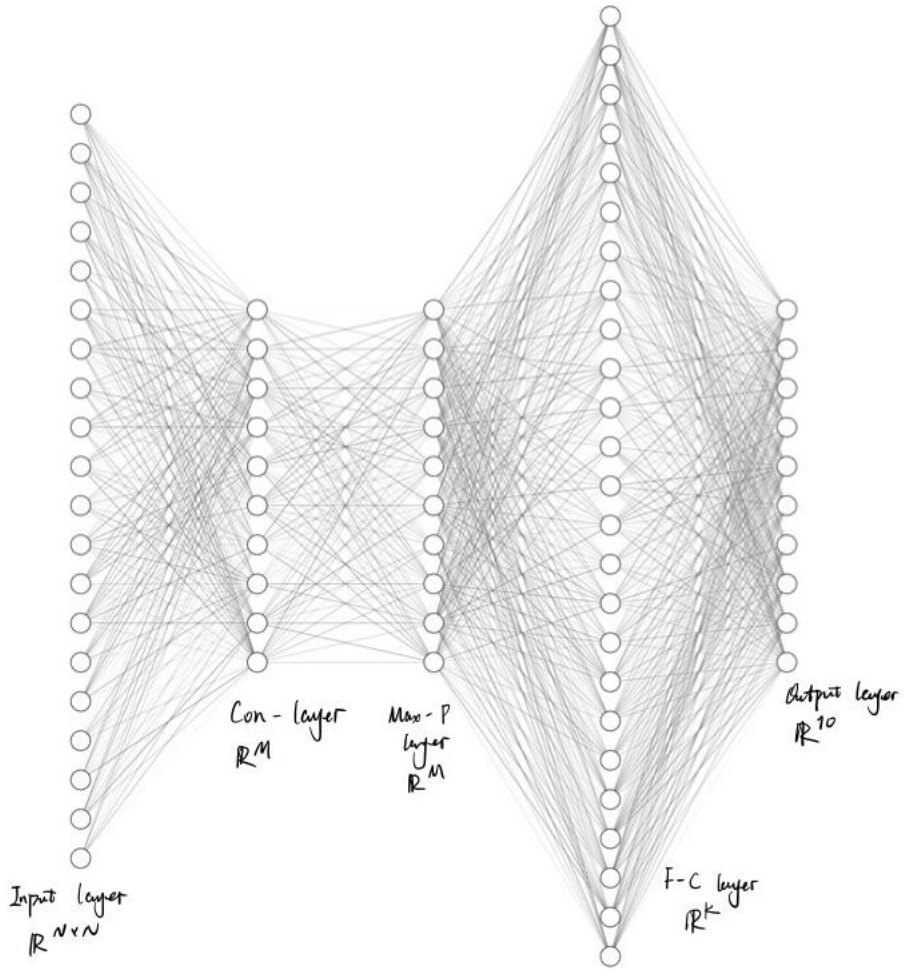
Exercise3

3.1

When w x + b = 0 for one of the perceptron even though multiply a positive c→ ∞, always get the c (w x + b) = 0. The Sigmoid function at the position 0, always have the value of 0.5. This value can not be judged, so it fails.

3.2

(a)



(b)

convolution layer:

M*5*5 = 20*25 = 500

max pooling layer: None

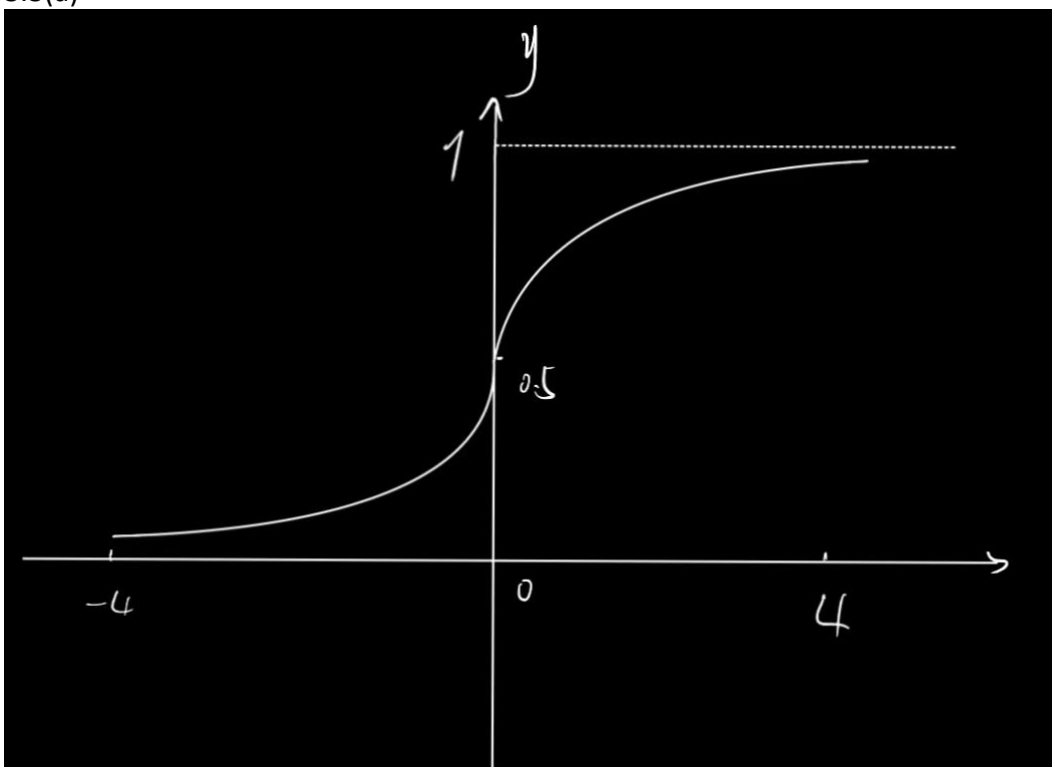
fully connected layer: K*M=100*20=2000

(c) shared weights mean that using the same weights in different layers. (Filtering the images by the same filters)

They used in a CNN to detect different features in an image by using sliding convolution on the same layer.

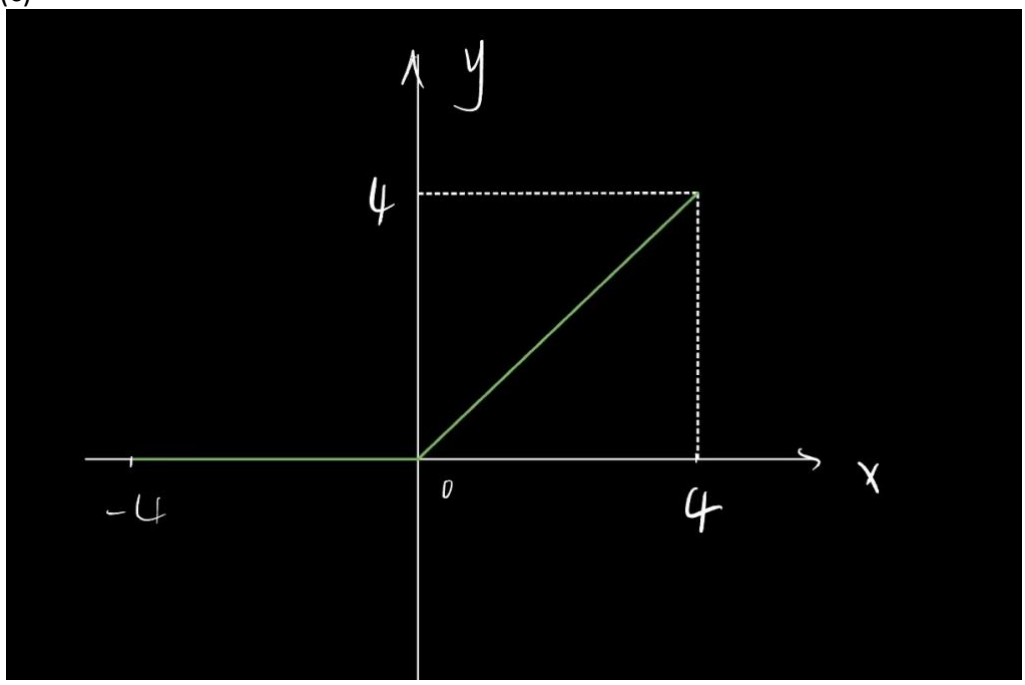
The reason of translation invariance of CNN is the convolution operation is a linear operation and the basic features are lines.

3.3(a)



(b)At the points far from origin, the slope of the function is too small, it will decrease the speed of backward propogation.

(c)



(d)$S(z)i = \frac{e^{zi}}{\sum_{k=1}^{K} e^{zk}}$

Softmax function satisfies the sum of 1, and the outputs are between 0 and 1.

3.4

(a)

The classification error is entirely determined by the prediction mechanism and the actual behavior of the observed quantity. Loss" is a quantitative measure of how bad it is to get an error of a particular size/direction, which is affected by the negative consequences of an inaccurate prediction.

(b)train dataset: using to train the model

Validation dataset:

1.Evaluating model effects

2. adjusting the hyperparameters so that the model works best on the validation set

Test dataset: the final evluate to the model.

(c)data augmentation is extend the training data set by allowing limited data to produce more equivalent data.

Example : Geometric Transformations; Color Space; Noise Injection

(d) Regularization adds a priori to the model parameters, making the model less complex and relatively less perturbed for noise as well as inputs of outlier.

(e)By discarding some unimportant features and keep the important ones