:

$$\hat{x}^t\Sigma^{-1}(\mu_0-\mu_1) < \ln\left(\frac{1-p}{p}\right)-\mu_1\Sigma^{-1}\mu_1+\mu_0\Sigma^{-1}\mu_0$$

ב. עבור $d=1$ נקבל ש- $\Sigma=\text{var}(X)$, כלומר:

$$\hat{x}\,\text{var}(x)(\mu_0-\mu_1) < \ln\left(\frac{1-p}{p}\right)+(\mu_0^2-\mu_1^2)\text{var}(x)$$

$$\Downarrow$$

$$\hat{x} < \frac{\ln\left(\frac{1-p}{p}\right)}{\text{var}(x)(\mu_0-\mu_1)}+\mu_0+\mu_1$$

(קיבלנו שבדרך ה-bound הוא (בעצם) נקודה על ציר ה-X.

באופן דומה עבור $d=2$ כאשר $\hat{x}=\begin{pmatrix}x_1\\x_2\end{pmatrix}$

נקבל שה-bound הוא פונקציה של x_1,x_2

(כלומר פונקציה מ- $\{0,1\}\to\mathbb{R}^2$) לכן נקבל צורה

של קו ישר ב- $\mathbb{R}^2$, שמחלק את המישור.

לסיכום בצורת ההכללה במימד d הוא יהיה hyperplane,

מרחב מסדר במימד $d-1$ שמחלק את המרחב.

ד. (4

$$E(R_i)=E\left(\sum_{j=1}^{n}R_{ij}\right)=\sum_{j=1}^{n}E(R_{ij})=\sum_{j=1}^{n}L_j\cdot\frac{1}{m}=\frac{1}{m}\sum_{j=1}^{n}L_j=\frac{L}{m}$$

ה. כאשר $\delta=\frac{1}{10}$, $E(R_i)=\frac{L}{m}$ (ג׳):

$$P(R_i\ge(1+\delta)E(R_i))=P\left(R_i\ge\left(1+\tfrac{1}{10}\right)\tfrac{L}{m}\right)\le$$

לפי צ׳רנוף:

$$\le\left(\frac{e^{\frac{1}{10}}}{(1+\frac{1}{10})^{(1+\frac{1}{10})}}\right)^{\frac{L}{m}}=\left(\frac{1.105}{1.110}\right)^{\frac{L}{m}}=0.995^{\frac{L}{m}}$$

ו. לפי חסם האיחוד:

$$P\left(\bigvee_{i=1}^{m}P(R_i\ge(1+\delta)E(R_i))\right)\le\sum_{i=1}^{m}P(R_i\ge(1+\delta)E(R_i))=$$

$$=\sum_{i=1}^{m}0.995^{\frac{L}{m}}=m\cdot0.995^{\frac{L}{m}}$$

אלגברה לינארית

א) כיוון 1 - נניח שלכל וקטור v מתקיים $v^tAv\ge0$. אז

לפי הרמז יש Q,D כך ש- $v^tQ^tDQv\ge0$ לכל v.

כלומר איברי D, הע"ע של A הם אי-שליליים

לכן קיימת מטריצה E כך ש- $E^2=D$. אז:

$$A=Q^tDQ=Q^tE^2Q=Q^tE^tEQ=(QE)^t\cdot(QE)$$

כנדרש. נניח $A=X^tX$, ($A_{n\times n}$)

כיוון 2 - נסמן את האיבר ה-i ב- v^tAv:

$$(v^tX^t)_i=\sum_{j=1}^{n}v_jX^t_{ji},\quad (Xv)_i=\sum_{j=1}^{n}v_jX_{ij}$$

$$\Downarrow$$

$$(v^tAv)_i=\sum_{j=1}^{n}v_j^2X^t_{ji}\cdot X_{ij}=\sum_{j=1}^{n}(v_j\cdot X_{ij})^2\ge0$$

כנדרש.

ב) יהי וקטור v ו- A,B,θ כפי שהוגדר. אז:

$$v^t(\theta A+(1-\theta)B)v=v^t\theta Av+v^t(1-\theta)Bv$$

כאשר ברור ש- $v^tAv\ge0$ גורר ש- $v^t\theta Av\ge0$ ובנ"ל עבור

B ו- $1-\theta$ כי $0\le\theta\le1$.

בנוסף - $[\theta A+(1-\theta)B]^t=\theta A^t+(1-\theta)B^t\overset{*}{=}\theta A+(1-\theta)B$

כלומר קיבלנו מט׳ סימטרית כנדרש

* - נובע מסימטריות של A ו-B.

חלק א' הוכחת נגזרות

1) נשתמש בהגדרה של נגזרת - $\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$

כאשר $f(x)=x^tAx$:

$$f(x+h)=(x+h)^tA(x+h)=x^tAx+h^tAx+x^tAh+h^tAh=$$
$$=x^tAx+h^tAx+h^tA^tx+h^tAh$$

לכן:

$$f(x+h)-f(x)=\cancel{x^tAx}+h^tAx+h^tA^tx+h^tAh-\cancel{x^tAx}=$$
$$=h^t(A+A^t)x+h^tAh$$

סה"כ גם נגזרת חלקית של נחלק ב-h

ונקח גבול נקבל ש- h^tAh שואף ל-0 מהר יותר ה

מ-h ולכן יתאפס ונשאר (מגיע) רק $(A+A^t)x$ כנדרש.

---

2) נגדיר $g(p)=(\sum_{i=1}^{n}p_i)-1$ ו-λ כופל לגרנז'

לכן, נגדיר פונקציה h אותה נרצה למקסם:

$$h(p,\lambda)=-\sum_{i=1}^{n}p_i\ln(p_i)+\lambda g(p)=-\sum_{i=1}^{n}p_i\ln(p_i)+\lambda\left[\sum_{i=1}^{n}p_i-1\right]$$

נגזור ונשווה ל-0:

$\forall 1\le i\le n$:

$$\left(\frac{\partial h(p,\lambda)}{\partial p}\right)_i=-(\ln p_i+1)+\lambda=0\Rightarrow \ln p_i=\lambda-1$$

ונקבל ש- $\ln p_i=\ln p_j$ לכל $1\le i,j\le n$ ולכן

$p_i=p_j$ כלומר התפלגות אחידה.

ג) א. נגדיר $\hat{X}=\max_{i=1}^{n-1}(X_i)$ ונקבל:

$$F_{\hat{X}}(\alpha)=P(\hat{X}\le\alpha)=$$
$$=P\left(\bigcap_{i=1}^{n-1}X_i\le\alpha\right)=$$
$$=\prod_{i=1}^{n-1}P(X_i\le\alpha)=\left(F_X(\alpha)\right)^{n-1}$$

נגזור כדי לקבל את $f_{\hat{X}}(\alpha)$:

$$f_{\hat{X}}(\alpha)=(n-1)\left(F_X(\alpha)\right)^{n-1}\cdot f_X(\alpha)$$

נחשב את ההסתברות שהאיבר הנבחר $X_0\ge\max_{i=1}^{n-1}(X_i)$

כלומר $X_0\ge\hat{X}$:

$$P(X_0\ge\hat{X})=\int_0^\infty\int_0^{x_0}f_{X_0}(x)\cdot f_{\hat{X}}(\hat{x})\,dx\,d\hat{x}=$$

$$=\int_0^\infty f_{X_0}(x)\cdot\int_0^{x_0}f_{\hat{X}}(\hat{x})\,dx\,d\hat{x}=$$

$$=\int_0^\infty f_{X_0}(x)\cdot F_{\hat{X}}(\hat{x})\,d\hat{x}=\int_0^\infty f_{X_0}(\hat{x})\cdot F_{X_0}(\hat{x})^{n-1}\,d\hat{x}$$

[illegible]

ב. נבצע החלפת משתנים – $F_X(x)=t$, $f_X(x)=dt$

ונקבל:

$$\int_0^\infty f_X(x)\cdot F_X(x)^{n-1}\,dx=\int_0^\infty t^{n-1}\,dt=\lim_{m\to\infty}\frac{t^n}{n}\Big|_0^m=$$

$$=\frac{1}{n}\lim_{m\to\infty}F_X(x)^n\Big|_0^m=\frac{1}{n}\left[\lim_{m\to\infty}F_X(m)-F_X(0)\right]=\frac{1}{n}[1-0]=\frac{1}{n}$$

כנדרש.