Issued: April 4, 2024 ISE 789/OR 791 HW #6 Due: April 25, 2024

Consider the dataset, letter-recognition.csv. Train on the first 15000 items and then use the resulting model to predict the letter category for the remaining 5000.

Consider 5 possible convolutional neural networks:

(N1) is a network with following information:

The fist layer is a 2-dimensioanl convolutional layer with the dimensionality of the output space being 64, kernel size being 2×2, activation function being "ReLu";

The second layer is 2-dimensioanl convolutional layer with the dimensionality of the output space being 32, kernel size being 2×2, activation function being "ReLu";

The third layer is 2-dimensioanl convolutional layer with the dimensionality of the output space being 16, kernel size being 2×2, activation function being "ReLu";

The fourth layer is a flattern layer;

The fifth layer is a dense layer whose dimensionality of the output space is 26, and the activation function is "softmax".

(N2) is a network with following information:

The fist layer is a 2-dimensioanl convolutional layer with the dimensionality of the output space being 64, kernel size being 2×2, activation function being "ReLu";

The second layer is 2-dimensioanl convolutional layer with the dimensionality of the output space being 32, kernel size being 2×2, activation function being "ReLu";

The third layer is 2-dimensioanl average pooling layer whose output is a 2×2 matrix

;
The forth layer is 2-dimensioanl convolutional layer with the dimensionality of the output space being 16, kernel size being 2×2, activation function being "ReLu";

The fifth layer is a flattern layer;

The sixth layer is a dense layer whose dimensionality of the output space is 26, and the activation function is "softmax".

(N3) is a network with following information:

The fist layer is a 2-dimensioanl convolutional layer with the dimensionality of the output space being 64, kernel size being 2×2, activation function being "ReLu";

The second layer is 2-dimensioanl convolutional layer with the dimensionality of the output space being 32, kernel size being 2×2, activation function being "ReLu";

The third layer is 2-dimensioanl max pooling layer whose output is a 2×2 matrix;

The forth layer is 2-dimensioanl convolutional layer with the dimensionality of the output space being 16, kernel size being 2×2, activation function being "ReLu";

The fifth layer is a flattern layer;

The sixth layer is a dense layer whose dimensionality of the output space is 26, and the activation function is "softmax".

(N4) is a network with following information:

The fist layer is a 2-dimensioanl convolutional layer with the dimensionality of the output space being 64, kernel size being 2×2, activation function being "Sigmoid";

The second layer is 2-dimensioanl convolutional layer with the dimensionality of the output space being 32, kernel size being 2×2, activation function being "Sigmoid";

The third layer is 2-dimensioanl max pooling layer whose output is a 2×2 matrix;

The forth layer is a flattern layer;

The fifth layer is a dense layer whose dimensionality of the output space is 26, and the activation function is "softmax".

(N5) is a network with following information:

The fist layer is a 2-dimensioanl convolutional layer with the dimensionality of the output space being 64, kernel size being 2×2, activation function being "hyperbolic tangent";

The second layer is 2-dimensioanl convolutional layer with the dimensionality of the output space being 32, kernel size being 2×2, activation function being "hyperbolic tangent";

The third layer is 2-dimensioanl max pooling layer whose output is a 2×2 matrix;

The forth layer is a flattern layer;

The fifth layer is a dense layer whose dimensionality of the output space is 26, and the activation function is "softmax".

1. (40 points) Using accuray score as the metric, fit the network with ADAM solver with epochs=3, 5, 10, 20. Record the accuray score over each experiment.

2. (40 points) Using accuray score as the metric, fit the network with SGD solver with epochs=3, 5, 10, 20. Record the accuray score over each experiment.

3. (20 points) Putting problem 1 and problem 2 together, what do you learn from doing this exercise in terms of number of epochs, solvers, layers, and activation function?