Dataset: Star

NN Classification

Markus Kinn

I dette dokumentet kommer jeg til å diskutere valg jeg har tatt og sammenligne resultater fra ulike algoritmer. Jeg har ett dokument for hvert datasett.

For å forbedre resultatene på modellene har jeg prøvd å implementere så mye so mulig fra *Machine Learning Performance Improvement Cheat Sheet.*

**Generelt om datasettet:**

Datasettet er lagd for å kunne klassifisere stjerner ut ifra disse attributtene: Temperature, Luminosity, Radius, Absolute magnitude, Star color og Spectral class.

Datasettet har ingen feil.

Her måtte jeg også one-hot encode alle kategoriske variabler og deretter gjøre alle features om til tensors.

**Neural Network 1:**

**Før tuning:**

**Text

Description automatically generated**

**Etter tuning:**

**??????**

**Diskusjon av resultater:**

Modellen presterer bra, men sliter med noe over-fitting.

Jeg hadde opprinnelig problemer med at test scoren var høyere enn training scoren, men dette fikset jeg ved å først og fremst øke test størrelsen fra 0.1 tilbake til 0.25, men også ved å redusere batch størrelsen til 6 fra opprinnelige 10.

**Neural Network 2:**

**Før tuning:**

**Text

Description automatically generated**

**Etter tuning:**

**???????**

**Diskusjon av resultater:**

Modellen presterer generelt bedre på begge scoringene. Modellen har nesten fjernet all over-fitting grunnet tre drop out lag. Et drop out lag sender angitt prosent med 0 verdier til neste lag i nettverket for å skape en slags forvirring og dermed senke over-fitting. Jeg økte også batch size tilbake til 10

**Sammenligning av alle resultater:**

På dette datasettet presterte den andre arkitekturen desidert best. Jeg klarte å redusere over-fittingen, samtidig som jeg økte test scoren betraktelig.