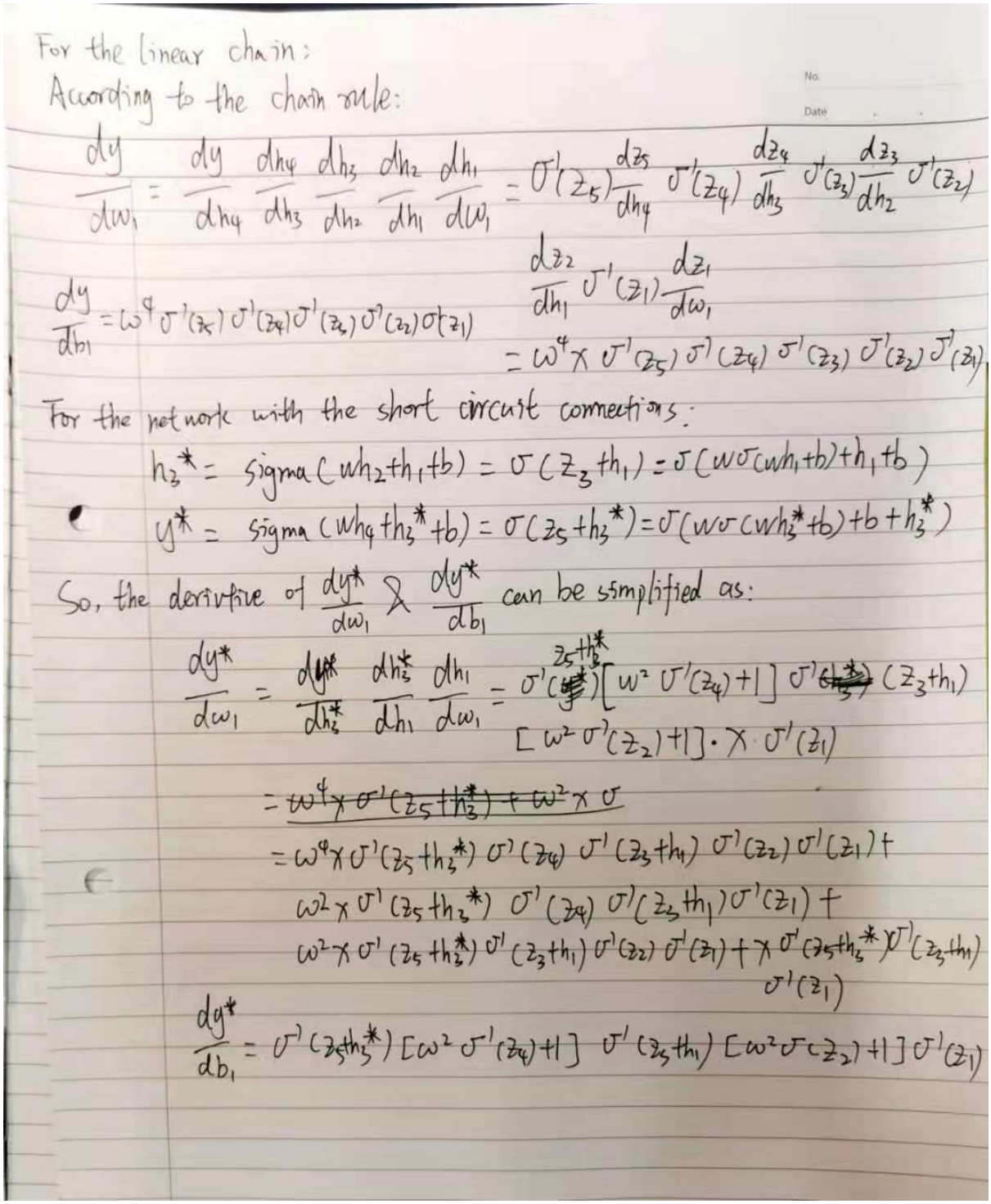
Question 1

Let z1=w1*x+b1, z2=w*h1+b, z3=w*h2+b, z4=w*h3+b, z5=w*h4+b. We can then generate the formula as below:



In dy*/dw1 & dy*/db1, we can find some additional element comparing with dy/dw1 & dy/db1. So, we can conclude that there are |dy*/dw1|>|dy/dw1| and |dy*/db1|>|dy/db1|.

Question 2

1. Let the initial value of θ as θ_0. Then the points θ will go are θ_0+0.3, θ_0+0.6, θ_0+0.9, θ_0+1.2, and then it will come back to θ_0+0.9, and converge.
2. Now we let θ_0=-1. If we keep on implementing our Adam optimizer without stopping, we will find a maximum the θ will reach, which is exactly equalling to the value of h.
   By implementing the Adam optimization by code, we generated a series of value of θ(as below), and we can conclude that the maximum is 0.4101, meaning the maximum of h is also the same as 0.4101.

```
[-1,
 -0.7,
 -0.39999999999999786,
 -0.09999999999999776,
 0.20000000000000367,
 0.35348343141804006,
 0.41018429512997384,
 0.39851277161401233,
 0.3362158088702168,
 0.23511522423117243,
 0.10347681809094891]
```

Question 3