

Az agy vizuális rendszerében zajló komputációk vizsgálata generatív probablisztikus modell használatával

Kun Mátyás

Témavezetők: Csikor Ferenc, Orbán Gergő

Komputációs Rendszerszintű Idegtudomány Kutatócsoport, Komputációs Tudományok Osztálya, Wigner FK

2021. december 9.

1. Bevezetés

Az idei félévben önálló kutatási projektnek az agy vizuális kérgének vizsgálatát választottam. Ennek során a kiindulópontom az, hogy ez egy összetett hierarchikus rendszert alkot [1]. Az információ először a V1-be jut el, melyet aztán továbbít a hierarchikusan magasabb rendű területekre: V2-be, majd V4-be és végül az IT-be, ahol olyan magasszintű reprezentációk alakulnak ki, mint például arcok nézőpont-, skála-, és megvilágítás-invariáns megjelenítése. Ahhoz, hogy a különböző rétegeket külön tudjam vizsgálni, egy olyan generatív modellre volt szükségem, ami nemlineáris, hierarchikus és probablisztikus inferencát tud megvalósítani. Az általam választott modell egy hierarchikus Variational AutoEncoder (későbbiekben csak VAE). A VAE modellekkel, és azzal, hogy hogyan működnek, az alábbi forrásból ismerkedtem meg [2]. A továbbiakban ismertetem az általam használt VAE-t és részletezem a vizsgálatok menetét, jelentőségét.

2. Variational Autoencoder

2.1. Unsupervised learning, probablisztikus generatív modell

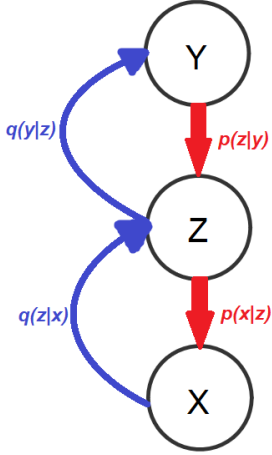
Az *unsupervised learning* egy olyan ága a gépi tanulásnak, melynek során az adataink nincsenek ellátva címkékkel. Az unsupervised learningen belül több különböző modell létezik, azonban, mivel egy *probablisztikus modellt* tervezek használni, a továbbiakban csak ezek működését fogom részletezni. A probablisztikus modellek célja, hogy olyan eloszlásokat hozzanak létre, melyek a lehető legpontosabban írják le a tanulódátokat, és melyekből ha mintavételezünk, az eredetihez hasonló adatpontokat kapunk. Ezért tanulás során olyan θ paramétereket akarunk megtalálni, melyek felhasználásával a lehető legpontosabban tudjuk közelíteni az eredeti eloszlást. Tehát a cél, hogy $p_{\theta}(x) \approx p(x)$ legyen. Ezen modellek egyik változata, ahol a megfigyeléseink statisztikáját látens változók segítségével írjuk le (pl.: $p(x) = \int p(x, z) dz = \int p(x|z)p(z) dz$ (ahol $p(z)$ a *prior eloszlás*)). Fontos tulajdonságuk, hogy tudunk az eredeti adatainkhoz hasonló adatokat is létrehozni csupán mintavételezéssel (*generatív modell*).

2.2. Közelítő poszterior, hierarchia a modellben

Annak érdekében, hogy a tanulási problémákat kezelhetővé tudjuk alakítani, szükségünk lesz egy inferencia modellre, $q_{\phi}(z|x)$ -re. Ezt szokás *közelítő poszteriornak*, vagy *kódolóznak* (encoder) is

nevezni. Úgy szeretnénk optimalizálni a Φ paramétereket, hogy $q_{\Phi}(z|x) \approx p_{\theta}(z|x)$ legyen. Mivel olyan struktúrát vizsgálok, mely hierarchikus (vizuális látókéreg), olyan modellre van szükségem, ami figyelembe veszi ezt a hierarchiát. Ennek érdekében az általam használt VAE két inferencia és két generatív leképezést tartalmaz, amit az 1. ábra szemléltet.

1. ábra. Hierarchikus VAE, ahol Y reprezentálja V2-t, Z pedig V1-t. X az adathalmaz.



Ahhoz, hogy a VAE működését optimalizálni tudjuk, szükségünk lesz az *Evidence lower bound* fogalmára (későbbiekben csak ELBO). Ezt a $p(x)$ logaritmus-valószínűségéből vezetem le:

$$\begin{aligned} \log p(x) &= \int dzdy \cdot q(z, y|x) \cdot \log p(x) = \\ &= \int dzdy \cdot q(z, y|x) \cdot \log \left(\frac{q(z, y|x)}{p(z, y|x)} \cdot \frac{p(z, y, x)}{q(z, y|x)} \right) = \\ &= \text{KL} [q(z, y|x) || p(z, y|x)] + \int dzdy \cdot q(z, y|x) \cdot \log \left(\frac{p(x|z) \cdot p(z|y) \cdot p(y)}{q(z|y) \cdot q(z|x)} \right) \end{aligned}$$

A végeredményben látható első tag a $q(z, y|z)$ és a $p(z, y|z)$ együttes eloszlások közötti *Kullback-Leibler divergencia*, a második tag pedig az ELBO. A Kullback-Leibler divergencia egy nemnegatív érték, ami két eloszlás hasonlóságát jellemzi (0-t pontosan akkor vesz fel, ha a két eloszlás megegyezik). Ha az ELBO-t tovább egyszerűsítjük, akkor megkapjuk a következő tagokat:

$$\text{ELBO} = \mathbb{E}_{q(z|x)} [\log(p(x|z))] - \mathbb{E}_{q(z|x)} [\text{KL}[q(y|z) || p(y)]] - \mathbb{E}_{q(y|z)} [\text{KL}[q(z|x) || p(z|y)]] \quad .$$

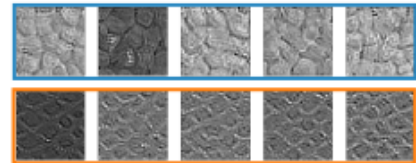
Tanítás során az a célunk, hogy ezeknek a tagoknak az együttes értékét maximalizáljuk. Így olyan eloszlást tudunk létrehozni, mely a lehető legpontosabban reprezentálja az eredeti adatunkat. Az adatok, amiket a betanított modell kiértékelése során használok, olyan természetes képek, melyek előre fehéritettek. Ezek a képek egyes textúra-családokat határoznak meg, melyeken belül különböző (azonos textúrájú) képek találhatóak (2. ábra).

3. Összefoglalás, kitekintés

A modellben a poszterior akkor is többféleképpen partícionálható, ha a generatív modellt rögzítjük. Az a fajta partícionálás, amit kutatásom során használni fogok, az úgynevezett Chain ELBO. A kutatócsoport egy másik fajta ELBO-val is foglalkozik, amit úgy hívnak, hogy Top-down inference ELBO. Mivel abban az esetben gyakrabban igénybe van véve a felső réteg (a top-down hatás miatt), ezért várható, hogy a modellnek mindegyik rétege használva lesz tanítás során, azonban a Chain ELBO-nál könnyen megvan az esély arra, hogy a felső réteg csak minimális mennyiségben tanuljon. Azonban erre csak a kutatás után fogok tudni pontos választ adni, ez egyelőre még csak sejtés.

Ahhoz, hogy jobban megismerjük agyunk vizuális látókérgét, elengedhetetlen mindegyik fajta ELBO vizsgálata. Kutatásom során remélem, sikerül majd hozzájárulnom ahhoz, hogy olyan modellt tudjunk alkotni, mely a lehető legpontosabban reprezentálja agyunk működését.

2. ábra. A keretek jelölik az egyes textúra-családokat, azon belül pedig a különböző képek láthatóak.



Hivatkozások

- [1] Corey M Ziemba, Jeremy Freeman, J Anthony Movshon, and Eero P Simoncelli. Selectivity and tolerance for visual texture in macaque v2. *Proceedings of the National Academy of Sciences*, 113(22):E3140–E3149, 2016.
- [2] Diederik P Kingma and Max Welling. An introduction to variational autoencoders. *arXiv preprint arXiv:1906.02691*, 2019.