ICA ES TSAI

ANDRIS, BARNA, AND CSABA

Abstract.

1. Elozmenyek

Unsupervised learning-et szeretnenk individualis framework-ben modellezni. Az otlet az, hogy az eddigi framework-ot a feje tetejere allitjuk: A tanulo mondja a loss fuggvenyt, a kornyezet prezental egy mintat. A loss fuggveny valami megszoritott osztalybol jon. Ha pl. surusegfuggveny becslest akarunk ebben a framework-ben modellezni, akkor a loss fuggvenyek halmaza olyan l fuggvenyeket tartalmaz, amelyekre

$$\int \exp(-l(x))dx = 1, \quad \text{(vagy konstans)} \tag{1.1}$$

[CS: Szerintem ez onmagaban eleg szep dolog, emlekszel, hogy mit szenvedtunk azon, hogy hogyan lehetne nyelvtanulast individualis sorozatos framework-ben kezelni? Hat igy lehetne..)]

Nomost, klasszikusan az ICA arrol szol, hogy van d fuggetlen forrasod, ezek surusegfuggvenyei $p_{S_1}, \ldots, p_{S_d}$. Legyen $p_{\mathbf{S}}$ ezek szorzata. Dobjuk $\mathbf{S}$-t ebbol ($\mathbf{S}$ tehat d-dim vektor). Van egy ismeretlen nemszingularis $\mathbf{A}^*$ matrix. A megfigyelesed

$$\mathbf{X} = \mathbf{A}^*\mathbf{S} \tag{1.2}$$

keverek. A cel, hogy talaljunk egy olyan $\mathbf{W}$ matrixot, hogy $\mathbf{WX}$ komponensei fuggetlenek. Az adatok $\mathbf{X}_1, \ldots, \mathbf{X}_T$, ahol $\mathbf{X}_i = \mathbf{A}^*\mathbf{S}_i$, $\mathbf{S}_1, \ldots, \mathbf{S}_T$ fuggetlen mintak $p_{\mathbf{S}}$-bol.

Tegyuk fel hogy van valamilyen $p_{\tilde{\mathbf{S}}} = \prod_{i=1}^d p_{\tilde{S}_i}$ tippunk a $p_{\mathbf{S}}$ eloszlasra. Idealis esetben persze $p_{\tilde{S}_i} = p_{S_i}$ lenne. A tovabbiakban a $p_{\tilde{S}_i}$ eloszlasokat modeling eloszlasoknak fogjuk hivni. Ekkor a problemat meg lehet egy ML formalizmussal kozeliteni:

Hasznaljuk ki, hogy ha $\mathbf{X} = \mathbf{A}\tilde{\mathbf{S}}$, akkor

$$p_{\mathbf{X}}(\mathbf{x}) = p_{\mathbf{A}\tilde{\mathbf{S}}}(\mathbf{x}) = \frac{p_{\tilde{\mathbf{S}}}(\mathbf{A}^{-1}\mathbf{x})}{|\det \mathbf{A}|} = p_{\tilde{\mathbf{S}}}(\mathbf{W}\mathbf{x})|\det \mathbf{W}|$$

Keressunk tehat egy olyan $\mathbf{W} \doteq \mathbf{A}^{-1}$ matrixot, amelyre

$$L_T(\mathbf{W}) = -\frac{1}{T} \log p_{\mathbf{A}\tilde{\mathbf{S}}}(\mathbf{X}_1, \ldots, \mathbf{X}_T)$$

minimalis.

Date: March 4, 2009.

Tegyuk fel, hogy a mintak iid-k. **[B: ez nem feltelenul kellene, vannak eredmenyek stacinoarius sorozatokra is]**. Kihasznalva, hogy

$$\log p_{\mathbf{A}\tilde{\mathbf{S}}}(\mathbf{X}_t) = \log p_{\tilde{\mathbf{S}}}(\mathbf{W}\mathbf{X}_t) + \log|\det \mathbf{W}|,$$

kapjuk hogy

$$L_T(\mathbf{W}) = -\log|\det \mathbf{W}| - \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{d}\log p_{\tilde{S}_i}(\mathbf{w}_i^T\mathbf{X}_t) \tag{1.3}$$

ahol $\mathbf{w}_i^T$ az i^{th} sora $\mathbf{W}$-nek. **[B: ez a keplet megtalalhato** [Hyv99] **13.oldalan, eredetileg pedig talan** [PGJ92] **cikkben.]**

Ha az (1.3) egyenlet $\mathbf{W} = (\mathbf{w}_1^T, .., \mathbf{w}_d^T)$-ben szigoruan konvex, akkor valamely konvex tartomanyon, akkor a [CBL06] konyv Thm 3.1 ad egy regret korlatot a follow the leader algoritmus regretjere.

[CS: Ezzel persze meg nem oldottuk meg a teljes puzzle-t. De errol meg majd kesobben. Eloljaroban annyit, hogy nem is azt banom, hogy $p_1, \ldots, p_d$-t ismertnek tetelezzuk fel ehhez az analizishez, hanem inkabb az erdekel (es erre nem tudom a valaszt), hogy egy algoritmus, ami a fentieket implementalja, miert mukodik "jol" meg akkor is, ha az $\mathbf{S}_1, \ldots, \mathbf{S}_n$ sorozat korrelalt, es a marginalisok nem is a $p_1, \ldots, p_d - t$ kovetik. (Tudod, amikor egy sinust es egy negyszogjelet kettevalaszt az algoritmus.)]

[B: A sinus es negyszogjel szetvalasztasahoz erdemes lenne a Wiskott fele Slow Feature Analizist (SFA) tanulmanyozni [WS02]**. Ez arrol szol, hogy ha 2 sima fuggvenyt osszekeverunk, akkor a keverek kevesbe lesz ezek utan sima, emiatt a visszallitashoz a becsult komponensek simasagat kell minimalizalni. E modszer bizonyitottan kapcsolatban all az ICAval** [BBW06]**, es vannak erdekes alkalmazasai pl bizonyos invarianciak tanulasa** [WS02]**. Az, hogy a keverekbol pontosan mikor allithatok vissza az eredeti sima komponensek viszont tudtommal meg nem ismert. Erdekes lenne egy Darmois-Skitowich tetel szeru elmeletet kidolgozni SFA-ra is.]**

2. Az ICA egyertelmusegi kerdesei

E fejezetben osszefoglaljuk hogy mikor, es milyen undeterminanciaval allithatok vissza az ICA komponensek.

A Darmois Skitowich tetel, Cardoso eredmenyei, Fabain Theis eredmenyei. Max 1 forras lehet Gauss i.i.d estben. Bizonyos idosorok eseten tobb Gauss forras is lehet. Review on iid and on non-iid cases.

Definition 2.1 (2 matrix ekvivalens)**.** Azt mondjuk, hogy $\mathbf{W}_1$ es $\mathbf{W}_2$ matrixok ekvivalensek ($\mathbf{W}_1 \sim \mathbf{W}_2$), ha megkaphatok egymasbol a sorok permutalasaval es a sorok atskalazasaval, ahol e skalaparameterek valosak.

3. Az ML koltsegfuggveny tulajdonsagai

3.1. Az ML koltsegfuggveny konvexitasarol.

Lemma 3.1. *A positive definite matrixok halmaza konvex.*

[B: szigoruan? konvex]

Lemma 3.2. $-\log|\det(\mathbf{W})|$ *fuggveny szigoruan konvex a positive definite* $\mathbf{W}$ *matrixok halmazan.*

Proof. A konvexitas bizonyitasa megtalalhato pl itt: 8. tetel a [DCT91] cikknek az 1506. oldalan, valamint [Fan50]. A szigorusag belathato az eredeti bizonyitas elemzesevel: Let $\mathbf{X}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_1)$, $\mathbf{X}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_2)$, where $\mathbf{K}_1 \neq \mathbf{K}_2$ are positive definite matrices. Kepezzunk mixture eloszlast ezekbol, azaz legyen θ val. valtozo $\mathbf{P}(\theta = 1) = \lambda$, $\mathbf{P}(\theta = 2) = 1 - \lambda$ eloszlassal valamilyen $0 \leq \lambda \leq 1$ ertekkel. Legyen $\theta, \mathbf{X}_1, \mathbf{X}_2$ fuggetlenek, es legyen a $\mathbf{Z}$ mixture eloszlas $\mathbf{Z} = \mathbf{X}_\theta$. Ekkor $\mathbf{Z}$ surusegfuggvenye:

$$p_{\mathbf{Z}}(\mathbf{z}) = \lambda \mathcal{N}_{\mathbf{z}}(\mathbf{0}, \mathbf{K}_1) + (1-\lambda)\mathcal{N}_{\mathbf{z}}(\mathbf{0}, \mathbf{K}_2)$$

mig a kovariancia matrixa pedig: $\mathbf{K} = \lambda \mathbf{K}_1 + (1-\lambda)\mathbf{K}_2$. Fix varianciaju eloszlasok kozott a normalisnak a legnagyobb az entropiaja, ezert

$$\begin{aligned} (3.1)\quad \frac{1}{2}\log(2\pi e)^n|\lambda\mathbf{K}_1 + (1-\lambda)\mathbf{K}_2| &\geq H(\mathbf{Z}) \\ &\geq H(\mathbf{Z}|\theta) \\ &= \frac{\lambda}{2}\log(2\pi e)^n|\mathbf{K}_1| + \frac{1-\lambda}{2}\log(2\pi e)^n|\mathbf{K}_2| \end{aligned}$$

Emiatt $\log|\lambda\mathbf{K}_1 + (1-\lambda)\mathbf{K}_2| \geq \lambda \log|\mathbf{K}_1| + (1-\lambda)\log|\mathbf{K}_2|$, tehat $\log|\det(\mathbf{K})|$ konkav. Tovabba akkor lehetne csak nem szigoruan konkav, ha (3.1) egyenletben egyenloseg allna, azaz a $\mathbf{Z}$ mixture of Gauss is Gauss eloszlasu lenne:

$$\lambda\mathcal{N}_{\mathbf{z}}(\mathbf{0}, \mathbf{K}_1) + (1-\lambda)\mathcal{N}_{\mathbf{z}}(\mathbf{0}, \mathbf{K}_2) = \mathcal{N}_{\mathbf{z}}(\mathbf{0}, \lambda\mathbf{K}_1 + (1-\lambda)\mathbf{K}_2), \quad \mathbb{R}^n$$

$$\lambda C_{\mathbf{K}_1}\exp(-\frac{1}{2}\mathbf{z}^T\mathbf{K}_1^{-1}\mathbf{z}) + (1-\lambda)C_{\mathbf{K}_2}\exp(-\frac{1}{2}\mathbf{z}^T\mathbf{K}_2^{-1}\mathbf{z}) = C_{\mathbf{K}}\exp(-\frac{1}{2}\mathbf{z}^T\mathbf{K}^{-1}\mathbf{z})$$

$$\lambda C_{\mathbf{K}_1}\exp(-\frac{1}{2}\mathbf{z}^T(\mathbf{K}_1^{-1} - \mathbf{K}^{-1})\mathbf{z}) + (1-\lambda)C_{\mathbf{K}_2}\exp(-\frac{1}{2}\mathbf{z}^T(\mathbf{K}_2^{-1} - \mathbf{K}^{-1})\mathbf{z}) = C_{\mathbf{K}}$$

A jobb oldal konstans $\mathbf{z}$-ben. A bal oldal viszont csak akkor lehet konstans, ha $\mathbf{K}_1 = \mathbf{K}$, $\mathbf{K}_2 = \mathbf{K}$ azaz $\mathbf{K}_1 = \mathbf{K}_2$ ami ellentmondas. □

Theorem 3.3 ($L(\mathbf{W})$ konvex)**.** *Ha a p_i szigoruan log-konkav $\forall i = 1, \ldots, d$, akkor $L(\mathbf{W})$ (1.3) koltsegfuggveny konvex a positive definite $\mathbf{W}$ matrixok konvex halmazan.*

[B: sziguran? konvex. illetve kell-e itt a szigorusag a konkavsagban? ... eleg lenne ha a $\log|\det(\mathbf{W})|$ szigoruan konvex...]

3.2. **Az ML koltsegfuggveny aszimptotikus tulajdonsagai.** Felmerul a kerdes, hogy a (1.3) ML koltsegfuggveny mennyire jo, azaz globalis minimuma vissza adja-e az eredeti $\mathbf{A}^*$ kevero matrixot, es a $\mathbf{S}$ rejtett forrasokat. E kerdessel kapocsolatban Cardoso a kovetkozo eredmenyre jut [Car97] cikkben:

$$(3.2)\quad L_T(\mathbf{W}) = -\frac{1}{T}\sum_{t=1}^{T}\log p_{\mathbf{A}\tilde{\mathbf{S}}}(\mathbf{X}_t) \quad \underrightarrow{T \to \infty} \quad -\int p_{\mathbf{X}}(\mathbf{x})\log p_{\mathbf{A}\tilde{\mathbf{S}}}(\mathbf{x})d\mathbf{x}$$

$$\begin{aligned} L_\infty(\mathbf{W}) &= -\int p_{\mathbf{X}}(\mathbf{x}) \log p_{\mathbf{A}\tilde{\mathbf{S}}}(\mathbf{x}) d\mathbf{x} \\ &= \int p_{\mathbf{X}}(\mathbf{x}) \log \frac{p_{\mathbf{X}}(\mathbf{x})}{p_{\mathbf{A}\tilde{\mathbf{S}}}(\mathbf{x})} d\mathbf{x} + \int p_{\mathbf{X}}(\mathbf{x}) \log p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= D(p_{\mathbf{X}} || p_{\mathbf{A}\tilde{\mathbf{S}}}) + H(p_{\mathbf{X}}) \\ &= D(\mathbf{A}^*\mathbf{S} || \mathbf{A}\tilde{\mathbf{S}}) + H(\mathbf{A}^*\mathbf{S}) \end{aligned}$$

ahol $D(\cdot||\cdot)$ a Kullback-Leibler tavolsag, H a differential entropy. Mivel az $\mathbf{A}^*\mathbf{S}$ eloszlas, es igy $H(\mathbf{A}^*\mathbf{S})$ konstans, ezert $L_\infty(\mathbf{W})$ pontosan akkor minimalis ha az $\mathbf{A}^*\mathbf{S}$ és $\mathbf{A}\tilde{\mathbf{S}}$ eloszlasa azonos.

Corollary 3.4. *Ha az $p_{\tilde{\mathbf{S}}} = p_{\mathbf{S}}$, tehat a modeling density perfect, es az osszes marginalis eloszlas egyforma $p_{S_i} = p_{S_j}$, asszimptotikusan meg akkor is visszatudunk allitani egy $\mathbf{W} \sim (\mathbf{A}^*)^{-1}$ matrixot az ML koltsegfuggveny globalis minimumakent, azaz az ML koltsegfuggveny konzisztens. Ha a p_{S_i} eloszlasok kulonboznek es sorrendjuk rogzitett, akkor raadasul meg a $\mathbf{W}$ soraiban sincs permutacios bizonytalansag.*

3.3. **Az ML koltsegfuggveny kapcsolata egyeb koltsegfuggvenyekkel.** Vannak mas ICA koltsegfuggvenyek is: Mutual Information kriterium, marginalis entropia osszeg kirterium, kurtozis osszeg, infomax ICA, stb. Vizsgaljuk meg ezek kapcsolatat az ML koltsegfuggvennyel. [Car97] [HO00]

3.4. **Az ML becslesben levo forrasok sorrendjerol.** Az ICAban levo permutacios bizonytalansag azt jelenti, hogy adott koltsegfuggveny eseten az optimalis $\mathbf{W}$ mtx sorainak tetszoleges permutacioja is globalis optimuma a koltsegfuggvenynek.

Theorem 3.5 (Positive definite halamazon konvex L_T eseten a szeparacio egyertelmu)**.** *Tegyuk fel, hogy maximum egy forras Gauss, es a $p_{\tilde{\mathbf{S}}}$ surusegfuggvenyek legalabb annyira pontosak, hogy L_∞ minden globalis minimuma szeparalja a forrasokat, azaz $\mathbf{W} \sim (\mathbf{A}^*)^{-1}$, es ezek kozott letezik legalabb egy positive definite. Valamelyiket jelolje $\mathbf{W}_1$. Megmutatjuk, hogy abban az esetben ha a $\mathbf{W}$ matrixokat megszoritjuk a positivie definte matrixok halmazara, es azokat a $p_{\tilde{\mathbf{S}}}$ surusegfuggveneket tekintjuk melyek mellett $L_T(\mathbf{W})$ konvex ezen a konvex tartomanyon, akkor ez esetben mar nem lesz igaz a permutacios bizonytalansag, azaz hogy szeparacio eseten $\mathbf{W}$ sorainak sorrendje tetszoleges lehetne, hanem pontosan egy ilyen $\mathbf{W}$ letezik ($\mathbf{W}_1$), meg abban az esetben is ha az osszes $p_{\tilde{S}_i}$ azonos.*

Proof. Mivel $L_T(\mathbf{W})$ konvex minden T-re, ezert $L_\infty(\mathbf{W}) = \lim_{T\to\infty} L_T(\mathbf{W})$ is konvex $\mathbf{W}$-ben. Ebbol kovetkezoen vagy 1 darab, vagy vegtelen sok globalis minuma van, es a globalis minimumok konvex tartomanyt alkotnak. Azt kell belatnunk, hogy a $\mathbf{W}_1$ sorainak semelyik permutaltja nem lehet globalis minimum. Ha ugyanis $\mathbf{W}_2$ a $\mathbf{W}_1$ sorainak egy ilyen globalis minimot ado permutaltja lenne, akkor $\lambda\mathbf{W}_1 + (1-\lambda)\mathbf{W}_2$ is globalis minimum lenne, s igy ICA szeparalo matrix minden $\lambda \in [0, 1]$ eseten. Ez azonban ellentmondas mert tudjuk hogy az osszes ICA megoldas a $\mathbf{W}_1$-gyel ekivalens matrix – azaz sorainak permutacioja es estleg atskalozottja – kell hogy legyen, mig a $\lambda\mathbf{W}_1 + (1-\lambda)\mathbf{W}_2$ matrixok kozott nyilvan van amelyik nem ilyen alaku. □

Osszefoglalva az a helyzet, hogy ha az $\mathbf{A}^*$ kevero matrix egy permutacioja pozitiv definit, akkor letezik pozitiv definit megoldas a $\mathbf{W}$-re, es ez egyertelmu. Ebben az esetben lehet hatekony az algoritmus. **[B: az latom, hogy ha $\mathbf{A}^*$ positive**

definit az jo, hisz akkor inverze is az. De miert jo az is, ha valamely P permutacios mtxra PA* positive definite?]

[B: Egyebekent az is eleg lenne, ha tudnank ugy tovabb keverni vmilyen B matrixszal X = A*S megfigyeleseket (preprocessing), hogy a BX = BA*S-ben BA mar positive definite legyen, meg ha A nem is volt az.]

[CS: Mik azok a matrixok amiknek van olyan permutacioja ami pozitiv definit? (Ugy ertem, intuitive.. Vagy valami ekvivalens atfogalmazasban..)]

Megoldas lehetne meg, hogy feltetelezzuk a forrasok mar feherek, es akkor nem kellene a $\log|\det \mathbf{W}|$-t hasznalni az (1.3) egyenletben, ekkor azonban az ortogonalis mtxok kozt kellene keresni, ez a manifold viszont nem konvex...

[B: erdekes, hogy altalaban az ICA algoritmusok azzal kezdodnek, hogy feheritsuk az adatokat, mert ezutan mar csak egy ortogonalis matrixot kell keresni. A mi modszerunkben viszont nem szabad feheriteni, mert azutan a domain amin optimalizalnuk mar nem lenne konvex :)]

[Cs: Lehet, hogy tobbszoros sajatertekek vannak es minden p_i azonos. Pl. legyen $A = I$. Ekkor minden paros permutaciohoz tartozo permutacios matrix is megoldas, nem? A determinansa ezeknek a mx-oknak 1, es a kriterium masodik reszet nem valtoztatjak tegyuk fel, hogy a minta szimmetrikus, vagy vegtelen sok van belole, vagy $\sum_i -E[logp_i((WX)_i)]$-re csereltuk ezt. Azaz ki mondta, hogy $log \det(\mathbf{W})$ szigoruan konkav? Nem az. Mint ahogy a masodik resz sem az. Ugy tunik legalabbis. Hmm, ez erdekes. Megjegyzem a paros permutaciokhoz tartozo matrixok nem pozitiv definitek, vegyuk pl azt aki $1 <-> 2$ $3 <-> 4$ cseret csinalja. Az $1x1$-es minor a (0) matrix, igy nem pozitiv definit, stb. Viszont mivel a paratlan permutaciok determinansa -1, ezekre a kriterium vegtelent ad. Ezek nem jonnek szoba sem! Hu de furcsa ez.]

4. Egy Linearis Algebrai Kovetkezmeny

Theorem 4.1 (Positive definit matriox sorainak permutaciojarol)**.** *Legyen adott egy tetszoleges positive definite $\mathbf{W}_1$ matrix, es egy tetszoleges σ permutacio. Ha $\mathbf{W}_1$ sorait megpertaljuk σ szerint akkor a keletkezo $\mathbf{W}_2$ matrix pontosan akkor lesz positive definite, ha a σ permutacio az identias volt.*

Proof. A feladathoz hozzarendelheto egy olyan ICA feladat melyben $\mathbf{A}^* = \mathbf{W}_1^{-1}$, es minden p_{S_i} surusegfuggveny azonos log-konkav, de nem Gauss. Ekkor $\mathbf{W}_1$ globalis minimuma lesz L_∞-nek, ami permutacio invarians $\mathbf{W}$ soraira nezve, igy $\mathbf{W}_2$ is globalis minimum lesz. No de L_∞ konvex, igy $\lambda\mathbf{W}_1 + (1-\lambda)\mathbf{W}_2$ is globalis minimum lenne, s igy ez ICA szeparalo matrix lenne minden $\lambda \in [0, 1]$ eseten. Ez viszont ellentmondas, mert e matrixok kozott kell lenni olyannak ami nem ekvivalens $\mathbf{W}_1$-gyel, es igy nem lehet szeparalo matrix. Emiatt $\mathbf{W}_2$ nem lehet eleme a positive definite matrixok halmazanak. □

[B: Ismert ennek az allitasnak mas elemi bizonyitasa is?... Kihasznalva pl. valahogy, hogy tudjuk, a permutacio utan is W-nek szimmetrikusnak kell maradnia, es minden principa minorjal positive definite kell hogy legyen...]

[B: Vajon itt kell e szimmetria W-ben, vagy eleg az, hogy a sajatertkek mind pozitivak?]

5. ICA-ban Lokalis es Globalis Optimumok Vizsgalata

Irodalmi attekintes, one bit matching problema, ML-ICA with modelling density (amikor az igazi denisty-t nem tudjuk). Eredmenyek, and open questions.

Bizonyos forrasokra es koltsegfuggvenyekre egyebken van eredmeny arra, hogy annak a koltsegfuggvenynek minden lokalis optimum jo ICA megoldast ad, tehat gradiens modszerek hasznalhatok. De ez egyelore meg csak 2 forras eseten bizonyitott, tobb forras eseten nyitott problema [GM08]. Ez a cikk arra az esetre bizonyit, amikor a koltsegfuggveny a kurtozis, es a forrasok kozott nincs zero kurtozisu.
E cikkben szinten a lokalis optimumokat akarjak elkerulni: [BZBJ05]

Az ismert, hogy bizonyos ICA koltegfuggvenyek bizonyos forrasokon sok lokalis es not feasible optimummal rendelkeznek. A feladatunk az hogy olyan koltsegfuggvenyt, eloszlasokat, es a kevero matrixra domaint allapitsunk meg, ahol a keletkezo feladat egy konvex halmazon torteno konvex optimalizalas lesz.

6. Csaba Osoreg Kerdese, non-iid sources

Hogy lehet, hogy az iid sorozatokra (azaz minden forras egy iid sorozat) kigyezett ICA algoritmusok nagyon jol mukodnek akkor is, ha a sorozatok nem iid-k?

[CS: Asszem ezt most megertettem, legalabbis egy aspektusat (egy csomo mindent irtam kozben, amitol megertettem, amit most kitoroltem, mert mar ertem mi van.. milyen jo, hogy levelezunk;) Nyilvan ez is kozismert, stb. Mindegy.. Itt van amit kigondoltam:]

[B: Latom, jo hangulatban voltal ◡̈]

[B: errol a kerdesrol lasd meg a Section 1 beli SFA-s megjegyzesem.]

Legyen az egyszeruseg kedveert ket forrasunk: $\{X_1, X_2, \ldots\}$ es $\{Y_1, Y_2, \ldots\}$. Lehet, hogy ezek nem iid sorozatok. Viszont tegyuk fel, hogy azert fuggetlenek, hiszen ICA-rol van szo.

A konkretsag kedveert vegyuk a korabban diszkutalt modszert. Legyen az X_t, Y_t marginalisai a t-edik pillanatban $p_{X,t}$, illetve $p_{Y,t}$ (lehet, hogy ezek azonosak, de ezt most nem tesszuk fel). Itt most feltettuk, hogy a surusegfuggvenyek leteznek stb.

Ha a $\mathbf{W}$-vel visszeforgatott folyamatok X'_t, Y'_t, akkor egy lehetseges minimalizalando kriterium $\mathbf{W}$-ben

$$J_n(\mathbf{W}) = F(\mathbf{W}) - \sum_{t=1}^{n} \log(p_{X,t}(X'_t) p_{Y,t}(Y'_t)).$$

Itt $F(\mathbf{W})$ az ominozus $\log\det$-es tag. Itt a csavar az elozoekhez kepest az, hogy ez nem ugyanaz, mint ami a maximum likelihood-bol kijonne. Ugyanis ez

$$F(\mathbf{W}) - \log p_{X,1:n}(X'_1, .., X'_n) p_{Y,1:n}(Y'_1, .., Y'_n),$$

lenne, ahol $p_{X,1:n}$ (ill. $p_{Y,1:n}$) az $(X_1, .., X_n)$ (ill. $(Y_1, .., Y_n)$) egyuttes surusegfuggvenyei.

$$J_n(\mathbf{W}) = F(\mathbf{W}) - \sum_{t=1}^{n} \log(p_{X,t}(X'_t)p_{Y,t}(Y'_t)), \tag{6.1}$$

$$J'_n(W) = F(\mathbf{W}) - \log p_{X,1:n}(X'_1, .., X'_n)p_{Y,1:n}(Y'_1, .., Y'_n) \tag{6.2}$$

Azt allitom, hogy (*) megis ertelmes kriterium. Ez mindosszesen annyin mulik, hogy minden rogzitett t-re az X_t, Y_t egyuttes surusegfuggvenye is szorzat alakra bomlik, amikben $p_{X,t}$ es $p_{Y,t}$ fognak szerepelni. Ez ugye trivi. Nomost, miert ne vehetnek ezek szorzatat, annak a logaritmusat, stb.

Ebbol talan kijon, hogy minden olyan $\mathbf{W}$, ami valoban visszaallitja az X_t, Y_t-t $J_n(\mathbf{W})$-t minimalizalja (tartassuk n-t vegtelenbe, tegyunk fel mindenfele korlatossagot, legyen ergodicitas, stb., mittomen). Azaz $J_n(\mathbf{W})$ ertelmes kriterium.

A kulcs valahogy az ebben a gondolatmenetben, hogy a marginalisok szorzata minden rogzitett t-re eloallitja X_t, Y_t egyuttes eloszlasfuggvenyet. Es ehhez nem kell semmit feltenni arrol, hogy X_t (vagy Y_t) maga iid. A masik kulcs, hogy $\mathbf{W}$-ben nincs olyan sok szabadsagi fok (ezt meg nem hasznaltuk). Emiatt az ember azt gondolja, hogy ahogy $n \to \infty$, a kriteriumfuggveny egyre definitebb feluletet ad. Meg azt is megkockaztatom, hogy ertelmes esetekben $\arg\max_{\mathbf{W}} J_n(\mathbf{W})$ egy valoszinuseggel tart az ICA feladat megoldasainak halmazahoz (azaz a modszer konzisztens).

Ez egy kicsit arra is ramutatna, hogy ha az iid esetre be tudja azonositani az ember azokat a kriteriumokat, amik konzisztenciahoz vezetnek, akkor ebbol mellesleg kijohet a nem iid folyamatokra is egy-egy konzisztencia bizonyitas.

Azaz, altalanosabban, tegyuk fel hogy a kriteriumunk

$$J_n(\mathbf{W}) = F(\mathbf{W}) - \sum_{t=1}^{n} L_t(X'_t, Y'_t)$$

alaku es belattuk, hogy iid folyamatokra $argmax_{\mathbf{W}} J_n(\mathbf{W})$ konzisztens. Kerdes: Kovetkezik-e ebbol, hogy nem iid de pl. (stac) ergodikus folyamatokra is konzisztens a becsles. **[CS:Barna: Van ilyen eredmeny? Vagy nem igaz?] [B: Ugy emlekszem Pham-nak vannak eredmenyei ezzel kapcsolatban, megnezem, beirom majd hogy pontosan mit csintalt]**

[CS:Most eltavolodtunk az individualis sorozat temakortol, de talan nem baj. Majd meg visszaterhetunk rajuk, de asszem eloszor ilyen dolgokat erdemes tisztazni.]

Folytatva az elozo gondolatmenetet nezzuk csak, mi lesz, ha $n \to \infty$. Ekkor, ergodicitast felteve,

$$\lim_{n\to\infty} J'_n(\mathbf{W})/n = \tag{6.3}$$

$$\lim_{n\to\infty} -1/n \sum (\log P_{X_i}(X'_i) - \log P_{X'_i}(X'_i) + \log P_{X'_i}(X'_i) + \text{ugyanez Y-ra}$$

$$= D(P_{X'}||P_X) + H(X') + D(P_{Y'}||P_Y) + H(Y')$$

ahol $D(p||q)$ a KL-divergencia-rata, H pedig az (differential)entropy-rate. Ennek a minimuma pedig a $H(X,Y) = H(X) + H(Y)$ - legalabbis diszkret esetben -, hiszen az eredeti J koltsegfuggvenyunk (az eltuno $F(\mathbf{W})$ tagtol eltekintve) eppen az, hogy milyen hosszan tudjuk tomoriteni az X', Y' sorozatot, ami viszont - mivel $\mathbf{W}$ invertalhato - ugyanannyi kell legyen, mint amennyire az X, Y-t tudjuk.

Ha a J_n - betunkenti vagy fuggetlenseget feltetelezo - kriteriumot vesszuk, akkor (*)-ban nem a ratakat, hanem az 1D marginalisok divergenciajat es entropiajat kapjuk. Mivel a $P_X = P'_X$ es $P_Y = P'_Y$ minimalizalja (*)-ot, az igazi fuggetlenito $\mathbf{W}$ mindket esetben jo megoldas, es nagyjabol az is latszik, hogy ugyanolyan feltetelek mellett nem lesz mas megoldas. **[A: de azt most nem latom, hogy ez miert lenne mindig egyedi. Az a gyanum, hogy erre eleg sok eredmenyt lehetne talalni (kicsit beleneztem ica-s cikkekbe - blind source separation cimszo alatt), es ott az latszik, hogy jo sokat foglalkoznak azzal, hogy az algoritmusok mikor kerulnek lokalis minimumba.]**

[B: igen sok cikk vizsgalja ezeket, es sok reszeredmeny van. Megprobalok majd utana nezni ezeknek. De nagy altalanossagban eredmenyek meg nem ismertek szerintem.]

[B: Vannak olyan eredmenyek is, amik az ICA ML megkozeliteseben azt vizsgaljak, hogy mi tortenik ha a surusegfugvenyeket nem ismerjuk pontosan csak kozelitoleg. Lasd pl Cardoso eredmenyei: [Car97]**.**

A feltelezett suruseg fuggvenyeket szokas modeling surusegfuggvenyeknek is hivni. A sejtes az, hogy eleg azt tudni (jol megbecsulni) hogy a surusegfuggvenyek sub vagy supergauss-ok. Cardoso and Laheld, 1996; Hyvärinen and Oja, 1998; Lee et al., 1999]

References

[BBW06] T. Blaschke, P. Berkes, and L. Wiskott. What is the relation between slow feature analysis and independent component analysis? *Neural Computation*, 18(10):2495–2508, 2006.

[BZBJ05] M. Babaie-Zadeh, B. Bahmani, and C. Jutten. Ica by mutual information minimization: An approach for avoiding local minima. In *EUSIPCO2005*, 2005.

[Car97] J. F. Cardoso. Infomax and maximum likelihood for source separation. *IEEE Signal Processing Letters*, 4(4):112 – 114, 1997.

[CBL06] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[DCT91] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *IEEE Transactions on Information Theory*, 37(6):1501–1518, 1991.

[Fan50] K. Fan. Information theoretic inequalities. *Proc. Natl. Acad. Sci. USA*, 36(1):31Ű–35, 1950.

[GM08] F. Ge and J. Ma. Analysis of the kurtosis-sum objective function for ica. In *ISNN '08: Proceedings of the 5th international symposium on Neural Networks*, pages 579–588, Berlin, Heidelberg, 2008. Springer-Verlag.

[HO00] A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4-5):411–430, 2000.

[Hyv99] A. Hyvarinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999.

[PGJ92] D. T. Pham, Ph. Garat, and C. Jutten. Separation of mixture of independent sources through a maximum likelihood approach. *Signal Processing,Proceeding EUSIPCO '92*, 6:771–774, 1992.

[WS02] L. Wiskott and T. J. Sejnowski. Slow feature analysis: unsupervised learning of invariances. *Neural Computation*, 14(4):715–770, 2002.