
Wielowarstwowa sieć jednokierunkowa MLP w Tensorflow

Piotr Grzybowski
220931
p.grzybowski2@gmail.com

Abstract

Celem ćwiczenia jest wprowadzenie do biblioteki Tensorflow, przypomnienie podstawowej sieci MLP, wpływu hiperparametrów na uczenie oraz ocena otrzymywanych wyników. Sieć powinna rozwiązywać problem klasyfikacji obrazów ze zbioru CIFAR-10.

1 Zbadanie regularyzacji na proces uczenia

W tej sekcji została poddana analiza procesu uczenia przy braku regularyzacji funkcji kosztu oraz zastosowaniu regularyzacji l2.

1.1 Eksperyment 1

W tym eksperymencie zastosowano konfigurację jak w tabeli 3.

Tabela 1: Konfiguracja eksperymentu pierwszego.

Model	Metoda optymalizacji	Stała uczenia	Wielkość paczki	Regularyzacja
[500]	Momentum 0.9	0.001	64	Brak

Na rysunku 1 znajduje się wykres wartości funkcji straty w czasie uczenia dla danych treningowych i testowych. Można zauważyć, że wraz z upływem kolejnych epok model dopasowuje się bardziej do danych treningowych co sugeruje dążenie wartości funkcji kosztu do zera oraz zwiększanie się wartości funkcji kosztu dla danych testowych.



Rysunek 1: Wykres zależności wartości funkcji straty w czasie dla danych treningowych i testowych.

Na rysunku 2 znajduje się wykres skuteczności modelu w czasie uczenia dla danych treningowych i testowych. Można zauważyć, że wraz z upływem kolejnych epok model dopasowuje się bardziej do danych treningowych co sugeruje dążenie metryki skuteczności do 1.0 bez poprawy skuteczności dla danych testowych.



Rysunek 2: Wykres zależności skuteczności modelu w czasie dla danych treningowych i testowych.

Osiągnięta skuteczność to około 0.36 na zbiorze testowym. W następnym eksperymencie zostanie zastosowana regularyzacja L2 ze współczynnikiem $\beta = 0.1$

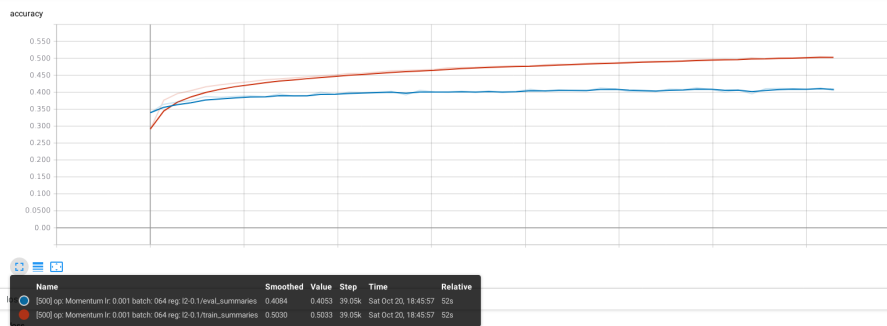
1.2 Eksperyment 2

W tym eksperymencie zastosowano konfigurację jak w tabeli 3.

Tabela 2: Konfiguracja eksperymentu pierwszego.

Model	Metoda optymalizacji	Stała uczenia	Wielkość paczki	Regularyzacja
[500]	Momentum 0.9	0.001	64	L2 $\beta = 0.1$

Na rysunku 3 znajduje się wykres wartości funkcji straty w czasie uczenia dla danych treningowych i testowych. Można zauważyć, że wraz z upływem kolejnych epok model dopasowuje się bardziej do danych treningowych co sugeruje dążenie wartości funkcji kosztu do zera oraz zwiększanie się wartości funkcji kosztu dla danych testowych.



Rysunek 3: Wykres zależności wartości funkcji straty w czasie dla danych treningowych i testowych.

Na rysunku 4 znajduje się wykres skuteczności modelu w czasie uczenia dla danych treningowych i testowych. Można zauważyć, że wraz z upływem kolejnych epok model dopasowuje się bardziej do danych treningowych co sugeruje dążenie metryki skuteczności do 1.0 bez poprawy skuteczności dla danych testowych.



Rysunek 4: Wykres zależności skuteczności modelu w czasie dla danych treningowych i testowych.

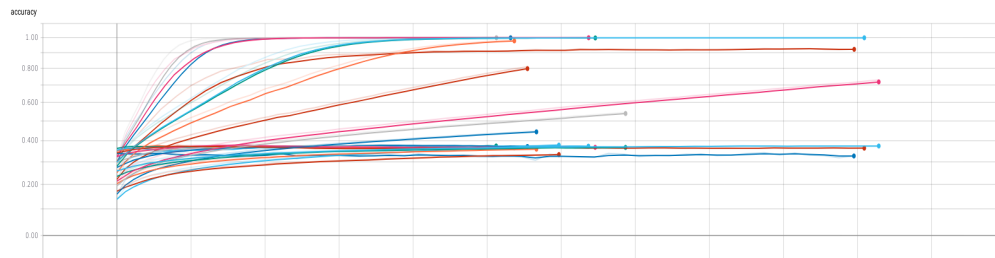
Dzięki zastosowaniu regularyzacji udało się wyrównać wartość funkcji straty dla zbiorów treningowego i testowego w okolicy wartości 2.0. Skuteczność modelu na zbiorze testowym była większa niż 0.40 przy równoczesnej niewiele większej skuteczności na zbiorze treningowym. Na tej podstawie można wyciągnąć wniosek, że model lepiej generalizuje dane.

W następnym eksperymencie przedstawione zostaną wartości przy zastosowaniu i braku regularyzacji dla różnych przedziałów hiperparametrów.

1.3 Eksperyment 3

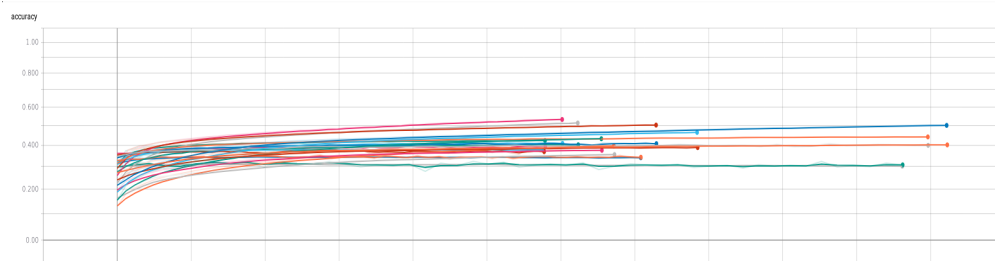
- Model: [500]
- Metoda optymalizacji: Momentum 0.9
- Stała uczenia: [0.1, 0.01, 0.001, 0.0001]
- Wielkość paczki: [16, 32, 64, 128, 256]
- Regularyzacja: [Brak, L2 0.1, L2 0.5]

Na rysunku 5 znajduje się wykres skuteczności modelu w czasie dla wielu kombinacji hiperparametrów przy braku regularyzacji. W prosty sposób można rozróżnić serie treningowe (dążące do 1.0) oraz serie testowe (te w okolicach 0.4).



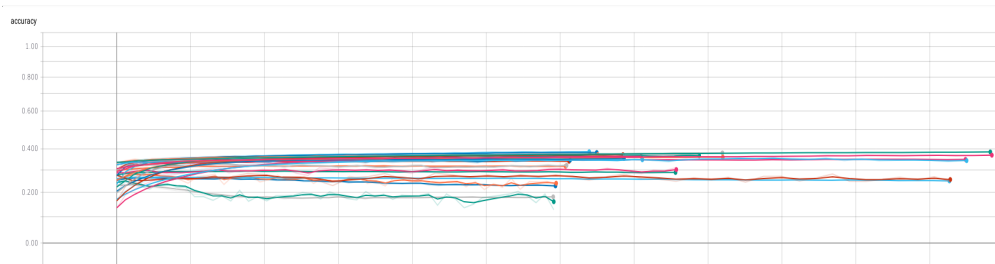
Rysunek 5: Wykres zależności skuteczności modelu w czasie dla danych treningowych i testowych.

Na rysunku 6 znajduje się wykres skuteczności modelu w czasie przy zastosowaniu regularyzacji L2 z parametrem $\beta = 0.1$. Wszystkie serie dla danych treningowych i testowych zawierają się w przedziale od 0.35 do 0.45.



Rysunek 6: Wykres zależności skuteczności modelu w czasie dla danych treningowych i testowych.

Na rysunku 7 znajduje się wykres skuteczności modelu w czasie uczenia dla danych treningowych i testowych dla wielu kombinacji hiperparametrów przy zastosowaniu regularyzacji L2 z parametrem $\beta = 0.1$. Wszystkie serie dla danych treningowych i testowych zawierają poniżej 0.40. Została zastosowana zbyt duża regularyzacja. Najlepszy osiągnięty wynik na zbiorze testowym to 0.37.



Rysunek 7: Wykres zależności skuteczności modelu w czasie dla danych treningowych i testowych.

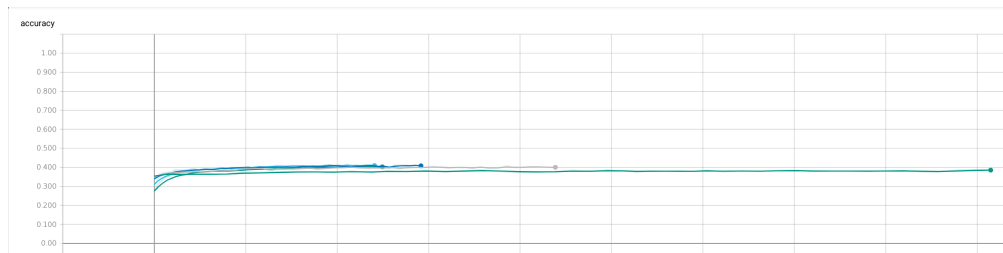
2 Zbadanie wpływu wielkości paczek na proces uczenia

W tej sekcji została poddana analiza procesu uczenia przy różnej wielkości paczek.

Tabela 3: Konfiguracja eksperymentu pierwszego.

Model	Metoda optymalizacji	Stała uczenia	Wielkość paczki	Regularyzacja
[500]	Momentum 0.9	0.001	16, 32, 64, 128, 256	L2 0.1

Na rysunku 13 znajduje się wykres skuteczności modelu w czasie uczenia dla danych treningowych i testowych dla różnej wielkości paczki. Można zauważyć że skuteczność na zbiorze testowym nie różni się bardzo, natomiast należy zwrócić uwagę na czas uczenia.



Rysunek 8: Wykres zależności skuteczności modelu w czasie dla danych testowych.

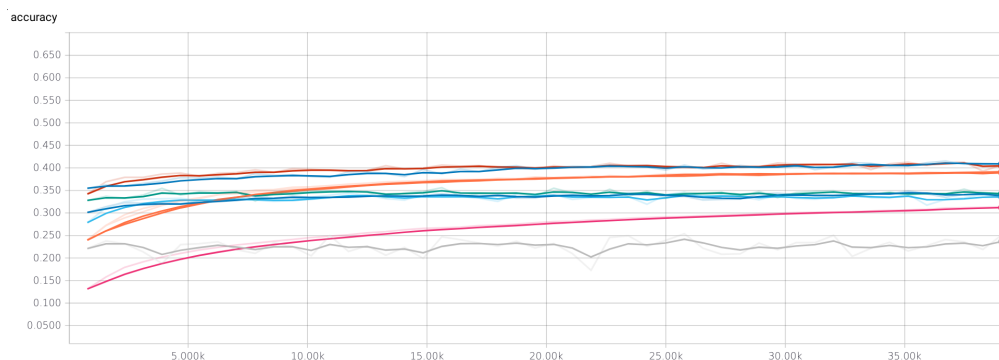
Wraz z zastosowaniem mniejszej paczki czas uczenia zwiększa się znacznie. Stosowanie większych paczek nie skutkowało znaczącej poprawy czasu uczenia.

- paczka 8 - 4m 59s
- paczka 16 - 2m 43s
- paczka 32 - 1m 18s
- paczka 64 - 54s
- paczka 128 - 44s
- paczka 256 - 43s

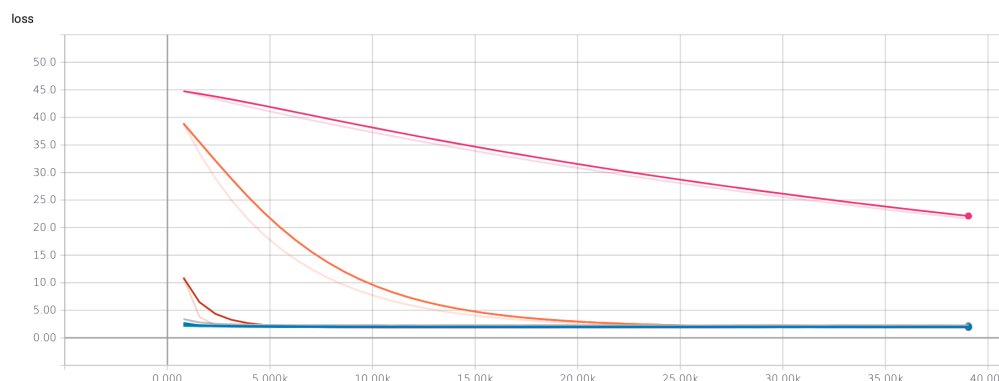
3 Zbadanie wpływu metody optymalizacji i stałej uczenia

W tej sekcji została poddana analiza różnych metod optymalizacji wraz z różnymi stałymi uczenia w celu uzyskania najlepszego wyniku. Konfiguracja poniżej

- Model: [500]
- Metoda optymalizacji: Momentum, Adam, SGD
- Stała uczenia: [0.01, 0.001, 0.0001]
- Wielkość paczki: 64
- Regularyzacja: L2 0.1



Rysunek 9: Wykres zależności skuteczności modelu w czasie dla danych testowych.



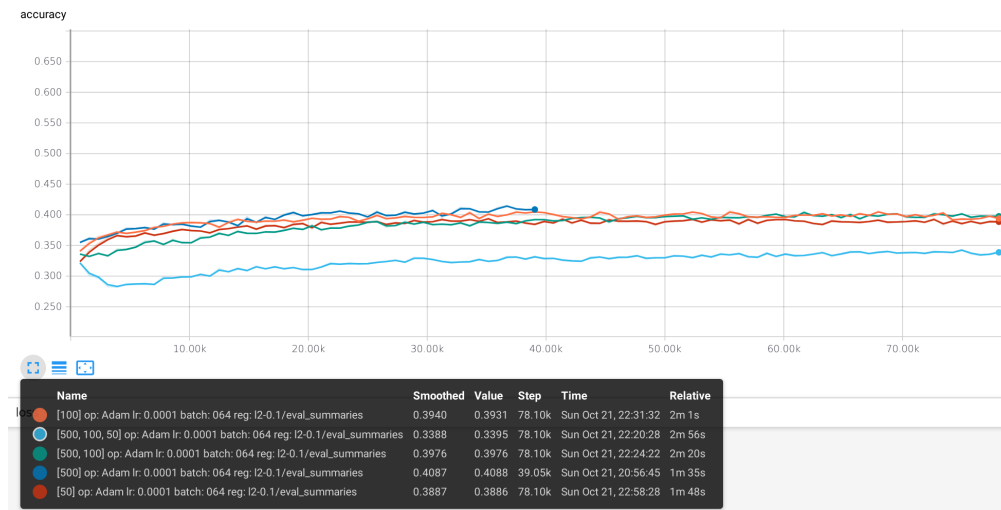
Rysunek 10: Wykres zależności skuteczności modelu w czasie dla danych testowych.

Name	Smoothed	Value	Step	Time	Relative
[500] op: Adam lr: 0.0001 batch: 064 reg: l2-0.1/eval_summaries	1.919	1.918	39.05k	Sun Oct 21, 20:56:45	1m 35s
[500] op: Adam lr: 0.001 batch: 064 reg: l2-0.1/eval_summaries	2.022	2.032	39.05k	Sun Oct 21, 19:54:24	1m 23s
[500] op: Adam lr: 0.01 batch: 064 reg: l2-0.1/eval_summaries	2.281	2.272	39.05k	Sun Oct 21, 20:55:04	1m 30s
[500] op: Momentum lr: 0.0001 batch: 064 reg: l2-0.1/eval_summaries	1.901	1.893	39.05k	Sat Oct 20, 18:50:00	56s
[500] op: Momentum lr: 0.01 batch: 064 reg: l2-0.1/eval_summaries	2.103	2.101	39.05k	Sat Oct 20, 18:42:06	50s
[500] op: RMSProb lr: 0.001 batch: 064 reg: l2-0.1/eval_summaries	2.057	2.058	39.05k	Sun Oct 21, 19:59:23	1m 11s
[500] op: SGD lr: 0.0001 batch: 064 reg: l2-0.1/eval_summaries	22.11	21.63	39.05k	Sun Oct 21, 20:46:09	1m 25s
[500] op: SGD lr: 0.001 batch: 064 reg: l2-0.1/eval_summaries	1.900	1.893	39.05k	Sun Oct 21, 20:05:23	1m 21s
[500] op: SGD lr: 0.01 batch: 064 reg: l2-0.1/eval_summaries	1.961	1.966	39.05k	Sun Oct 21, 20:44:40	1m 13s

Rysunek 11: Wykres zależności skuteczności modelu w czasie dla danych testowych.

Najwolniej zbiega SGD najszybciej Adam. Najszybciej działa momentum, najwolniej działa Adam.

4 Zbadanie wpływu liczby warstw i wielkości modelu na proces uczenia



Rysunek 12: Wykres zależności skuteczności modelu w czasie dla danych testowych.

Zbyt skomplikowany model dłużej się uczy. Brak wyraźnej różnicy między modelem z jedną warstwą i 500 jednostkami od modelu z 50 neuronami lub dwiema warstwami.

5 Prezentacja najlepszego modelu

[436	74	56	10	32	16	47	33	177	118]
[42	457	8	13	33	32	71	48	64	232]
[115	33	247	74	153	59	164	76	27	51]
[61	48	63	166	77	138	213	77	33	123]
[73	27	92	35	350	45	195	82	34	65]
[50	46	67	108	75	303	138	77	48	86]
[19	25	56	54	104	34	569	40	16	79]
[54	64	38	51	103	49	76	390	32	139]
[128	105	12	23	23	35	30	15	488	141]
[34	162	9	24	19	18	71	54	84	524]]

Rysunek 13: Tablica pomyłek dla zbioru testowego dla najlepszego modelu.