Assignment 3: Report bonus

a) Improvements

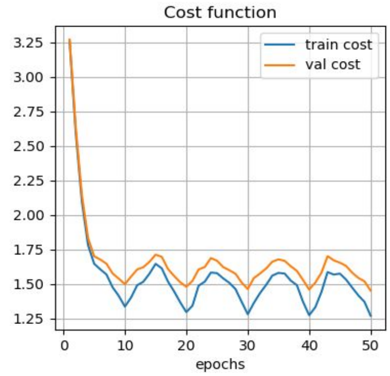
	lambda	batch	Architecture	eta_min	eta_max	n_s	acc_train	acc_valid	acc_test
Basic	0.005164	100	3NN	1e-5	0.1	2250	0.61871	0.5372	**0.5355**
1) Augmentation	0.005164	100	3NN	1e-5	0.1	2250	0.6091	0.545	**0.5351**
2) BN after ReLu	0.005164	100	3NN	1e-5	0.1	2250	0.6326	0.5492	**0.5376**
3) Dropout	0.005164	80	3NN	1e-5	0.1	2250	0.67675	0.5424	**0.5357**
4) All above + deeper NN	0.001164	100	4NN	1e-5	0.1	2250	0.67675	0.5628	**0.5523**

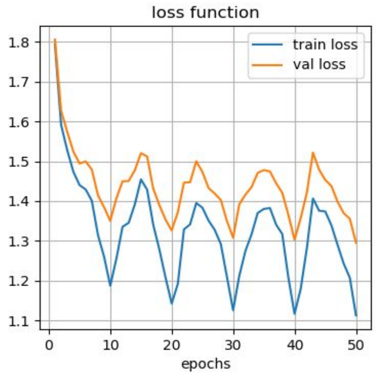
1) Augmentation with Gauss noise(std=0.125)

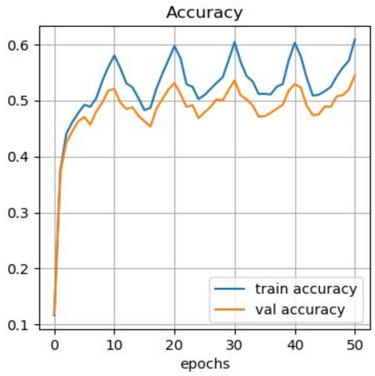
```
nn = Network()
    1.  nn.add(Dense(n_in, 50)), nn.add(ReLU()), nn.add(BatchNorm(50))
    2.  nn.add(Dense(50, 50)), nn.add(ReLU()), nn.add(BatchNorm(50))
    3.  nn.add(Dense(50, 10))
```

N_batch = 100
n_s = 5*n / n_batch, eta_min = 1e-5, eta_max= 1e-1
n_epochs = (2*n_s*n_batch/n)*5 = 5 cycles
λ = 0.005164

Train set accuracy: 0.6091
Validation set accuracy: 0.545
Test set accuracy: 0.5351







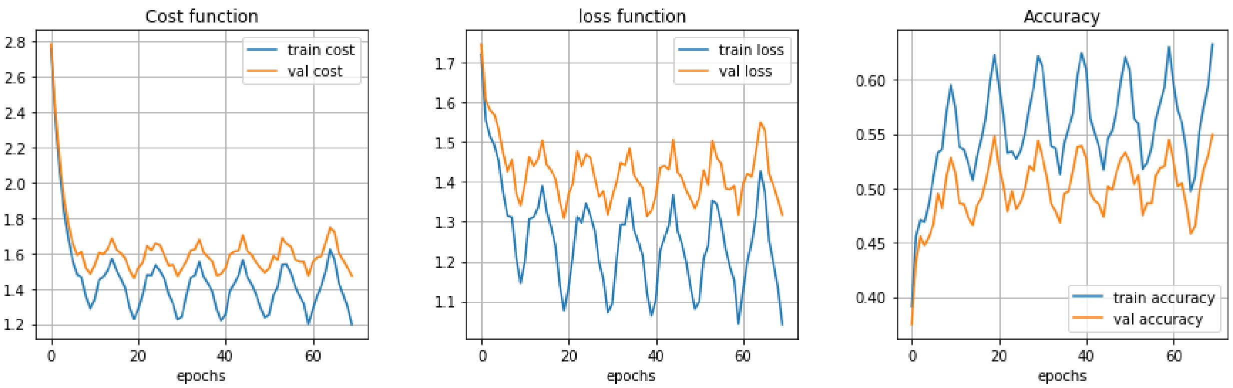
2) BN after ReLu

nn = Network()

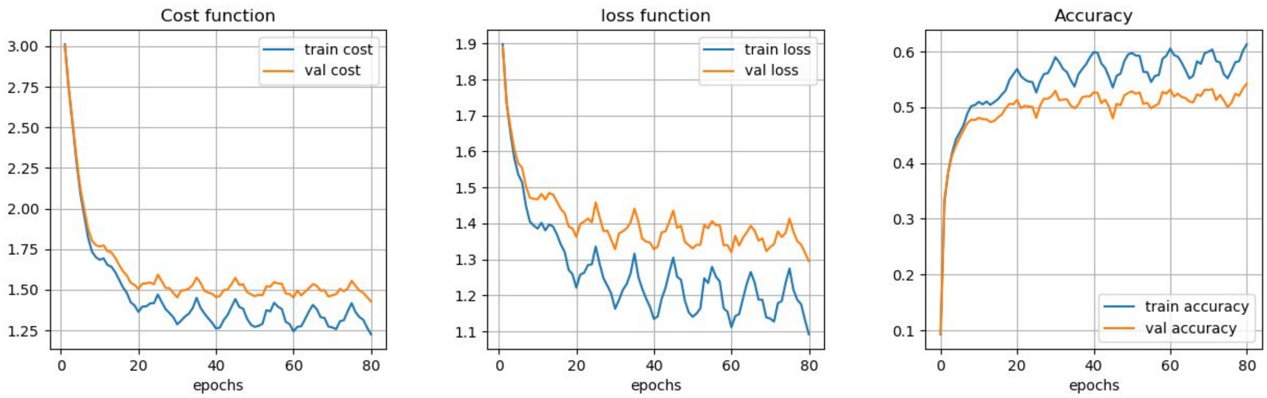
4. nn.add(Dense(n_in, **50**)), nn.add(ReLU()), nn.add(**BatchNorm**(**50**))
5. nn.add(Dense(**50**, **50**)), nn.add(ReLU()), nn.add(**BatchNorm**(**50**))
6. nn.add(Dense(**50**, **10**))

Train set accuracy: 0.6326222
Validation set accuracy: 0.5492
Test set accuracy: 0.5376

N_batch = 100
n_s = 5*n / n_batch, eta_min = 1e-5, eta_max= 1e-1
n_epochs = (2*n_s*n_batch/n)*7 = 7 cycles
λ = 0.005164



3) Dropout



nn = Network()

1. nn.add((**Dropout(0.3)**), nn.add(Dense(n_in, **50**)), nn.add(ReLU()), nn.add((**BatchNorm**(**50**))
2. nn.add(Dense(**50**, **50**)), nn.add(ReLU()), nn.add((**BatchNorm**(**50**))
3. nn.add((**Dropout(0.3)**), nn.add(Dense(**50**, **10**))

Train set accuracy: 0.61306
Validation set accuracy: 0.5424
Test set accuracy: 0.5357

N_batch = 100
n_s = 5*n / n_batch, eta_min = 1e-5, eta_max= 1e-1
n_epochs = (2*n_s*n_batch/n)*8 = 8 cycles
λ = 0.005164

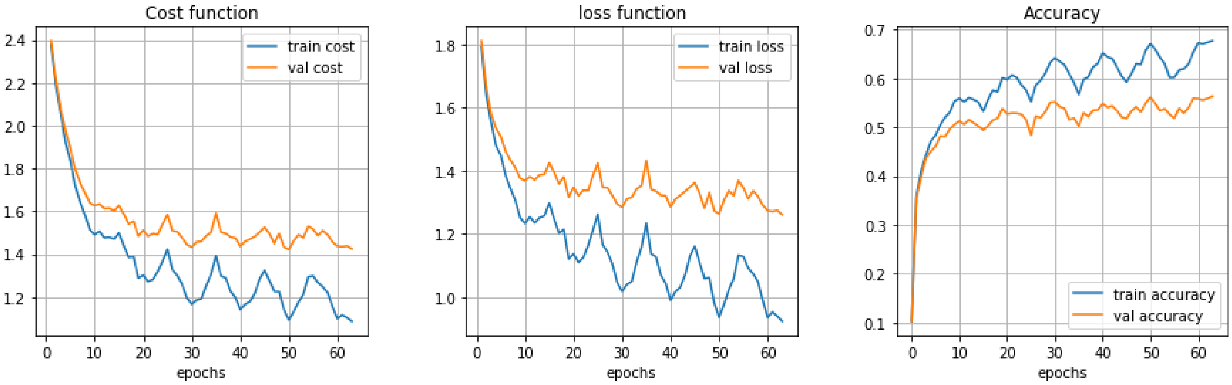
4) All improvements combined

nn = Network()

1. nn.add(**Dropout(0.3)**), nn.add(Dense(n_in, **100**)), nn.add(ReLU()), nn.add(**BatchNorm(100)**)
2. nn.add(Dense(**100**, **50**)), nn.add(ReLU()), nn.add(**BatchNorm(50)**)
3. nn.add(**Dropout(0.3)**), nn.add(Dense(**50, 100**)), nn.add(ReLU()), nn.add(**BatchNorm(100)**)
4. nn.add(**Dropout(0.2)**), nn.add(Dense(**100**, **10**))

Train set accuracy: 0.67675
Validation set accuracy: 0.5628
Test set accuracy: 0.5523

N_batch = 100
n_s = 5*n / n_batch, eta_min = 1e-5, eta_max= 1e-1
n_epochs = (2*n_s*n_batch/n)*6 = 6 cycles
λ = 0.001164



In order to achieve this accuracy I trained for 4 more epochs after the 6th cycle with eta=1e-7.