CSC 420 A2

Yinjun Zheng

Oct, 2021

1 Part 1

1.1 Q1

Let X_0 be the input, X_i be the output after i^{th} layer
Let W_i be the weight for layer i, b_i be the bias for layer i
Let W_σ, b_σ be the coefficients for activation function
then we have:

$$\begin{aligned} X_{n+1} &= \sigma(W_n * X_n + b_n) \\ &= W_\sigma * (W_n * X_n + b_n) + b_\sigma \\ &= (W_{\sigma * W_n) X_n + W_\sigma * b_n + b_\sigma} \\ &= W' * X_n + b' \end{aligned}$$

Therefore, it is still a linear relationship between output and input after the activation function. We have:

$$\begin{aligned} X_{n+1} &= \sigma o(W_n, n_n) X_n \\ &= \sigma o(W_n, n_n) o \sigma o(W_{n-1}, n_{n-1}) o X_{n-1} \\ &= ... \\ &= \sigma o(W_n, n_n) o \sigma o(W_{n-1}, n_{n-1}) o ... o \sigma o(W_0, n_0) o X_0 \\ &= W' X_0 + b' \end{aligned}$$

So no matter how many layers there are, the output is still a linear combination of the input, which is equal to the effect of just one layer.
The number of layers has effectively no impact on the network.

1.2 Q2

The forward pass is:

$$
\begin{aligned}
X_{12} &= W_1X_1 + W_2X_2 \\
X_5 &= \sigma(X_{12}) \\
X_{34} &= W_3X_3 + W_4X_4 \\
X_6 &= \sigma(X_{34}) \\
X_7 &= W_5X_5 + W_6X_6 \\
y &= \sigma(X_7) \\
L_2 &= \|y - \hat{y}\|_2^2
\end{aligned}
$$

The value for each step is:

$$
\begin{aligned}
X_{12} &= 0.75 * 0.9 + 0.63 * 1.1 = 1.368 \\
X_5 &= (1.368) = 0.797 \\
X_{34} &= 0.24 * 0.3 + -1.7 * 0.8 = -1.432 \\
X_6 &= \sigma(-1.432) = 0.193 \\
X_7 &= 0.8 * 0.797 + -0.2 * 0.193 = 0.599 \\
y &= \sigma(0.599) = 0.645
\end{aligned}
$$

Therefore, by backward propagation we get:

$$
\begin{aligned}
\frac{\partial L}{\partial w_3} &= \frac{\partial L}{\partial y} * \frac{\partial y}{\partial w_3} \\
&= \frac{\partial L}{\partial y} * \frac{\partial y}{\partial x_7} * \frac{\partial x_7}{\partial x_6} * \frac{\partial x_6}{\partial x_{34}} * \frac{\partial x_{34}}{\partial w_3} \\
&= 2 * \|y - \hat{y}\| * y * (1 - y) * w_6 * x_6 * (1 - x_6) * x_3 \\
&= 2 * \|0.645 - 0.5\| * 0.645 * (1 - 0.645) * -0.2 * 0.193 * (1 - 0.193) * -0.3 \\
&= 0.00062
\end{aligned}
$$

Therefore, $\frac{\partial L}{\partial w_3} = 0.00062$.

1.3 Q3

Input size is $12 * 12 * 50$, then we have $(12 - 4 + 2)/2 + 1 = 6$.
So output size is $6 * 6 * 20$

Case 1: If we consider the bias, then:
Layer C has $2 * 6 * 6 * 4^2 * 20 * 50 = 1152000$ flops.
Layer P has $4 * 4 * 20 * (3^2 - 1) = 2560$ flops.
So there are $1152000 + 2560 = 1154560$ flops.
Case 2: If we do not consider the bias, then:
there are $1154560 + 6 * 6 * 20 * 50 = 1190560$ flops.
Therefore, there are 1154560 flops with bias, and 1190560 flops without bias.

1.4 Q4

We have the kernel size $K = 5$
So the trainable parameters for each layer is listed below:

$$
\begin{aligned}
C1 &: 5 * 5 * 1 * 6 = 150 \\
S2 &: 0 \\
C3 &: 5 * 5 * 6 * 16 = 2400 \\
S4 &: 0 \\
C5 &: 5 * 5 * 16 * 120 + 120 = 48120 \\
C6 &: 120 * 84 + 84 = 10164 \\
C7 &: 84 * 10 + 10 = 850
\end{aligned}
$$

There are $150 + 2400 + 48120 + 10164 + 850 = 61684$ trainable parameters in total.

1.5 Q5

Let $y = \frac{1}{1+e^{-x}}$
then:

$$
\begin{aligned}
\frac{\partial y}{\partial x} &= \frac{0 + e^{-x}}{(1 + e^{-x})^2} \\
&= \frac{1 + e^{-x} - 1}{(1 + e^{-x})^2} \\
&= \frac{1}{1 + e^{-x}} - \frac{1}{(1 + e^{-x})^2} \\
&= y - y^2 \\
&= y(1 - y)
\end{aligned}
$$

Therefore, we only need the output value to get the derivative of logistic activation function. If we do the backward propagation, there is no need to know the input.

1.6 Q6

(a)
$tanh = \frac{1-e^{-2x}}{1+e^{-2x}} = -1 + \frac{2}{1+e^{-2x}}$
Since $\frac{1}{1+e^{-2x}} \in (0,1)$
Therefore, $tanh \in (-1,1)$
Also, we know that sigmoid(x) $\in$ (0,1). So tanh and sigmoid differs in the function range.

(b)
Since we know $\delta(x) = \frac{1}{1+e^{-x}}$
So $e^{-x} = \frac{1}{\delta} - 1$
We have:

$$\begin{aligned}\frac{\partial y}{\partial x} &= 2 * \frac{2e^{-2x}}{(1+e^{-2x})^2} \\ &= \frac{4(\frac{1}{\delta}-1)^2}{(1+(\frac{1}{\delta}-1)^2)^2} \\ &= \frac{4(\frac{1}{\delta^2}-\frac{2}{\delta}+1)}{(\frac{1}{\delta}^2-\frac{2}{\delta}+2)^2}\end{aligned}$$

Therefore, the gradient can be formulated as a function of logistic function.

(c)
Tanh has range (-1,1), so it can be used to classify between two classes because negative inputs will be mapped strongly negative and zero points mapped near zero. However, sigmoid has range (0,1), so it can be used in classifying more classed compared with Tanh.

2 Part 2

2.1 Task 1

I transfer the images into gray scale and store them as Pytorch tensors. The code is shown as follows:
I split the dataset into training, validation and test set as required.

```
# Define a transform to tensor gray data
transform = transforms.Compose([transforms.Grayscale(),transforms.ToTensor()
PATH = "notMNIST_small"
# Put data into dataloader
dataset = datasets.ImageFolder(root=PATH, transform=transform)
dataloader = torch.utils.data.DataLoader(dataset, batch_size=1, shuffle=True
plt.imshow(dataset[0][0].squeeze())
plt.show()

# split data into training, validation and test set
train_set, val_set, test_set = torch.utils.data.random_split(dataset, [15000

train_set = torch.utils.data.DataLoader(train_set, batch_size=1, shuffle=Tru
val_set = torch.utils.data.DataLoader(val_set, batch_size=1, shuffle=True)
test_set = torch.utils.data.DataLoader(test_set, batch_size=1, shuffle=True)
```

2.2 Task 2

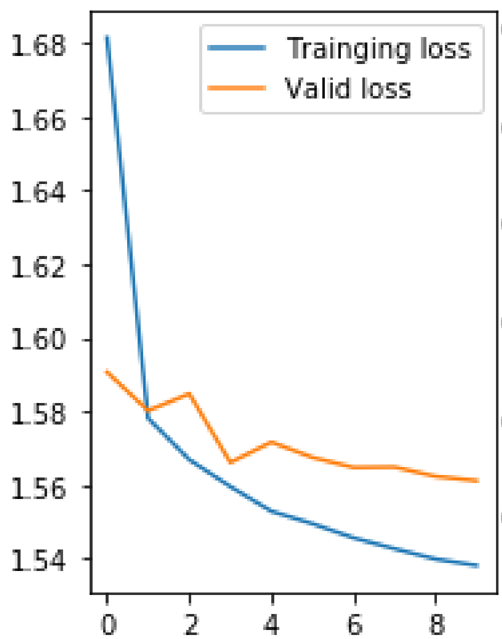
The least validation error occurs when learning rate = 0.01. Therefore, I choose learning rate = 0.01.
The loss and accuracy for each epoch is shown as follows:

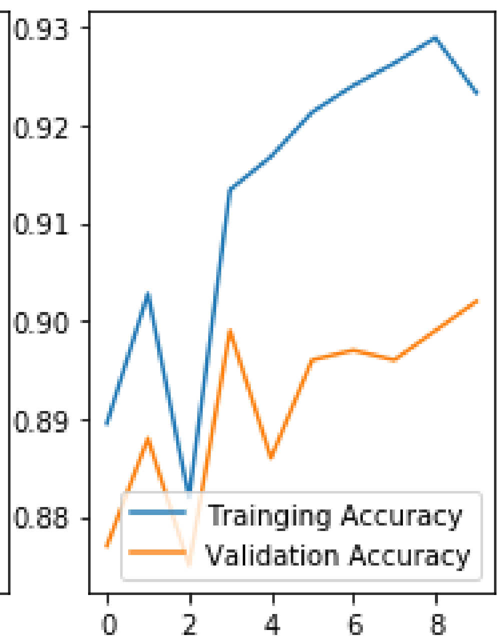
```
Epoch 1, Training loss: 1.669443848133087, Train accuracy: 0.8913333333333333
Epoch 1, Valid Loss: 1.5939233303070068, Valid Accuracy: 0.872
Epoch 1, Test Loss: 1.580641508102417, Test Accuracy: 0.888
Epoch 2, Training loss: 1.5763329716523489, Train accuracy: 0.9084
Epoch 2, Valid Loss: 1.5741633176803589, Valid Accuracy: 0.890
Epoch 2, Test Loss: 1.5673942565917969, Test Accuracy: 0.898
Epoch 3, Training loss: 1.5638115560770034, Train accuracy: 0.9088
Epoch 3, Valid Loss: 1.576158881187439, Valid Accuracy: 0.890
Epoch 3, Test Loss: 1.566970705986023, Test Accuracy: 0.895
Epoch 4, Training loss: 1.5551799175024033, Train accuracy: 0.9178
Epoch 4, Valid Loss: 1.5633509159088135, Valid Accuracy: 0.899
Epoch 4, Test Loss: 1.5602895021438599, Test Accuracy: 0.905
Epoch 5, Training loss: 1.5496447923739751, Train accuracy: 0.9204
Epoch 5, Valid Loss: 1.5657017230987549, Valid Accuracy: 0.898
Epoch 5, Test Loss: 1.559640884399414, Test Accuracy: 0.904
Epoch 6, Training loss: 1.5447507649739582, Train accuracy: 0.9250666666666667
Epoch 6, Valid Loss: 1.5644890069961548, Valid Accuracy: 0.896
Epoch 6, Test Loss: 1.5552014112472534, Test Accuracy: 0.909
Epoch 7, Training loss: 1.5422889921188354, Train accuracy: 0.9311333333333334
Epoch 7, Valid Loss: 1.5641446113586426, Valid Accuracy: 0.896
Epoch 7, Test Loss: 1.551956057548523, Test Accuracy: 0.913
Epoch 8, Training loss: 1.5390977177381515, Train accuracy: 0.9317333333333333
Epoch 8, Valid Loss: 1.5598759651184082, Valid Accuracy: 0.903
Epoch 8, Test Loss: 1.55512273311615, Test Accuracy: 0.909
Epoch 9, Training loss: 1.5359413376569748, Train accuracy: 0.9343333333333333

show more (open the raw output data in a text editor) ...

Epoch 9, Valid Loss: 1.5614333152770996, Valid Accuracy: 0.900
Epoch 9, Test Loss: 1.553298830986023, Test Accuracy: 0.910
Epoch 10, Training loss: 1.533662000187238, Train accuracy: 0.9336666666666666
Epoch 10, Valid Loss: 1.5618765354156494, Valid Accuracy: 0.903
Epoch 10, Test Loss: 1.5540754795074463, Test Accuracy: 0.909
```

The graph for loss and accuracy are shown as follows:



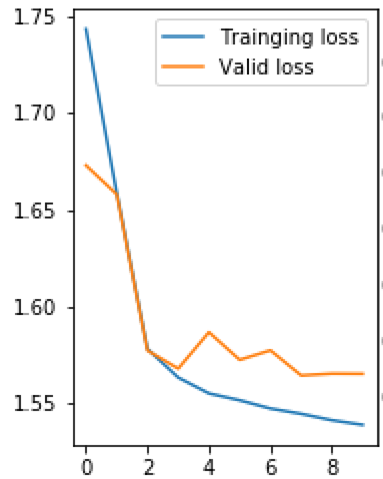


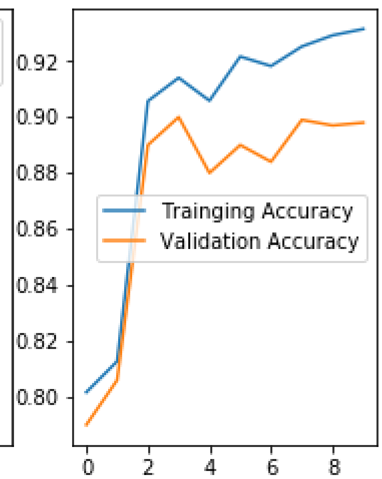
The training loss is 1.5336, training accuracy is 0.934.
The validation loss is 1.5619, validation accuracy is 0.903
The test loss is 1.5541, test accuracy is 0.909.

2.3 Task 3

The training process and accuracy for hidden unit = 100 is shown as follows:

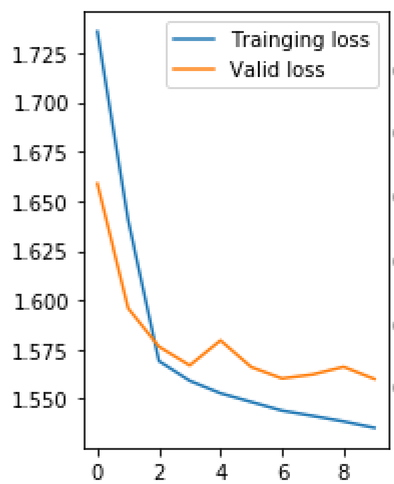
```
Epoch 1, Training loss: 1.7435459966818492, Train accuracy: 0.8016666666666666
Epoch 1, Valid Loss: 1.6728988885879517, Valid Accuracy: 0.790
Epoch 2, Training loss: 1.658864379843076, Train accuracy: 0.8127333333333333
Epoch 2, Valid Loss: 1.6579276323318481, Valid Accuracy: 0.806
Epoch 3, Training loss: 1.578160778093338, Train accuracy: 0.9057333333333333
Epoch 3, Valid Loss: 1.577335238456726, Valid Accuracy: 0.890
Epoch 4, Training loss: 1.5633172588268915, Train accuracy: 0.9140666666666667
Epoch 4, Valid Loss: 1.5679429769515991, Valid Accuracy: 0.900
Epoch 5, Training loss: 1.5550981132825215, Train accuracy: 0.9058666666666667
Epoch 5, Valid Loss: 1.586798906326294, Valid Accuracy: 0.880
Epoch 6, Training loss: 1.5515501772960028, Train accuracy: 0.9216666666666666
Epoch 6, Valid Loss: 1.5724867582321167, Valid Accuracy: 0.890
Epoch 7, Training loss: 1.547275187834104, Train accuracy: 0.9182666666666667
Epoch 7, Valid Loss: 1.577408790588379, Valid Accuracy: 0.884
Epoch 8, Training loss: 1.5445282136122385, Train accuracy: 0.9252666666666667
Epoch 8, Valid Loss: 1.5644092559814453, Valid Accuracy: 0.899
Epoch 9, Training loss: 1.5412188506523767, Train accuracy: 0.9292666666666667
Epoch 9, Valid Loss: 1.565346360206604, Valid Accuracy: 0.897
Epoch 10, Training loss: 1.5388338572740554, Train accuracy: 0.9315333333333333
Epoch 10, Valid Loss: 1.5652652978897095, Valid Accuracy: 0.898
```

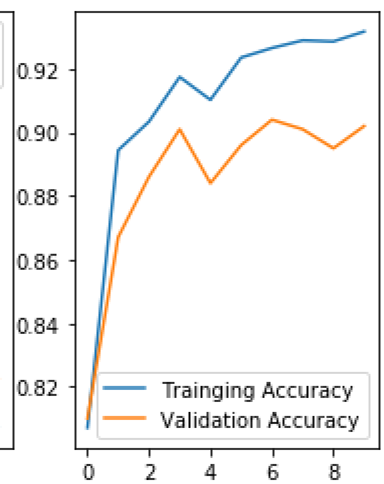




The training process and accuracy for hidden unit = 500 is shown as follows:

```
Epoch 1, Training loss: 1.7359175867557526, Train accuracy: 0.8070666666666667
Epoch 1, Valid Loss: 1.6587797403335571, Valid Accuracy: 0.810
Epoch 2, Training loss: 1.6403543238162994, Train accuracy: 0.8944
Epoch 2, Valid Loss: 1.5958607196807861, Valid Accuracy: 0.867
Epoch 3, Training loss: 1.5691008810122808, Train accuracy: 0.9034
Epoch 3, Valid Loss: 1.5763987302780151, Valid Accuracy: 0.886
Epoch 4, Training loss: 1.5590854446013769, Train accuracy: 0.9174666666666667
Epoch 4, Valid Loss: 1.5668898820877075, Valid Accuracy: 0.901
Epoch 5, Training loss: 1.55275145055453, Train accuracy: 0.9102666666666667
Epoch 5, Valid Loss: 1.5795304775238037, Valid Accuracy: 0.884
Epoch 6, Training loss: 1.5483512242794037, Train accuracy: 0.9236
Epoch 6, Valid Loss: 1.5660333633422852, Valid Accuracy: 0.896
Epoch 7, Training loss: 1.5439512927691141, Train accuracy: 0.9266
Epoch 7, Valid Loss: 1.5602589845657349, Valid Accuracy: 0.904
Epoch 8, Training loss: 1.5412697958310446, Train accuracy: 0.929
Epoch 8, Valid Loss: 1.5622998476028442, Valid Accuracy: 0.901
Epoch 9, Training loss: 1.5384637343962988, Train accuracy: 0.9286666666666666
Epoch 9, Valid Loss: 1.5661519765853882, Valid Accuracy: 0.895
Epoch 10, Training loss: 1.5352415783087412, Train accuracy: 0.9318
Epoch 10, Valid Loss: 1.5599983930587769, Valid Accuracy: 0.902
```





If unit = 100, the validation loss is 1.565, validation accuracy is 0.898.

If unit = 500, the validation loss is 1.560, validation accuracy is 0.902.
If unit = 1000, the validation loss is 1.560, validation accuracy is 0.903.
If unit = 100, the test accuracy is 0.900.
If unit = 500, the test accuracy is 0.901.
If unit = 1000, the test accuracy is 0.909.

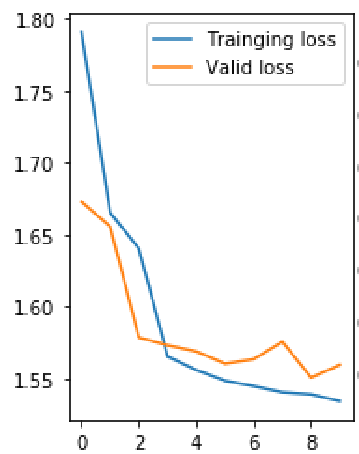
The best validation error occurs when hidden unit = 1000, with test accuracy = 0.909.

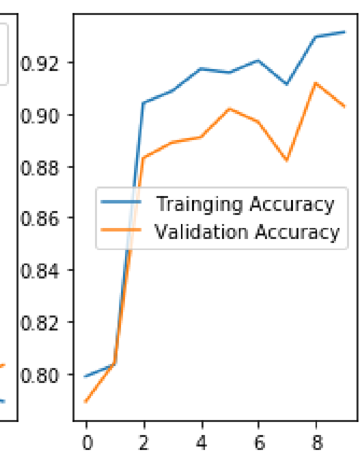
Observation: When the number of hidden units increase from 100 to 1000, the model has a better validation and test accuracy.

2.4 Task 4

The training result is shown as follows:

```
Epoch 1, Training loss: 1.7909219048817953, Train accuracy: 0.7986666666666666
Epoch 1, Valid Loss: 1.6730022430419922, Valid Accuracy: 0.789
Epoch 2, Training loss: 1.6654353649218878, Train accuracy: 0.8032
Epoch 2, Valid Loss: 1.6557978391647339, Valid Accuracy: 0.804
Epoch 3, Training loss: 1.6405191870053608, Train accuracy: 0.9042
Epoch 3, Valid Loss: 1.578701376914978, Valid Accuracy: 0.883
Epoch 4, Training loss: 1.5655361478090286, Train accuracy: 0.9088666666666667
Epoch 4, Valid Loss: 1.5732589960098267, Valid Accuracy: 0.889
Epoch 5, Training loss: 1.5561897614955902, Train accuracy: 0.9174
Epoch 5, Valid Loss: 1.5690351724624634, Valid Accuracy: 0.891
Epoch 6, Training loss: 1.5487092996994654, Train accuracy: 0.916
Epoch 6, Valid Loss: 1.5605688095092773, Valid Accuracy: 0.902
Epoch 7, Training loss: 1.544945517897606, Train accuracy: 0.9206666666666666
Epoch 7, Valid Loss: 1.5636616945266724, Valid Accuracy: 0.897
Epoch 8, Training loss: 1.5407059283892313, Train accuracy: 0.9114666666666666
Epoch 8, Valid Loss: 1.5757527351379395, Valid Accuracy: 0.882
Epoch 9, Training loss: 1.5392727221012115, Train accuracy: 0.9298
Epoch 9, Valid Loss: 1.5508739948272705, Valid Accuracy: 0.912
Epoch 10, Training loss: 1.5345542540391286, Train accuracy: 0.9316
Epoch 10, Valid Loss: 1.5597516298294067, Valid Accuracy: 0.903
```



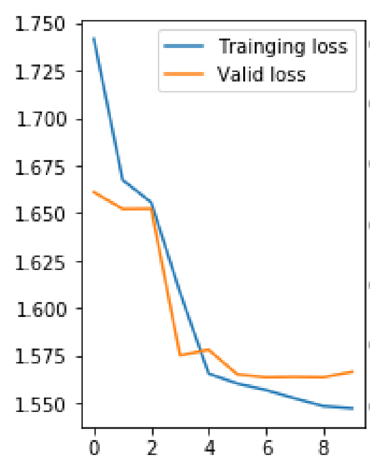


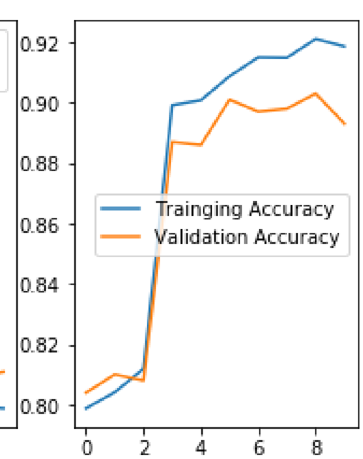
The validation accuracy is 0.903, test accuracy is 0.900.
The model with two layers has a similar validation error as the one-layer model. The test accuracy is a bit lower than the one-layer model.

2.5 Task 5

The training result for model with dropout is shown as follows:

```
Epoch 1, Training loss: 1.7415115685621898, Train accuracy: 0.7988
Epoch 1, Valid Loss: 1.6610021591186523, Valid Accuracy: 0.804
Epoch 2, Training loss: 1.6674253222465516, Train accuracy: 0.8041333333333334
Epoch 2, Valid Loss: 1.6521915197372437, Valid Accuracy: 0.810
Epoch 3, Training loss: 1.6555689032475154, Train accuracy: 0.8118666666666666
Epoch 3, Valid Loss: 1.652296781539917, Valid Accuracy: 0.808
Epoch 4, Training loss: 1.6083592546145122, Train accuracy: 0.8991333333333333
Epoch 4, Valid Loss: 1.5752291679382324, Valid Accuracy: 0.887
Epoch 5, Training loss: 1.5656225205024084, Train accuracy: 0.9008
Epoch 5, Valid Loss: 1.5781995058059692, Valid Accuracy: 0.886
Epoch 6, Training loss: 1.560445082449913, Train accuracy: 0.9087333333333333
Epoch 6, Valid Loss: 1.5652483701705933, Valid Accuracy: 0.901
Epoch 7, Training loss: 1.5570362313508987, Train accuracy: 0.915
Epoch 7, Valid Loss: 1.5637155771255493, Valid Accuracy: 0.897
Epoch 8, Training loss: 1.5526984762827556, Train accuracy: 0.9148666666666667
Epoch 8, Valid Loss: 1.563943862915039, Valid Accuracy: 0.898
Epoch 9, Training loss: 1.548561322983106, Train accuracy: 0.9210666666666667
Epoch 9, Valid Loss: 1.5637186765670776, Valid Accuracy: 0.903
Epoch 10, Training loss: 1.5474256283998489, Train accuracy: 0.9186
Epoch 10, Valid Loss: 1.5665515661239624, Valid Accuracy: 0.893
```





The validation accuracy is 0.903, validation loss is 1.564. The test accuracy is 0.911.
Conclusion: The test accuracy for the dropout model is higher than the original model. Therefore, dropout improves the model performance.