

Szakdolgozat kivonat

Magas dimenziós adat alacsonyabb dimenzióba való tömörítése

Alkalmazás: Bankkártyás csalás felderítése

Módos Dávid

Témavezető: Dr. Kovács Edith Alice

Szakdolgozatom fő témája a dimenziócsökkentés mely keretében egy mélytanulási algoritmust mutatok be, majd egy gépi tanulási algoritmust mutatok be, tárgyalom matematikai hátterüket, illetve alkalmazom őket banktranzakciós csalások felismerésére.

A bevezető részben a dimenziócsökkentésre térek ki külön tekintettel itt a dimenzióátkára („Curse of dimensionality”) melynél két példát is felhozok. A bevezetést a dimenziócsökkentési módszerek csoportosításával és azok népszerűségével zárom.

A második fejezetben az autoencoder mélytanulási algoritmussal foglalkozom. Bemutatom a szerkezeti felépítését és a matematikai hátterét. Ugyanitt tárgyalom a regularizált autoencodereket, a zajszűrő autoencodereket és néhány fontosabb alkalmazási területet, mint a képbesorolás, információ visszakeresés és nem utolsósorban anomália detektálás. Ez utóbbit a bemutatott alkalmazásban is felhasználom.

A neurális hálók legfontosabb beállításainak lehetőségére is kitértem, köztük az aktivációs függvények, a hibafüggvények és a mélytanulásban használatos gradiens csökkentési módszerekre. Az autoencodert különböző hiperparaméter (batch, epoch) beállítását is tárgyalom, ezeket futási eredmények összehasonlításával támasztom alá. A fejezet végén pedig az autoencoder és a PCA különbségeit és hasonlóságait tárgyalom.

A harmadik fejezetben az SNE és a t-SNE sztochasztikus szomszédsági leképezés algoritmust mutatok be, tárgyalom a matematikai hátterüket illetve egy adathalmazon illusztrálom a t-SNE működését.

A dolgozatom alkalmazási szempontból, leglátványosabb része a bankkártyás csalások detektálása. Egy tranzakciókat tartalmazó adathalmazon építék egy autoencodert, illetve ötvözöm az autoencoder és a t-SNE-ben rejlő lehetőségeket, így ezt a részt látványos és hatékony ábrákkal zárom.