

Päiväkirja

Contents

1 23.10.2023

Ajoin tosiaan ekat testit. Mukana oli kaikki vektorit, joille oli väestörakennedata ja vuosikate jossain määrin eheänä. Niitä tuli 6800. Ajoin ensimmäiseksi 100 epokkia 4-fold validationilla. Mallissa oli 2 keskikerrosta leveydellä 64 ja aktivaatiofunktio oli relu. Tulos näytti MAE (Mean Value Error) olevan 7 000 000.

Ajoin toisen testin vielä illalla. Kirjan esimerkkiä taas tapaillen tällä kertaa tutkittiin epokkien määrän vaikutusta. Muuten sama setuppi, mutta epokkeja 500 ja kerätään tulokset joka välissä analyysiä varten. Testit ottivat n 1,5h. Unohdin ottaa graafit ylös tällä kertaa, mutta niissä näkyi loiveneva lasku jostain 7400 \rightarrow 6800. Odotin että käyrä olisi lähtenyt nousuun ja olisi löytynyt selkeä ylisovituksen alku, mutta eipä näkynyt. Käyrä kuitenkin loiveni siihen malliin, että ei kovin paljoa parempaa tulosta saisi pelkällä epokkien määrän kasvattamisella.

2 27.10.2023

Konsultoituani chatGPT:tä ja muistettuani mm, että voisi BKT:n lisätä muutujaksi. Muovasin vähän hommia. Nyt pyörii vektoreina 3 peräkkäistä vuotta BKT:n kanssa. Otin myös suurimpien vuosikatteiden kunnat pois. Viimeisen vuoden vuosikate tulosteena. Vektorien määrä nyt n 4500. Pitäydyin vielä 4-foldissa, mutta lisäsin yhden kerroksen keskelle lisää ja asetin kerrosten leveydeksi 80. Tulokset toivottavasti tunnin sisään.

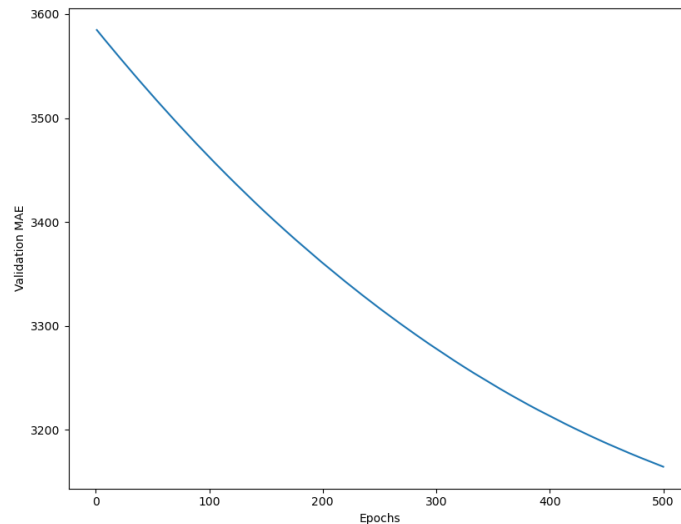


Figure 1: MAE - Epochs

Käyrän muoto on hyvin vastaava kuin aiemmalla. Selkeähköä laskua, mutta jänskättää laittaa epokkeja isoja määriä k-fold validationilla, kun sieltä vuotaa tietoa kaikei. Vuosikatteiden keskiarvo on tällä skaalalla 3440, joten tässä ei vielä taida olla liikaa hurraamista. Tämä testi söi 45 min. Seuraavaan testiin kerroksien leveydeksi 120 ja epokkeja 1000. Testi pyörimään ja lounaalle.

2 tuntia myöhemmin

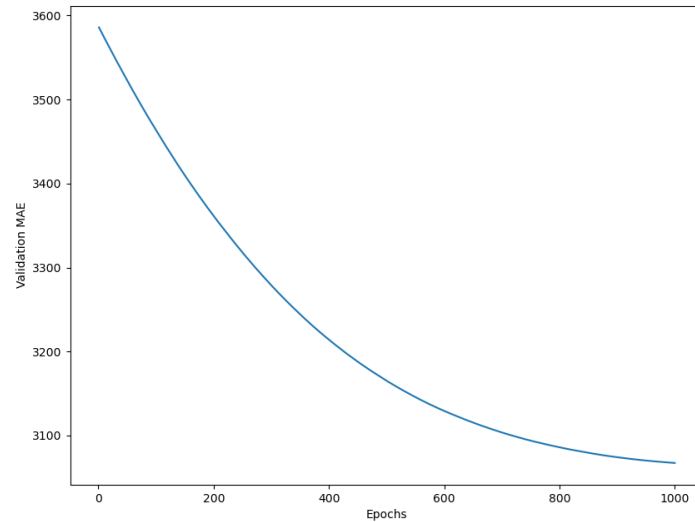


Figure 2: MAE - Epochs

Ei vieläkkään ylisovitusta, eikä vieläkkään virhe alle 3000. Muita juttuja pitäis keksiä. Onneksi niitä on tyrkyllä kanssa. Latailin vielä noita työssäkäyntitilastoja talteen oikeanmuotoisina exceleinä. Voisin askarrella ne vielä tietokantaan, mutten varmaan enää tänää aja enempää verkon koulutuksia läpi. Tuntuu, että toistaiseksi isoin muuttuja oli se, että otti isoimmat kunnat pois laskuista. Positiivista on, että asiat ovat jo testailtavissa.