

Najlepsze wyniki są, gdy jest dodana kolejna warstwa neuronów

Zmienienie optymalizatora na Adama sprawia, że przy tych samych wartościach współczynnika uczenia model się nie uczy jako powinien przy tej samej ilości epok(100) (następuje przy początkowych epokach gwałtowny spadek funkcji błędu (walidacyjnej i treningowej), a potem funkcja bładu jest praktycznie stała i niewielkimi wahaniami), reasumując model się szybko uczy na początku, ale potem już nie jest w stanie zaktualizować tak wag sieci by zmniejszyć wartość funkcji błędu.

Gdy używamy optymalizatora sgd (hiperparametry 1, 2 i 3) spadek funkcji błędu (walidacyjnej i treningowej) jest widoczny podczas całego procesu uczenia (na przełomie wszystkich epok)

Zmiana współczynnik uczenia przy optymalizatorze SGD sprawia, że model uczy się wcześniej (hiperparametry 2 i 3) niż przy domyślnym współczynniku uczenia zastosowanym w optymalizatorze SGD.

Dodanie kolejnej warstwy gęstej zwiększa szybkość uczenia kosztem czasu uczenia i zwiększa też dokładność modelu.