ALAKFELISMERÉS—HÁZI FELADAT PÉLDÁK

1. Feladat Bizonyítsd be, hogy ha X egy nemnegatív valószínűségi változó , akkor

$$\mathbf{E}X = \int_0^\infty \mathbf{P}\{X > t\}dt.$$

2. Feladat Legyenek $X_1, X_2, \ldots$ korlátos valószínűségi változók , amelyekre $\mathbf{P}\{|X_n| \leq M\} = 1$ valamely $M > 0$ számra. Tegyük fel, hogy valamely X valószínűségi változóra, $\lim_{n\to\infty} X_n = X$ valószínűségben. Bizonyítsd be, hogy ekkor $\mathbf{P}\{|X| \leq M\} = 1$.

3. Feladat Bizonyítsd be, hogy ha $X \in \mathcal{R}^d$ és $Y \in \{0,1\}$ független valószínűségi változók , akkor $L^* = \min(p, 1-p)$, ahol $p = \mathbf{P}\{Y = 1\}$ és L^* a Bayes hiba. Mutass egy peldát amikor X és Y nem függetlenek, és mégis $L^* = \min(p, 1-p)$.

4. Feladat Bizonyítsd be, hogy ha egy X valószínűségi változó olyan, hogy $\mathbf{P}\{X \in [a,b]\} = 1$, akkor $\mathbf{Var}\{X\} \leq (b-a)^2/4$.

5. Feladat Legyen $X' = (X_1, X_2)$ és jelölje $L^*_{X_1}$ az X_1 megfigyelésére alapozott, az Y bináris valószínűségi változóra adott Bayes döntés hibáját. Legyen $L^*_{X'}$ hasonlóan, az X_1 és X_2 együttes megfigyelésére alapozott Bayes hiba. Bizonyítsd be, hogy $L^*_{X_1} \geq L^*_{X'}$, illetve, ha X_2 független (X_1, Y)-tól, akkor $L^*_{X_1} = L^*_{X'}$.

6. Feladat (DÖNTÉS VISSZAUTASÍTÁSSAL.) Némely döntési feladatban megengedett, hogy azt mondjuk, hogy “nem tudom”, ha ez nem túl gyakran történik. Formálisan, egy döntésfüggvény három értéket vehet fel: 0, 1 és “nem tudom”. Ekkor két valószínűség méri a döntés jóságát: a visszautasítás valószínűsége $\mathbf{P}\{g(X) = \text{“nem tudom”}\}$, illetve a hibavalószínűség $\mathbf{P}\{g(X) \neq Y | g(X) \neq \text{“nem tudom”}\}$. Egy $0 < c < 1/2$ számra definiáljuk a következő döntést:

$$g_c(x) = \begin{cases} 1 & \text{ha } \eta(x) > 1/2 + c \\ 0 & \text{ha } \eta(x) \leq 1/2 - c \\ \text{“nem tudom”} & \text{egyébként.} \end{cases}$$

Mutasd meg, hogy ez a döntés optimális olyan értelemben, hogy ha bármely más g döntésre igaz, hogy

$$\mathbf{P}\{g(X) = \text{“nem tudom”}\} \leq \mathbf{P}\{g_c(X) = \text{“nem tudom”}\},$$

akkor

$$\mathbf{P}\{g(X) \neq Y | g(X) \neq \text{“nem tudom”}\} \geq \mathbf{P}\{g_c(X) \neq Y | g_c(X) \neq \text{“nem tudom”}\}.$$

Más szóval, a visszautasítás valószínűségét egy bizonyos szint alatt tartó döntések közül a minimális hibavalószínűségű döntés g_c alakú (valamilyen c-re).

7. Feladat Mutasd meg, hogy bármely két döntésre $|L(g_1) - L(g_2)| \leq \mathbf{P}\{g_1(X) \neq g_2(X)\}$. Mikor áll fenn egyenlőség?

ALAKFELISMERÉS—HÁZI FELADAT PÉLDÁK

8. Feladat Tegyük fel, hogy az X valószínűségi változó olyan, hogy minden $\epsilon > 0$-ra

$$\mathbf{P}\{X > \epsilon\} \leq ce^{-n\epsilon}.$$

Mutasd meg, hogy

$$\mathbf{E}X \leq \frac{\log c + 1}{n}.$$

9. Feladat Legyen $X \in \{0,1\}^d$, és tegyük fel, hogy $Y = g(X)$, ahol g egy Boole *monom* (egytagú polinom), azaz

$$g(x) = x^{(i_1)}x^{(i_2)}\cdots x^{(i_k)}$$

valamely $i_1, \ldots, i_k \leq d$ indexekre, ahol $x^{(i)}$ az x vektor i-edik bitjét jelöli. Vagyis $g(x)$ értéke pontosan akkor egy, ha az x vektor $i_1, \ldots, i_k$-adik bitjeinek értéke mind egy. A tanítóadat n darab véletlenül sorsolt X_i vektorból, és a hozzájuk tartozó $Y_i = g(X_i)$-ből áll. Legyen g_n az a monom, amelyet definiáló i_j koordináták éppen azok, amely pozícióban minden olyan X_i vektor, amelyre $Y_i = 1$, egyes. Az órán tanult tétel felhasználásával mutasd meg, hogy

$$\mathbf{P}\{L(g_n) > \epsilon\} \leq 2^d e^{-n\epsilon}, \quad \text{és} \quad \mathbf{E}L(g_n) \leq \frac{d\log 2 + 1}{n}.$$

10. Feladat Legyen B egy binomiális eloszlású valószínűségi változó n és p paraméterekkel, azaz $B = \sum_{i=1}^n I_{A_i}$, ahol $A_1, \ldots, A_n$ független események, és $\mathbf{P}\{A_i\} = p$ minden i-re. Mutasd meg a Chernoff technika felhasználásával, hogy bármely $k \geq np = m$ számra

$$\mathbf{P}\{B > k\} \leq \left(\frac{m}{k}\right)^k e^{k-m}.$$

Útmutatás: $\mathbf{E}\left[e^{sI_{A_i}}\right] \leq e^{p(e^s-1)}$. (Miért?)

11. Feladat Terjeszd ki az előző feladatot állítását tetszőleges $[0,1]$-beli független valószínűségi változók összegére. Pontosabban, legyen $B = \sum_{i=1}^n X_i$, ahol $X_i \in [0,1]$, és $\sum_{i=1}^n p_i = m$, ahol $p_i = \mathbf{E}X_i$. Ekkor $k \geq m$ esetén

$$\mathbf{P}\{B > k\} \leq \left(\frac{m}{k}\right)^k e^{k-m}.$$

Útmutatás: Hasonlóan a Hoeffding egyenlőtlenség bizonyításához, használd az e^{sx} függvény konvexitását.

12. Feladat Legyen most B egy m paraméterű Poisson eloszlású valószínűségi változó , azaz $\mathbf{P}\{B = i\} = m^i e^{-m}/i!$, ha $i = 0, 1, 2, \ldots$. Mutasd meg a Chernoff technika segítségével, hogy

$$\mathbf{P}\{B > k\} \leq \left(\frac{m}{k}\right)^k e^{k-m}.$$

Ez a korlát akkor érdekes, ha visszaemlékszel, hogy hogyan lehet a binomiális eloszlást Poissonnal közelíteni. *Útmutatás:* Itt a Chernoff felső korlát pontosan minimalizálható s-ben. Használd fel, hogy $\sum_{i=0}^{\infty} x^i/i! = e^x$.

13. Feladat Tegyük fel, hogy az X valószínűségi változó feltételes eloszlása mind az $\{Y=0\}$, mind az $\{Y=1\}$ feltétel mellett d dimenziós normális, azaz a két feltételes sűrűségfüggvény a következő alakú:

$$f_i(x) = f(x|Y=i) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma_i}} e^{-\frac{1}{2}(x-m_i)^T \Sigma_i^{-1}(x-m_i)}, \qquad i=0,1,$$

ahol $m_i = \mathbf{E}\{X|Y=i\}$ az i-edik osztályhoz tartozó várható érték vektor, és Σ_i a megfelelő kovarianciamátrix, amely j-edik sorának k-adik eleme $\mathbf{E}\{(X^{(j)}-m_i^{(j)})(X^{(k)}-m_i^{(k)})\}$. Mutasd meg, hogy a Bayes döntés kvadratikus. Mutasd meg, hogy abban a speciális estben, amikor $\Sigma_0 = \Sigma_1$ és a két osztály a priori valószínűsége megegyezik, a Bayes döntés lineáris.

14. Feladat Legyen $d=1$, és tekintsük azokat a döntésfüggvényeket, amelyek kettévágják a számegyenest, azaz, amelyek

$$g_a(x) = \begin{cases} 0 & \text{ha } x \le a \\ 1 & \text{ha } x > a \end{cases} \quad \text{vagy} \quad \begin{cases} 0 & \text{ha } x \ge a \\ 1 & \text{ha } x < a \end{cases}$$

alakúak. Mutasd meg, hogy ha L az ilyen döntések hibavalószínűsége közül a minimális, akkor $L \le \min\{p, 1-p\}$, ahol $p = \mathbf{P}\{Y=1\}$.

15. Feladat Mutasd meg, hogy ha $p=1/2$, akkor az előző feladatbeli L számra:

$$L = \frac{1}{2}\left(1 - \sup_x |F_1(x) - F_0(x)|\right),$$

ahol $F_0(x) = \mathbf{P}\{X \le x | Y=0\}$ és $F_1(x) = \mathbf{P}\{X \le x|Y=1\}$ a két osztályhoz tartozó feltételes eloszlásfüggvény. Vond le a következtetést, hogy $L=1/2$ akkor és csak akkor, ha $L^* = 1/2$.

16. Feladat A gyakorlatban gyakran nehéz megtalálni az empirikus minimumot szolgáltató döntésfüggvényt, és helyette egy közelítő megoldást kell használni. Legyen tehát $\mathcal{C}$ döntésfüggvények egy osztálya, és $\tilde{g}_n$ egy olyan osztályozási szabály, amelyre valamely ϵ_n és δ_n számokra,

$$\mathbf{P}\left\{\hat{L}_n(\tilde{g}_n) \le \hat{L}_n(g_n^*) + \epsilon_n\right\} \ge 1 - \delta_n,$$

ahol g_n^* egy az empirikus hibat minimalizáló döntés. ($\hat{L}_n$ jelöli az empirikus hibát.) Más szóval, $\tilde{g}_n$ közel minimalizálja az empirikus hibát nagy valószínűséggel. Bizonyítsd be, hogy minden $\epsilon \ge \epsilon_n$-re

$$\mathbf{P}\left\{L(\tilde{g}_n) - \inf_{g\in\mathcal{C}} L(g) > \epsilon\right\} \le \delta_n + \mathbf{P}\left\{2\sup_{g\in\mathcal{C}}\left|\hat{L}_n(g) - L(g)\right| > \epsilon - \epsilon_n\right\}.$$

17. Feladat "Fingering" téglákra. Tegyük fel, hogy az (X,Y) pár eloszlása olyan, hogy $L^*=0$, az X-nek létezik a sűrűségfüggvénye, és a Bayes döntés olyan alakú, hogy egy téglán (d dimenziós intervallumon) belül 1, azon kívül nulla. Tegyük fel, hogy adott n független mintánk ebből az eloszlásból. Tekintsük azt a g_n osztályozási szabályt, amely nulla azon a legkisebb téglán belül, amely tartalmazza az összes 1-es cimkéjű mintapontot, és a téglán kívül 0. Bizonyítsd be, hogy

$$\mathbf{P}\{L(g_n) > \epsilon\} \le n^{2d} e^{-(n-2d)\epsilon}.$$

Útmutatás: használd az órán tanult "fingering" módszert.

ALAKFELISMERÉS—HÁZI FELADAT PÉLDÁK

18. Feladat Legyen A és B két tetszőleges valószínűségi változó , x, y pedig két valós szám. Mutasd meg, hogy
$$\mathbf{P}\{A+B>x+y\} \leq \mathbf{P}\{A>x\}+\mathbf{P}\{B>y\}.$$

19. Feladat Legyen $\mathcal{A}=\{[a,b]: a \leq b\}$, azaz a számegyenes összes zárt intervallumának családja. Mennyi $s(\mathcal{A}, n)$, az $\mathcal{A}$ n-edik shatter koefficiense?

20. Feladat Legyen $\mathcal{A}_k$ a k intervallum úniójaként felírható halmazok családja. Mennyi az $\mathcal{A}_k$ Vapnik-Chervonenkis dimenziója?

21. Feladat Legyen $\mathcal{A}$ az $\mathcal{R}^d$-beli szimplexek ($d+1$ csúcsú testek) családja. Mekkora az $\mathcal{A}$ Vapnik-Chervonenkis dimenziója? (Legalább egy jó felső becslést adj!)

22. Feladat Adj felső becsléseket a következő halmazcsaládok Vapnik-Chervonenkis dimenziójára: (1) konvex k-szögek a síkon. (2) k-szögek a síkon.

23. Feladat Legyen $\mathcal{A}$ az $\mathcal{R}^d$-beli intervallumok (téglák) családja. Mutasd meg, hogy $V_{\mathcal{A}}=2d$.

24. Feladat A sík egy $A \subset [0,1]^2$ részhalmazát *monoton rétegnek* nevezzük, ha igaz az, hogy ha egy x pont eleme A-nak, akkor bármely $y \leq x$ pont is eleme A-nak, ahol az $y \leq x$ reláció úgy értendő, hogy y mindkét koordinátája nem nagyobb, mint az x megfelelő koordinátája. Mutasd meg, hogy az összes monoton réteget tartalmazó halmazosztálynak végtelen a Vapnik-Chervonenkis dimenziója.

25. Feladat Legyen
$$\mathcal{A}=\left\{\left\{x: \sum_{i=1}^{d} a_i x^{(i)} \geq 0\right\}; a_1, \ldots, a_d \in \mathcal{R}\right\},$$
azaz az origón átmenő hipersíkokkal definiált félterek családja. Bizonyítsd be, hogy minden n-re
$$s(\mathcal{A}, n)=2 \sum_{i=0}^{d-1}\binom{n-1}{i},$$
és hogy $V_{\mathcal{A}}=d$. Útmutatás: Használj teljes indukciót d és n szerint. Ez egy nehéz példa.

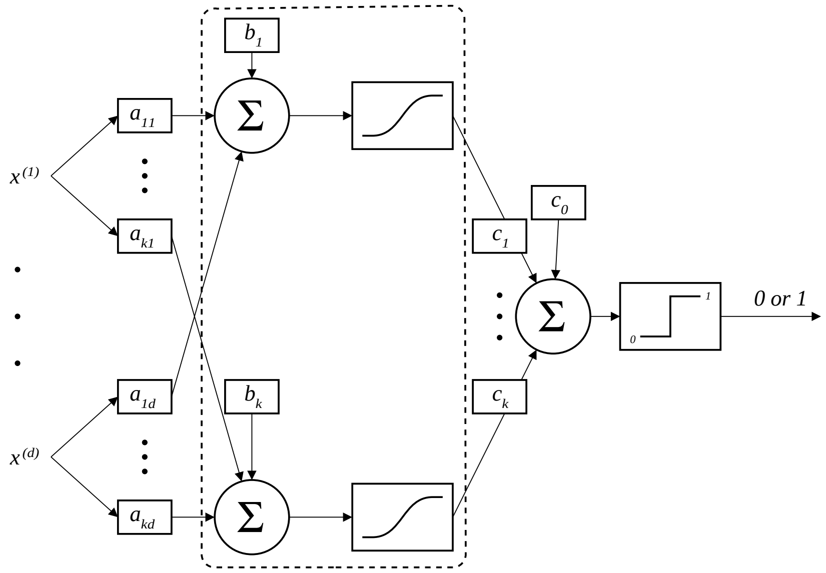
26. Feladat Mutasd meg, hogy a shatter koefficiens es a VC dimenzió között fennálló $s(\mathcal{A}, n) \leq \sum_{i=0}^{V_\mathcal{A}} \binom{n}{i}$ egyenlőtlenség éles, vagyis minden V pozitív egész számra mutass egy olyan $\mathcal{A}$ halmazosztályt, amelynek az n-edik shatter koefficiensére $s(\mathcal{A}, n) = \sum_{i=0}^{V_\mathcal{A}} \binom{n}{i}$ teljesül minden n-re! Útmutatás: Legyen $\mathcal{A}$ az összes legfeljebb V pontú halmaz osztálya.

27. Feladat Egy *egy rejtett rétegű neurális hálózat* (l. ábra) egy függvény $\mathcal{R}^d \to \mathcal{R}$, amely a következő alakba írható:

$$\psi(x) = c_0 + \sum_{i=1}^{k} c_i \sigma(\psi_i(x)),$$

ahol $x = (x^{(1)}, \ldots, x^{(d)})^T$, a bemenet vektor, σ az ugrásfüggvény: $\sigma(x) = -1$ ha $x < 0$, és $\sigma(x) = 1$ ha $x \geq 0$, továbbá

$$\psi_i(x) = b_i + \sum_{j=1}^{d} a_{i,j} x^{(j)}.$$



Tekintsük azon halmazok $\mathcal{A}$ családját, amelyek $\{x : \psi(x) > 0\}$ alakúak, ahol az $a_{i,j}, b_i$ és c_i paraméterek tetszőleges valós értékeket vehetnek fel. Mutasd meg, hogy $s(\mathcal{A}, n) \leq (n+1)^{k(d+1)+k+1}$. Útmutatás: Emlékezz vissza, hogy n pontot $\mathcal{R}^d$-ben lineáris félterekkel legfeljebb n^{d+1}-féleképpen lehet kettévágni.

28. Feladat Mutasd meg, hogy páros k esetén a fent definiált halmazosztály VC dimenziója nem nagyobb, mint kd. Útmutatás: Azt kell tehát belátni, hogy kd pont shatterelhető. Vegyünk kd általános helyzetű pontot. Feltehetjük, hogy a kiválasztandó halmaz legfeljebb $kd/2$ elemű. Osszuk be a pontokat $k/2$ legfeljebb d pontot tartalmazó csoportba. Minden ilyen d pontra ráfektethetünk egy hipersíkot, amely csakis ezeket a pontokat tartalmazza. Két neuron súlyainak megfelelő beállításával a ψ függvény értéke pozitívvá tehető ezeken a pontokon...

29. Feladat Legyenek $\tilde{\eta}_0, \tilde{\eta}_1 : \mathcal{R}^d \to [0,1]$ tetszőleges függvények, és definiáljuk a g döntést a következőképp:

$$g(x) = \begin{cases} 1 & \text{ha } \tilde{\eta}_1(x) \geq \tilde{\eta}_0(x) \\ 0 & \text{egyébként.} \end{cases}$$

Bizonyítsd be, hogy

$$\mathbf{P}\{g(X) \neq Y\} - L^* \leq \mathbf{E}|\eta(X) - \tilde{\eta}_1(X)| + \mathbf{E}|1 - \eta(X) - \tilde{\eta}_0(X)|.$$

30. Feladat Jelölje $f_0(x)$ és $f_1(x)$ az X feltételes sűrűségfüggvényét az $\{Y = 0\}$, illetve $\{Y = 1\}$ feltételekkel. Legyen továbbá $q_0 = \mathbf{P}\{Y = 0\}$, $q_1 = \mathbf{P}\{Y = 1\}$. Mutasd meg, hogy a Bayes döntés

$$g^*(x) = \begin{cases} 1 & \text{ha } q_1 f_1(x) \geq q_0 f_0(x) \\ 0 & \text{egyébként} \end{cases}$$

alakú. Mutasd meg továbbá, hogy a

$$g(x) = \begin{cases} 1 & \text{ha } \tilde{q}_1 \tilde{f}_1(x) \geq \tilde{q}_0 \tilde{f}_0(x) \\ 0 & \text{egyébként} \end{cases}$$

dönésre

$$\mathbf{P}\{g(X) \neq Y\} - L^* \leq \int \left| \tilde{q}_0 \tilde{f}_0(x) - q_0 f_0(x) \right| dx + \int \left| \tilde{q}_1 \tilde{f}_1(x) - q_1 f_1(x) \right| dx$$

teljesül.

31. Feladat Legyenek U, V és Z diszkrét valószínűségi változók úgy, hogy Z független (U, V)-től. Bizonyítsd be, hogy

$$\mathbf{E}\{f(U)|V, Z\} = \mathbf{E}\{f(U)|V\}.$$

32. Feladat Bizonyítsd be, hogy az 1-NN szabály *admisszibilis*. Ez azt jelenti, hogy létezik az (X, Y) párnak egy olyan eloszlása, amelyre az 1-NN szabály várható hibavalószínűsége $\mathbf{E}L(g_n)$ minden n-re szigorúan <u>kisebb</u>, mint a k-NN szabályé, minden $k > 1$-re. Útmutatás: Legyen X egyenletes eloszlású a $[0,1] \cup [3,4]$ halmazon és legyen $Y = 0$ ha $X \in [0,1]$ és $Y = 1$ ha $X \in [3,4]$. Ekkor az 1-NN szabály csak akkor hibázik, ha a $[0,1]$ és $[3,4]$ intervallumok egyikébe egy adatpont sem esik.

33. Feladat Legyen $\mathbf{P}\{Y = 0\} = \mathbf{P}\{Y = 1\} = 1/2$, és legyenek az X feltételes sűrűségfüggvényei $Y = 0$ illetve $Y = 1$ feltétellel $f_0(x) = 2x$ és $f_1(x) = 2 - 2x$ ha $x \in [0,1]$ és $f_0(x) = f_1(x) = 0$ ha $x \notin [0,1]$. Számítsd ki a Bayes hibát, és az 1-NN és 3-NN szabályok aszimptotikus hibavalószínűségét!

34. Feladat Mutasd meg, hogy ha $X_{(1)}$ az X legközelebbi szomszédja $X_1, \ldots, X_n$ között, akkor $\|X - X_{(1)}\|$ nullához tart valószínűségben, ahogy $n \to \infty$. Útmutatás: Rögzített $X = x$ esetén $\|x - X_{(1)}\| > \epsilon$ pontosan azt jelenti, hogy

$$\frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \in S_{x,\epsilon}\}} < \frac{1}{n},$$

ahol $S_{x,\epsilon}$ az x középpontú ϵ sugarú gömböt jelöli. Ha $\mu(S_{x,\epsilon}) > 0$, akkor a nagy számok törvénye szerint a bal oldal egy pozitív számhoz konvergál.

35. Feladat Egydimenziós ($d = 1$) osztályozási feladatok esetén a k-*spacing* osztályozási szabályt a következőképpen definiáljuk: legyen $k < n$ egy pozitív egész szám, és legyen $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ az $X_1, \ldots, X_n$ adatpontok nagyság szerinti sorbarendezése, azaz $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$. (Tegyük fel, hogy X-nek létezik a sűrűségfüggvénye, tehát a sorbarendezés egy valószínűséggel egyértemű.) A számegyenest N intervallumra particionáljuk: $A_1, \ldots, A_N$, ahol $N = \lceil n/k \rceil$, és $j = 1, \ldots, N-1$-re

$$X_{(k(j-1)+1)}, \ldots, X_{(kj)} \in A_j,$$

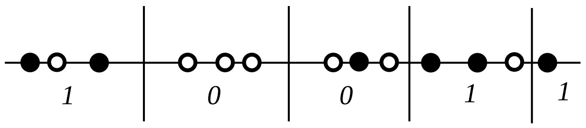
továbbá

$$X_{(k(N-1)+1)}, \ldots, X_{(n)} \in A_N.$$

Helyezzük az A_j és A_{j+1} intervallumok határát az A_j-beli legjobboldalibb és az A_{j+1}-beli legbaloldalibb adatpont közötti felezőpontba. Az osztályozási szabály ekkor többségi döntés az egyes halmazokon belül (l. ábra) , vagy precízebben:

$$g_n(x) = \begin{cases} 1 & \text{if} \quad \sum_{i=1}^n I_{\{X_i \in A(x), Y_i = 1\}} > \sum_{i=1}^n I_{\{X_i \in A(x), Y_i = 0\}} \\ 0 & \text{otherwise,} \end{cases}$$

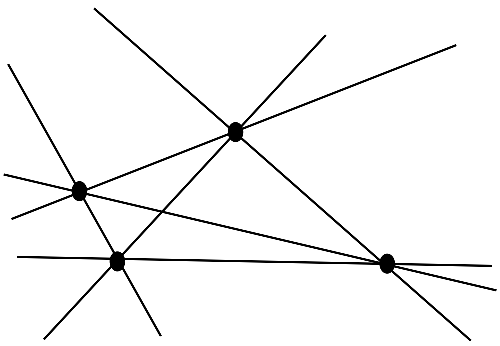
ahol $A(x)$ az x-et tartalmazó intervallumot jelöli.



Mutasd meg, hogy g_n univerzálisan konzisztens, ha $k \to \infty$ és $k/n \to 0$ ahogy $n \to \infty$.

Útmutatás: használd a partíciós szabályokra kimondott általános konzisztenciatételt.

36. Feladat Legyen $d = 2$. Vegyük az $X_1, \ldots, X_n$ adatpontok közül az első k darabot ($k < n$), és kössünk össze bármely két pontot. (Az ábrán ez $k = 4$ pontra látszik.) Az egyenesek a síkot $1 + k(k+1)/2$ halmazra particionálják. (Miért?) (Feltehetjük, hogy X-nek létezik a sűrűségfüggvénye.)



A maradék $n - k$ pont segítségével definiálhatunk egy osztályozási szabályt, amely többségi döntést hoz a partíció halmazai fölött. Mutasd meg, hogy a kapott szabály univerzálisan konzisztens, ha $k \to \infty$ és $k^2/n \to 0$ ahogy $n \to \infty$.

Útmutatás: használd a partíciós szabályokra kimondott általános konzisztenciatételt.