1. (1 point) Compare and Contrast the results of the Linear Classifier, KNN CLassifier, Fully Connected Classifier (your optimal solution), CNN Classifier (your optimal solution), and Transfer Learning (using ResNet-18). Why did you get the results you got (Write one line each about the classifier, as to why they performed well and didn't perform well)

Classifier	Comment
Linear Classifier	Linear classifier predicts the target class based on pre-trained weight and bias by linear equation It is easier to cause overfitting because of its linearity
KNN CLassifier	KNN classifiers predict the class based on its distance from each group, It shows lower accuracy for complex high dimensional data set due to the difficult of calculate all distance to the other groups.
Fully Connected Classifier	Accuracy: 54% Fully connected classifier. Show more accuracy then Linear and classifier. Takes so long time to
CNN Classifier	Accuracy: 58%. CNN Classifier shows a little bit improvement comparing to FC classifier.
Transfer Learning	Accuracy: 80.83% It shows the best performance (81.7% accuracy) thanks to its well pre-trained network. Increase of epoch slightly affects overall result because it's pre trained network.

2. (1 Point) With respect to Part 1, briefly explain how different regularization parameters and different optimization techniques affect the results of your training and validation set. (Mention 2 points each on your findings with regularization and optimization techniques. You can use any relevant part of the assignment for context)

Classifier	regularization and optimization techniques
Linear Classifier	Linear classifiers have a few parameters for tuning the NN. 1)Learning rate can let the system avoid the overfitting problem. 2) the initial value of parameters and biases also affect overall performance.
KNN CLassifier	1) The starting point of centroids are important to reduce the total number of estimation, because If centroids change, then we need to calculate the distance again

Fully Connected Classifier	1) Higher weight scalde for L2 regularization is useful to prevent model overfitting. 2) Dropout rate is not only good for preventing overfitting, but also performance.
CNN Classifier	1) L2 Regularization, drop out ratio, and weight scale are useful to prevent model overfitting.
Transfer Learning	1)Learning rate and 2) weight decay are useful when the training accuracy is higher than validation accuracy. (Yes, overfitting case)

3. (1 Point) With respect to the final model that you submitted in Part 2, how many algebraic operations (additions and multiplications) were performed in the forward pass for the first layer (Conv_ReLU_Pool)? Explicitly mention input image dimensions, number of filters, and size of the filter along with your calculations. For simplicity assume only one image is passed.

[Answer]

- <Abbreviation>
    - Input image= in, Filter=flt, H= Height, W=Width, C=Channel, D=Depth (The number of filters)
- <Calculation>
    - input image dimensions= C x H_in x W_in
    - Filter dimension = D x C x H_flt x w_flt
    - H_out = (H_in + 2 * pad - H_flt) // stride + 1
    - W_out = (W_in + 2 * pad - W_flt) // stride + 1
    - DxCxHxW convolution = (D*C*N*M)*2
    - **Total number of algebraic operations** =

      FN * H_out*W_out* ((D*C*H_flt * W_flt)

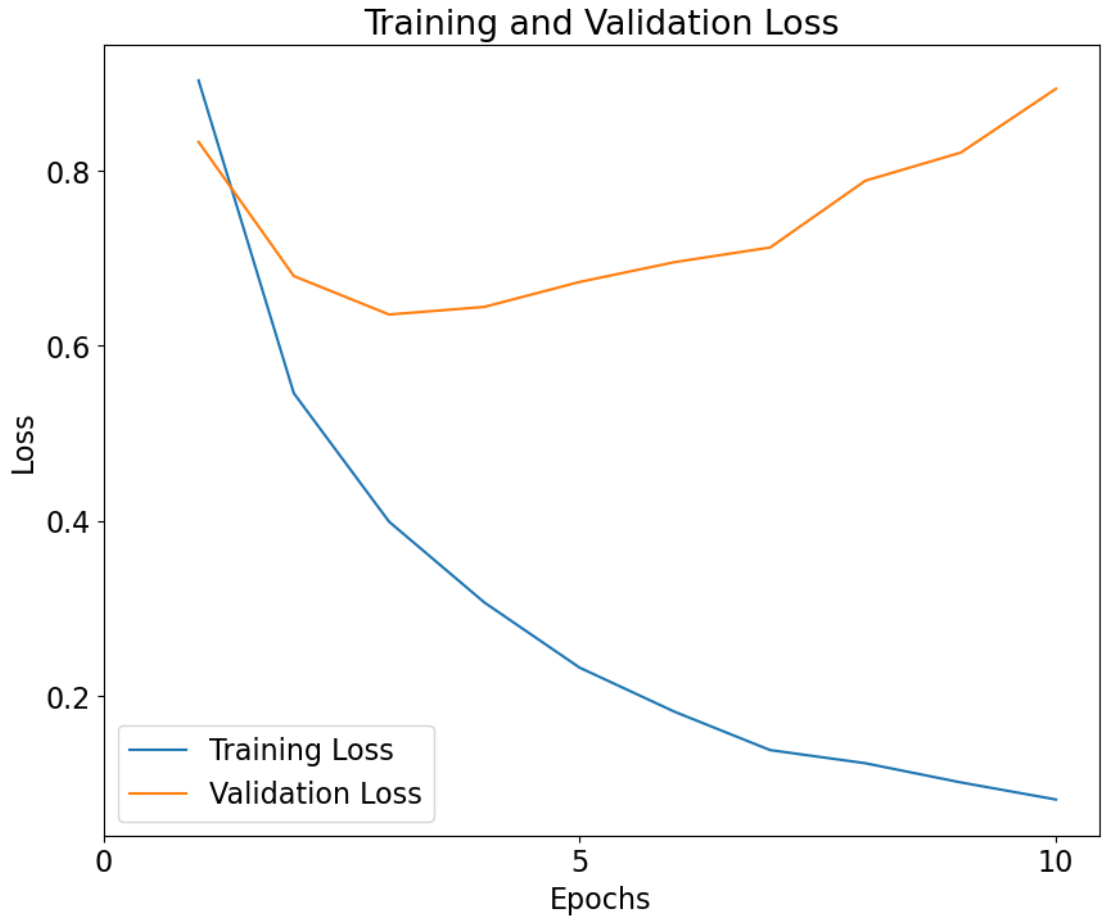
4. (1 Point) For Part 2, justify your choice of hyperparameters based on the observations you made while playing with the model.

[Answer] Usually the higher number of filters and filter size shows the higher accuracy of the system. To avoid the overfitting issue, the hidden dimension value and regularization value need to be set. (0 shows bad performance.) With cuda devices, batch size is very helpful to decrease the training time, but too large batch size with small epochs shows bad results due to insufficient training.

5. (1 Point) In Part 3, what changes do you observe when you continue training for longer? Comment on the training and validation losses and accuracies. What is this observed phenomenon called? What would your approach be to help with this?

Longer training time causes the pollution of the previous well pre-trained network. The graph below shows that the loss for training sharply decreases but the loss for validation increases. which results in

overall accuracy decrease.



6. (1 Point) You have learned about regularization in CNNs. However, not all networks use the same regularization parameters. Comment on the strength of regularization needed to train a model effectively, as the size of the model increases.

[Answer] With a larger model, we typically have more capacity to represent complex patterns in the data. However, if your dataset is relatively small, there's a higher risk of overfitting. In such cases, stronger regularization is necessary to prevent the model from fitting noise in the data.