

Podsumowanie modeli trenowanych na param.: bs:64, ep: 100, lr: 0.001

Model_A (MLP - ReLU):

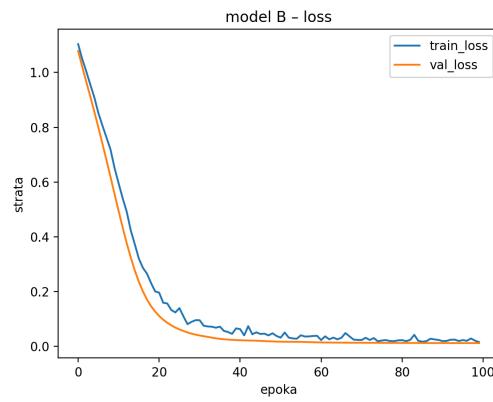
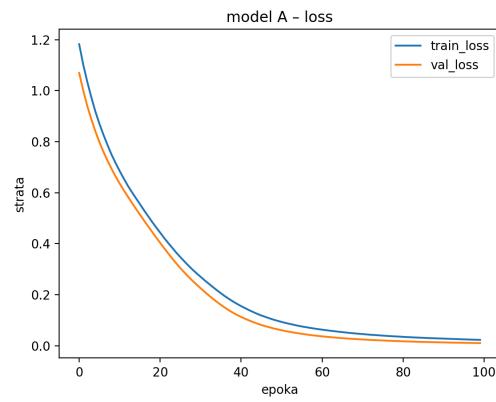
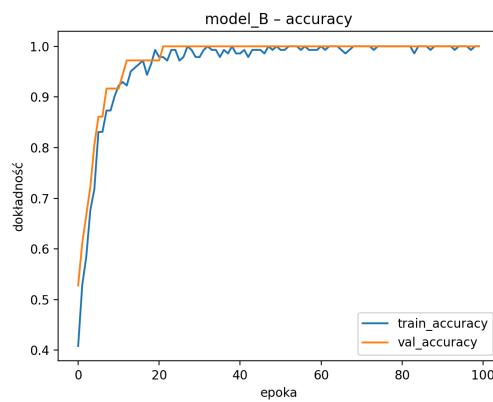
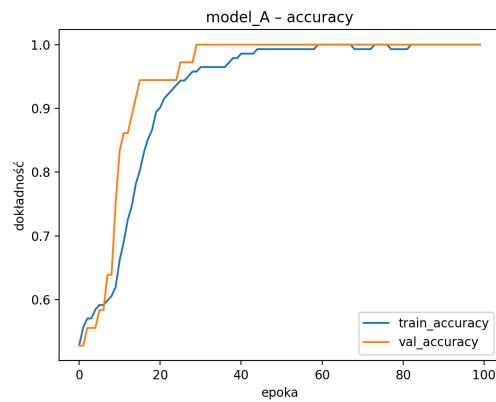
- ilość warstw: 4
- Input(13) → Dense(32, ReLU) → Dense(16, ReLU) → Dense(3, Softmax)
- optymalizator: adam
- brak regularyzacji
- funkcja aktywacji warstw: relu
- prosty, płaski model o małej liczbie paramaterów

Model_B (MLP - GELU + L2 + dropout):

- ilość warstw: 7
- optymalizator: adam
- inicjalizacja: glorot uniform
- funkcja aktywacji warstw: gelu
- regularizacja L2 + dropout

Oba modele:

- funkcja straty: `categorical_crossentropy`
- funkcja aktywacji ostatniej warstwy: softmax



Wnioski:

Oba modele osiągają pełną dokładność na zbiorze treningowym. Model B dzięki regularizacji L2 i dropout ma łagodniejszą krzywą val. Model B jest potencjalnie bardziej odporny na przeuczenie.