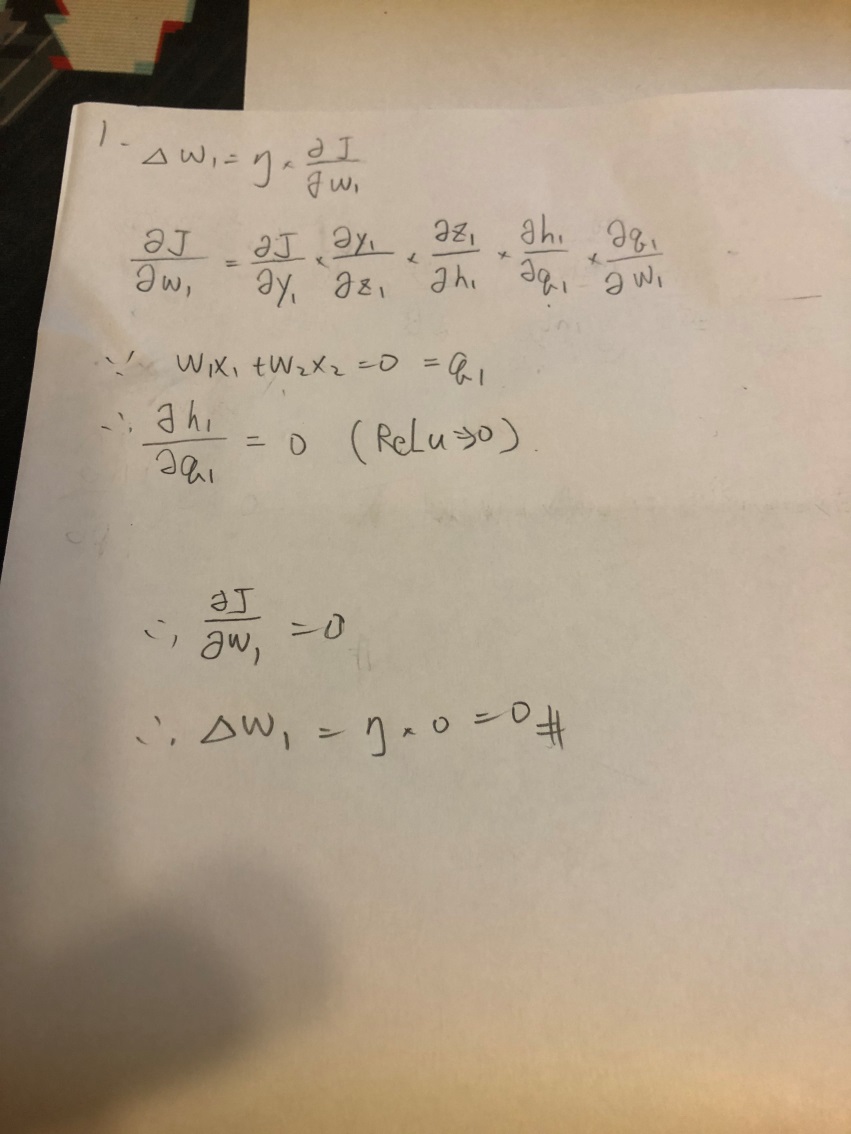
1.



2.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 0 | x1 | 0 | x2 | 0 | x3 | 0 | x4 | 0 | 0 | 0 |  |  |
| 1 | 1 | 2 | 3 | 4 | 5 |  |  |  |  |  |  |  | 3\*x1+5\*x2 | 2 |
| 2 |  | 1 | 2 | 3 | 4 | 5 |  |  |  |  |  |  | 2\*x1+4\*x2 | 2 |
| 3 |  |  | 1 | 2 | 3 | 4 | 5 |  |  |  |  |  | 1\*x1+3\*x2+5\*x3 | 3 |
| 4 |  |  |  | 1 | 2 | 3 | 4 | 5 |  |  |  |  | 2\*x2+4\*x3 | 2 |
| 5 |  |  |  |  | 1 | 2 | 3 | 4 | 5 |  |  |  | 2\*x3+4\*x4 | 2 |
| 6 |  |  |  |  |  | 1 | 2 | 3 | 4 | 5 |  |  | 2\*x3+4\*x4 | 2 |
| 7 |  |  |  |  |  |  | 1 | 2 | 3 | 4 | 5 |  | 1\*x3+3\*x4 | 2 |
| 8 |  |  |  |  |  |  |  | 1 | 2 | 3 | 4 | 5 | 2\*x4 | 1 |

Ans=16

3.

不行，因為權重共享的緣故，如果使用relu，只要有一個權重大於一，在連乘之下，輸出值會變得非常大。

4.

Try a smaller set of features

減少trainning data，讓model少學一點trainningdata的特徵。

Use a smaller value for λ in the L2 norm term

使權重值保持較小

Run gradient descent with a different optimizer (such as RMSprop or ADAM).

可緩解學習率下降過快

5.

