## **Computing derivatives w.r.t all weights in any layer**

1. Let’s take a simple example of a Wk ∊ ℝ3x3 and see what each entry looks like

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | |  | |  | |  |  |
|  | |  | |  | |
|  | |  | |  | |
|  |  |  | |  | |  | |  |  |
|  | |  | |  | |
|  | |  | |  | |

1. Thus we can update all the weights in one go using a
2. Finally coming to the biases,
3. We can now write the gradient w.r.t the vector bk

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | |  | = |
| .  .  . | |
|  | |

1. Thus, we can update all biases using a