Izvješće

**Uvod**

Izvješće pokriva opis aktivnosti projektnog zadatka, dobivene rezultate, naše komentare, zaključke te popis izvora.

**Opis projektnog zadatka**

Projektni zadatak je bio usporediti točnost ResNet-18 arhitekture naspram klasične CNN arhitekture pri treniranju nad CIFAR i MNIST skupom podataka. Glavna razlika između ovih mrežnih arhitektura je postojanje takozvanih shortcut connectiona (residual connection) između slojeva. Njihova uloga je da u sljedeći sloj prenose input trenutnog sloja, koji nije bio transformiran prolaskom kroz sloj. Proveli smo treniranje ResNet-18 mreže s različitim optimizerima (adam, stohastic gradient descent) te različitim batch size-om (128, 512) da saznamo kako ta dva koncepta utječu na njenu točnost. Treniranje je provedeno kroz 30 epoha i testirano na nezavisnom skupu.

**Dobiveni rezultati**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Točnost** | | | |
| optimizer | stohastic gradient descent | | Adam | |
| batch size | 128 | 512 | 128 | 512 |
| **CIFAR train** | 94.236% | 94.684% | 96.804% | 96.666% |
| **CIFAR test** | 88.820% | 88.010% | 90.630% | 88.190% |
| **MNIST train** | 99.417% | 99.458% | 99.287% | 99.312% |
| **MNIST test** | 98.960% | 98.670% | 99.260% | 98.210% |

**Komentari rezultata**

Uočena je razlika u točnosti mreže ovisno o skupu podataka. Točnost je veća nad skupom MNIST, nego nad skupom CIFAR. Smatramo da svaki skup podataka daje drugačiji n-dimenzionalni oblik koji utječe na traženje globalnog minimuma funckije i na djelovanje optimizera. Pretpostavljamo da skup podataka MNIST daje „pogodniji oblik“ za mrežu ResNet-18. Također, vidimo golim okom da je MNIST, skup koji prikazuje brojke, puno lakše raspoznati nego što je to u slučaju CIFAR-a koji prikazuje neke složenije objekte.

Za skup CIFAR (vrijedi i za SGD i za Adam) primjećujemo malo veću točnost nad testnim skupom kad koristimo batch size 128 u odnosu na 512 (+0.810% sa SGD, +2.440% s Adam). Za MNIST također primjećujemo bolju točnost na testnom skupu korištenjem batch size 128 (+0.290%, +0.950%). Iz priloženog vidimo značajnu razliku, te možemo komentirati da je veća točnost nad podacima koje nismo učili (testni skup) kada koristimo batch size 128. Razlog leži u tome što manji batch više šumi i tako ne dozvoljava modelu da se prenauči.

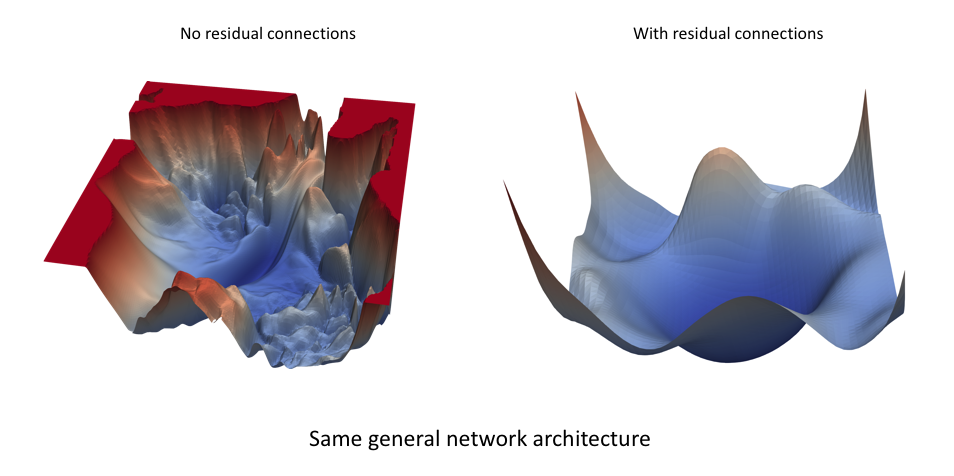
Nadalje, nad oba skupa bolju točnost postižemo korištenjem Adam optimizera, stoga možemo zaključiti da on brže pronalazi minimum.

**Usporedba točnosti ResNet-18 modela i osnovnog CNN modela**

--usporediti sa CNN i plottati--

**Zaključak**

Bolja performansa ResNet-18 mreže nad klasičnom CNN mrežom je posljedica postojanja shortcut connectiona. Njihov učinak na loss landscape neuronske mreže je vidljiv na priloženoj slici. Traženje minimuma i optimizacija funkcije gubitka je manje zahtjevna nad glatkijim loss landscape-om i manja je vjerojatnost da završimo u nekom lokalnom minimumu.



**Izvori:**

--linkati samo znanstvene radove--