

Ex.1

Count the number of parameters in the previously defined model.

Given n inputs and m outputs, the total number of weights is $n*m$. Adding a bias for each output node, the total number of weights is $(n+1)*m$.

$$32 * 32 * 3 * 100 + 100 + 100 * 10 + 10 = 308,310$$

Ex.2

In the following examples, two transformations are applied to the train/test sets: 1. ToTensor(), 2. Normalize([0.49139968, 0.48215841, 0.44653091], [0.24703223, 0.24348513, 0.26158784]); batch size is set to 256 and Adam optimizer is used.

Number of layers/Number of neurons per layer

SmallNet

[3072] → [150] → Sigmoid → [10]

Learning rate: 5e-4

Weights initialization: default

Results → Test -- Loss: 1.560, Acc: 46.95 %

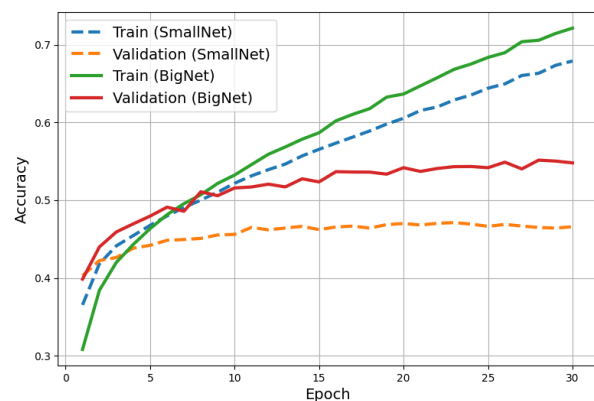
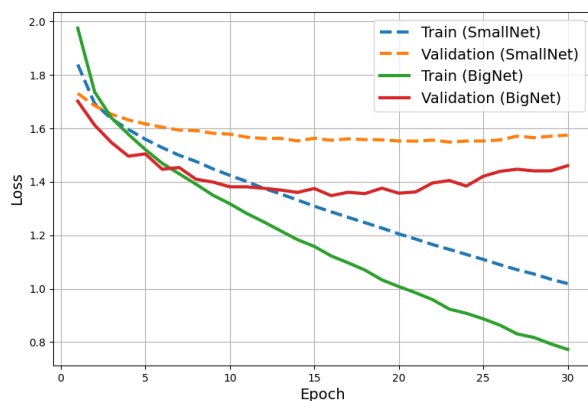
BigNet

[3072] → [2500] → ReLU → Dropout(0.3) → [1000] → ReLU → Dropout(0.3) → [250] → ReLU → Dropout(0.3) → [100] → ReLU → Dropout(0.3) → [10]

Learning rate: 5e-4

Weights initialization: xavier_normal_/fill_(0.1) → weight/bias

Results → Test -- Loss: 1.341, Acc: 54.09 %



Learning rate

BigNet

[3072] → [2500] → ReLU → Dropout(0.3) → [1000] → ReLU → Dropout(0.3) → [250]
→ ReLU → Dropout(0.3) → [100] → ReLU → Dropout(0.3) → [10]

Learning rate: 1e-3, 5e-4, 3e-7

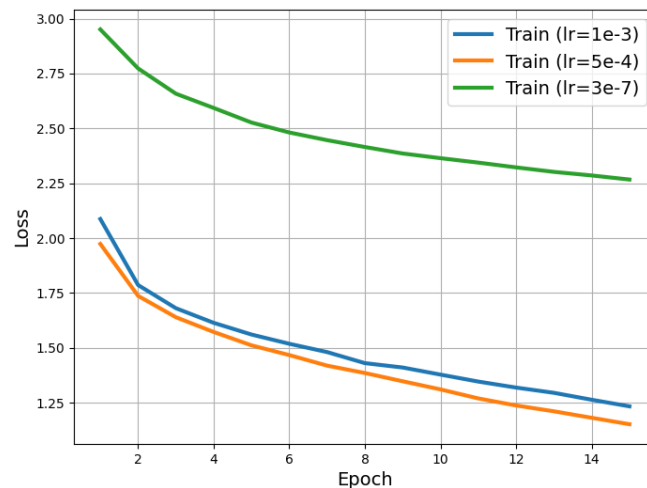
Weights initialization: xavier_normal_/fill_(0.1) → weight/bias

Epochs: 15

Results (lr=1e-3) → Test -- Loss: 1.374, Acc: 52.33 %

Results (lr=5e-4) → Test -- Loss: 1.331, Acc: 53.63 %

Results (lr=3e-7) → Test -- Loss: 2.009, Acc: 29.25 %



Weights Initialization

BigNet

[3072] → [2500] → ReLU → Dropout(0.3) → [1000] → ReLU → Dropout(0.3) → [250]
→ ReLU → Dropout(0.3) → [100] → ReLU → Dropout(0.3) → [10]

Learning rate: 5e-4

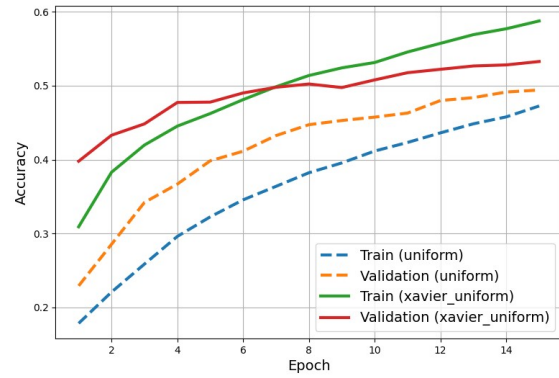
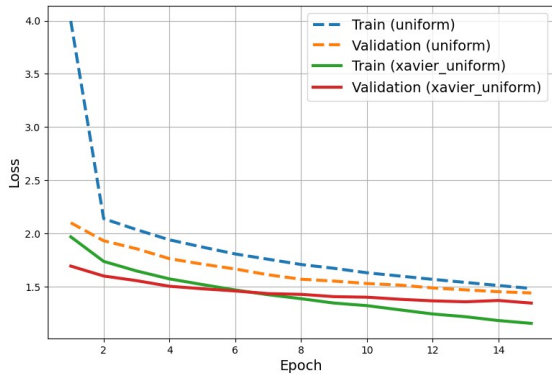
Weights initialization n. 1: uniform_(-0.12, 0.12)/constant_(0.01) → weight/bias

Weights initialization n. 2: xavier_uniform_/constant_(0.01) → weight/bias

Epochs: 15

Results (WI1) → Test -- Loss: 1.419, Acc: 50.95 %

Results (WI2) → Test -- Loss: 1.332, Acc: 53.88 %



Dropout

BigNet

[3072] → [2500] → ReLU → Dropout(X) → [1000] → ReLU → Dropout(X) → [250] → ReLU → Dropout(X) → [100] → ReLU → Dropout(X) → [10]

Learning rate: 5e-4

Weights initialization: xavier_uniform/constant_(0.01) → weight/bias

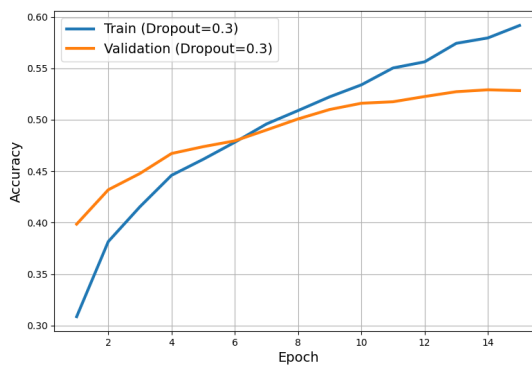
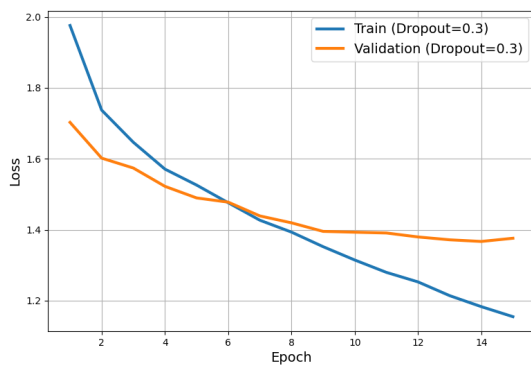
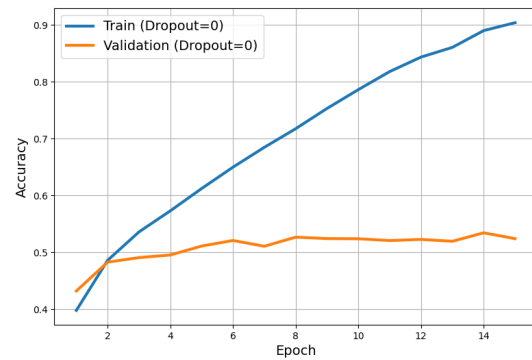
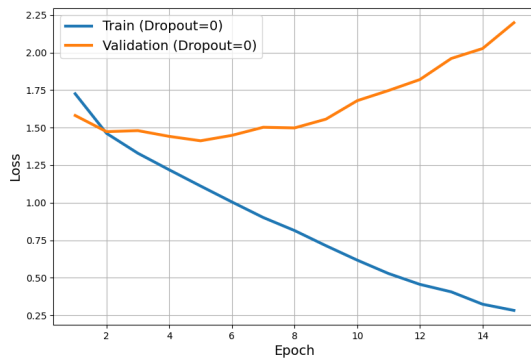
Dropout rates: 0, 0.3, 0.9

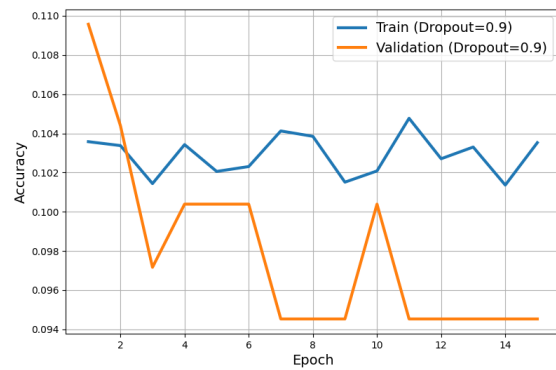
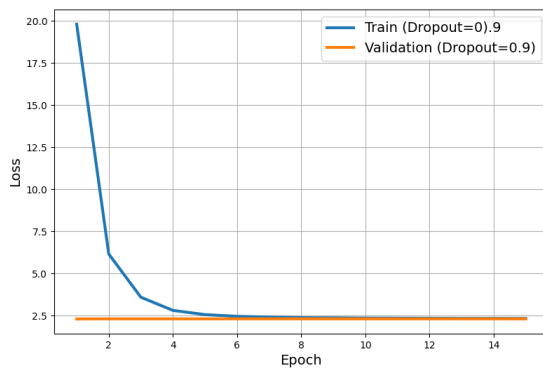
Epochs: 15

Results (Dropout=0) → Test -- Loss: 1.387, Acc: 52.20 %

Results (Dropout=0.3) → Test -- Loss: 1.349, Acc: 53.16 %

Results (Dropout=0.9) → Test -- Loss: 2.298, Acc: 10.88 %





Activation function

BigNet2

[3072] → [3000] → Act. Function → Dropout(0.2) → [2000] → Act. Function → Dropout(0.2) → [1500] → Act. Function → Dropout(0.2) → [1000] → Act. Function → Dropout(0.2) → [500] → Act. Function → Dropout(0.2) → [100] → Act. Function → Dropout(0.2) → [10]

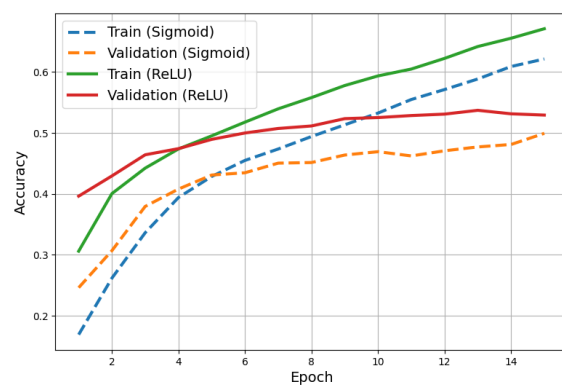
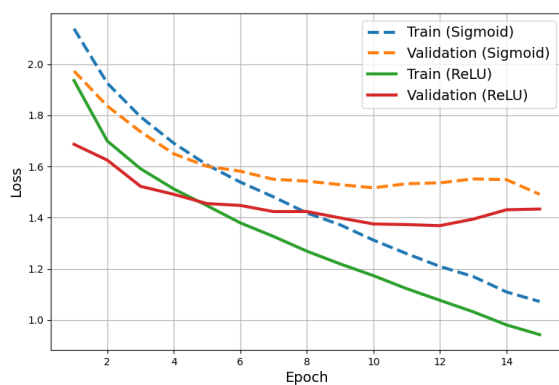
Learning rate: 5e-4

Weights initialization: xavier_uniform_/constant_(0.01) → weight/bias

Epochs: 15

Results (Act. Function: Sigmoid) → Test -- Loss: 1.471, Acc: 50.85 %

Results (Act. Function: ReLU) → Test -- Loss: 1.359, Acc: 53.18 %



Good Model

[3072] → [2500] → ReLU → Dropout(0.5) → [1000] → ReLU → Dropout(0.5) → [10]

Weights initialization: xavier_normal_/constant_(0.01) → weight/bias
Data Augmentation: RandomRotation(10),RandomHorizontalFlip()
Learning rate: 5e-4

Results → Test -- Loss: 1.268, Acc: 55.28 %

