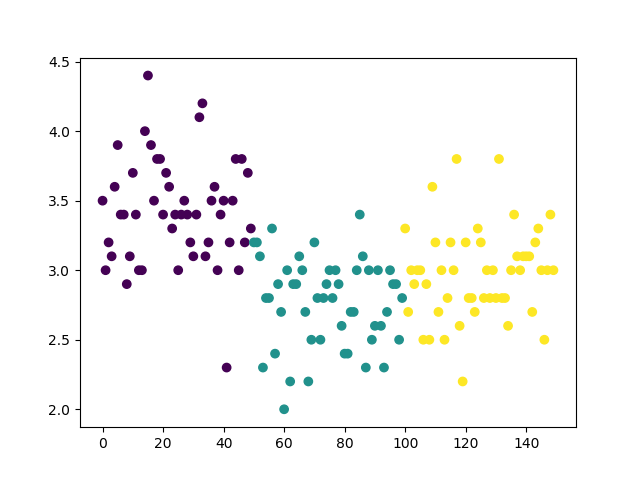
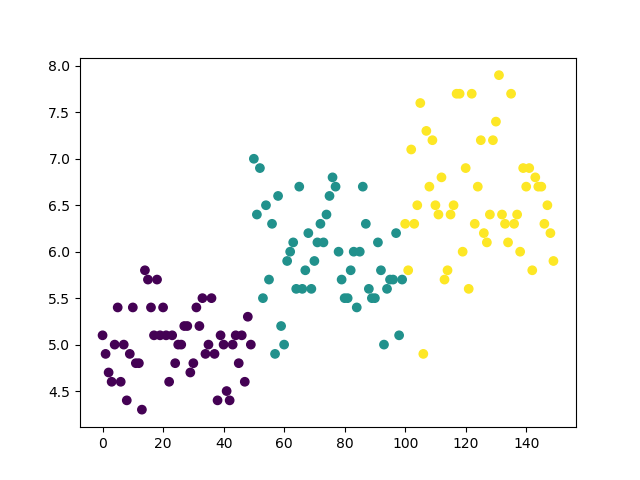
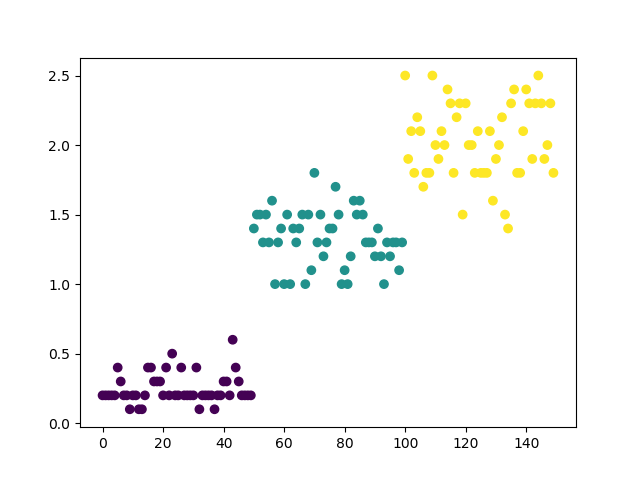
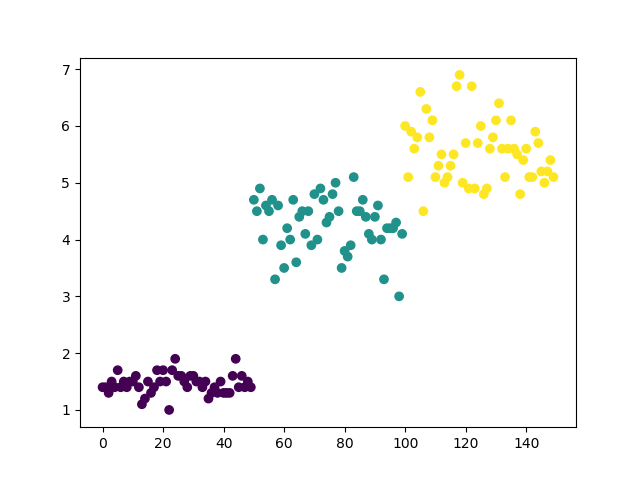
EDA:

The Iris Dataset has 4 features and 3 types of flowers. The four scatter plots below are respectively the 1st, 2nd, 3rd and 4th feature and the corresponding labels.

Clearly, the first 50 data belong to one label, 51~100 data and 101~150 data belong to other labels, respectively.

From the perspective of features, we can see that the label is mainly determined by the 3rd and 4th features.





Network Structure:

The network has 3 layers, each of which is a full connected layer.(We use the one hot label in the training.)

The first layer directly takes data saved in array as input and maps each 4 dimensional data to the hidden layer of dimension 30, followed by a ReLU layer as activation function.

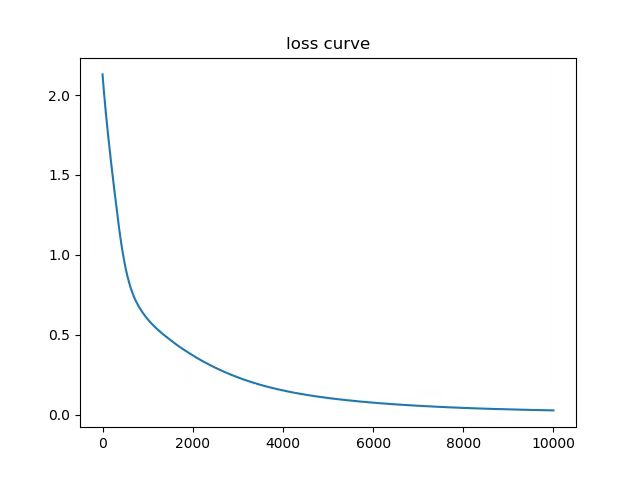
The second layer is a hidden layer, whose input size is 30 and output size is 10. All data mapped from the first layer to this hidden layer will be transformed to a 10-dimensional data. Finally, same with the first layer, a ReLU activation function is added at the end.

The third layer is a hidden layer, whose input size is 10 and output size is 3. Finally, a Softmax activation function is added at the end.

The loss function is cross entropy loss, the optimizer is the Adam optimizer, and the learning rate is 0.005.

Evaluation:

After 10000 iterations, the loss function is almost convergent. The loss curve is shown as follows:



The result accuracy on the validation set is 100 percent.