**JichenDai quiz 2**

**1.1** Dimension is **4x4**

**1.2** Dimension is **2x2**

**1.3** max pool

| | |
|---|---|
| 55 | 104 |
| 55 | 40 |

**1.4** average pool

| | |
|---|---|
| 29.5 | 102.5 |
| 55 | 25 |

**2.1** After been unrolled through time, it can be seen as a very deep network and we can use back-propagation to train it. Error is propagated back to the first input time step so that the error gradient can be calculated and the weight of the network can be updated.

**2.2** After been unrolled through time, it can be seen as a networks in which all the layers share the same weight.

**2.3** In the forward pass, the activation function can squashes the values, preventing them from blowing up.

However, the backwards pass is linear because activation function is not used in backward pass.

**2.4 Gradient Clipping** : clip the gradient g so that it has a norm if at most $\eta$ .