Exercise 6

Hou, Jue

014695647

1. Multilayer perceptron - structure and parameters

Assume that there is no recursive calculation and all the intermediate result will be stored.

Define the number of operation for all the derivative or partial derivative operations as 1 operation except the derivative for sigmoid.

Define the number of operation for sigmoid function as 3 operations and the number of operations for its derivative as 2 operations

|  |  |
| --- | --- |
| Two hidden layer of 5 and 8 nodes |  |
| number of parameters | 3\*5+5\*8+8\*2=71 |
| number of operation for outputs | 2\*5\*(3-1)+2\*8\*(5-1)+2\*2\*(8-1)=112 |
| number of operation for gradients | 8\*(2\*2+1)+5\*(2\*8+1)+8\*2\*(1+2+2)+5\*8+3\*5=260 |

|  |  |
| --- | --- |
| One hidden layer of 13 nodes |  |
| number of parameters | 3\*13+13\*2=65 |
| number of operation for outputs | 2\*13\*(3-1)+2\*2\*(13-1) =100 |
| number of operation for gradients | 13\*(2\*2+1) +13\*2\*(1+2+2)+13\*3=234 |

1. Multilayer perceptron - activation functions

Defineas ReLu function, also define the following functions:

,

Assume and

The network for multiplication can be showed as follow:



1. Character classification with MLP

(Pragramming)

1. Character classification with CNN

CNN have several advantages in the character classification. One is that it can process as a 2D image instead of a 1D vector so that more information can be preserved. Also, such an advantage can help to exploit distinguishing features of the characters and improve the classification.

Estimated complete time: 6 hours (Most time is used to count operations in gradients and test activation function layer)