

Toteutan projektin Pythonilla. Osaan hieman myös esimerkiksi C++:aa, mutta en mielestäni sillä tasolla, että osaisin välttämättä vertaisarvioida muiden opiskelijoiden sillä toteuttamia harjoitustöitä.

Projektin aiheena tulee olemaan käsinkirjoitettujen numeroiden tunnistus käyttäen neuroverkkoa. Verkon opettamisessa käytetään MNIST-tietokantaa, joka sisältää 28x28-kokoisia harmaasävykuvia. Yksittäinen kuva on esitetty bittijonona, jonka pituus on  $8 \times 784$  ja jokainen bittijonon tavu vastaa kokonaislukua välillä 0-255, missä arvo 0 vastaa valkoista ja 255 mustaa pikseliä. Tietokannassa on erillississä tiedostoissa koulutus- ja testauskuvat omina bittijonoinaan ja vastaavasti jokaisen kuvan luokittelut omina tiedostoinaan. Tulen todennäköisesti käyttämään jotain valmista kirjastoa näiden tiedostojen parsimiseen. Koulutuksen jälkeen algoritmille voidaan antaa syötteenä sopivalla tavalla esitetty 784-alkoinen vektori (esitettynä joko bittijonona tai muunnettuna suoraan esimerkiksi numpy-taulukoksi) verkon ulostulona on jokin luvuista 0-9 riippuen siitä, millaiseksi verkko sen luokittelee.

Harjoitustyön ydin on neuroverkon ja etenkin vastavirta-algoritmin (backpropagation) toteutus.

Algoritmin aikavaativuus on tyypillisesti luokkaa  $O(n \times t \times N)$ , missä  $n$  on koulutukseen käytettävien kierrosten (=epookkien) määrä,  $t$  koulutusesimerkkien määrä ja  $N$  verkon parametrien määrä. Takaisinvirtaus ei vaadi merkittävästi lisätilaa verrattuna itse verkon tallentamiseen, mutta takaisinvirtausalgoritmi vaatii sen, että lähempänä ulostuloa olevien kerrosten osittaisderivaatat säilytetään muistissa, jotta lähempänä syötettä olevien kerrosten virheet saadaan laskettua tehokkaasti. Takaisinvirtausalgoritmi ei vaadi mitään erityisiä tietorakenteita.

Lähteinä aion käyttää ainakin kurssisivulla tarjottuja aiheita, mutta uskoisin, että takaisinvirtausalgoritmin implementointi tulee vaatimaan jonkin verran lisälähteitä, jotta siihen pääsee kunnolla sisälle. Lisäksi uskoisin, että projektin edetessä tulen tutustumaan myös esimerkiksi yksikkötestaukseen tarkemmin. Lisäksi Poetry on itselleni uusi työkalu, joten varmasti senkin opetteluun menee oma aikansa. Olen tätä kirjoittaessani tutustunut jo esimerkiksi 3Blue1Brownin videosarjaan sekä lukenut kurssimateriaalissa mainittua Heli Tuomisen kirjoittamaa suomenkielistä kurssimateriaalia.

Opiskelen Tietojenkäsittelytieteen kandiohjelmassa ja projektin dokumentaatio tulee olemaan englanniksi.