Homework 7: Weight Initialization and Batch Normalization**縱軸橫軸單位與數量級皆與助教提供的一樣**

|  |  |  |  |
| --- | --- | --- | --- |
| Student id: | B06502028 | Name: | 莊立楷 |

1. **Plot the value distribution picture of output and gradient of output in very layer. (In three different weight initialization method, which are normal, Xavier and orthogonal weight initialization)**

**In normal weight initialization**

|  |  |
| --- | --- |
| value distribution picture of output | value distribution picture of output gradient |
|  |  |

**In Xavier weight initialization**

|  |  |
| --- | --- |
| value distribution picture of output | value distribution picture of output gradient |
|  |  |

1. **Plot the variance value picture of output and output gradient in very layer. (In two different weight initialization methods, which are normal and Xavier weight initialization)**

**In normal weight initialization**

|  |  |
| --- | --- |
| variance value picture of output | value variance picture of output gradient |
|  |  |

**In Xavier weight initialization**

|  |  |
| --- | --- |
| variance value picture of output | value variance picture of output gradient |
|  |  |

**Compare two results**

|  |  |
| --- | --- |
| variance value picture of output | value variance picture of output gradient |
| 藍色為xavier，橘色為Normal | 藍色為xavier，橘色為Normal |

1. **Plot the variance value picture of output and output gradient in very layer. (without and with batchNorm in every layer)**

**Without batchNorm**

|  |  |
| --- | --- |
| variance value picture of output | value variance picture of output gradient |
|  |  |

**With BatchNorm**

|  |  |
| --- | --- |
| variance value picture of output | value variance picture of output gradient |
|  |  |

**Compare two results**

|  |  |
| --- | --- |
| variance value picture of output | value variance picture of output gradient |
|  |  |

1. **Plot the variance value picture of output in very layer. (In two different weight initialization method, which are normal and Xavier weight initialization. Every layer has batchNorm layer)**

|  |  |
| --- | --- |
| variance value picture of output(Normal) | variance value picture of output(Xaxier) |
|  |  |

1. **Xavier weight initialization is for activation function tanh. Which weight initialization is for activation function relu?**

Ans: 使用xavier的變形，He initialization。