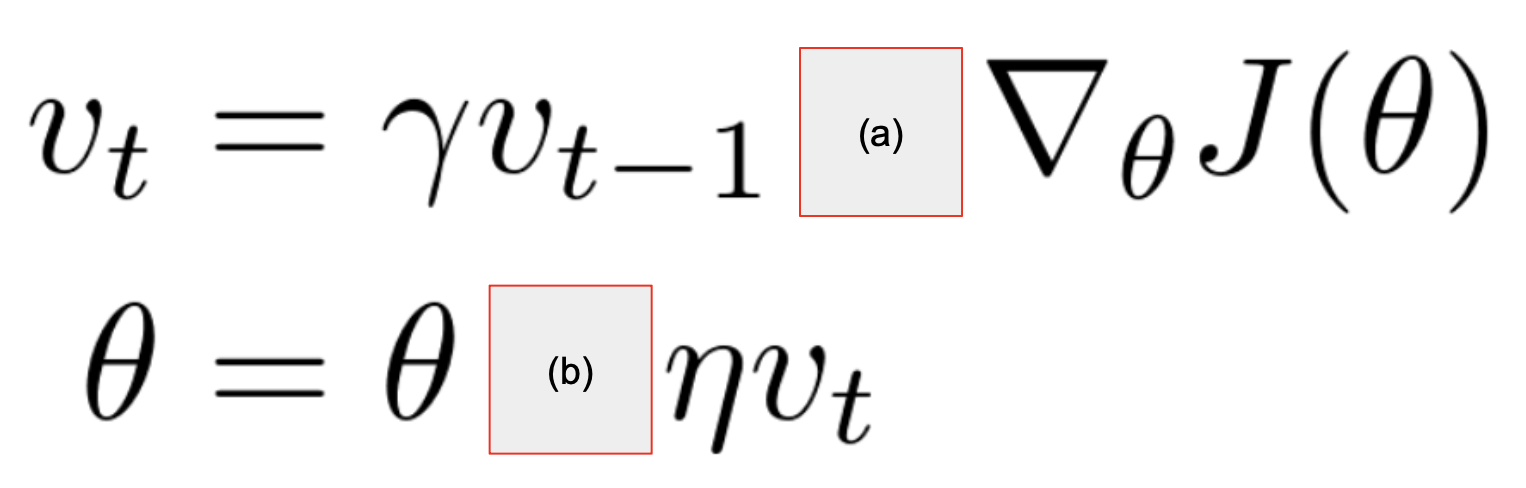
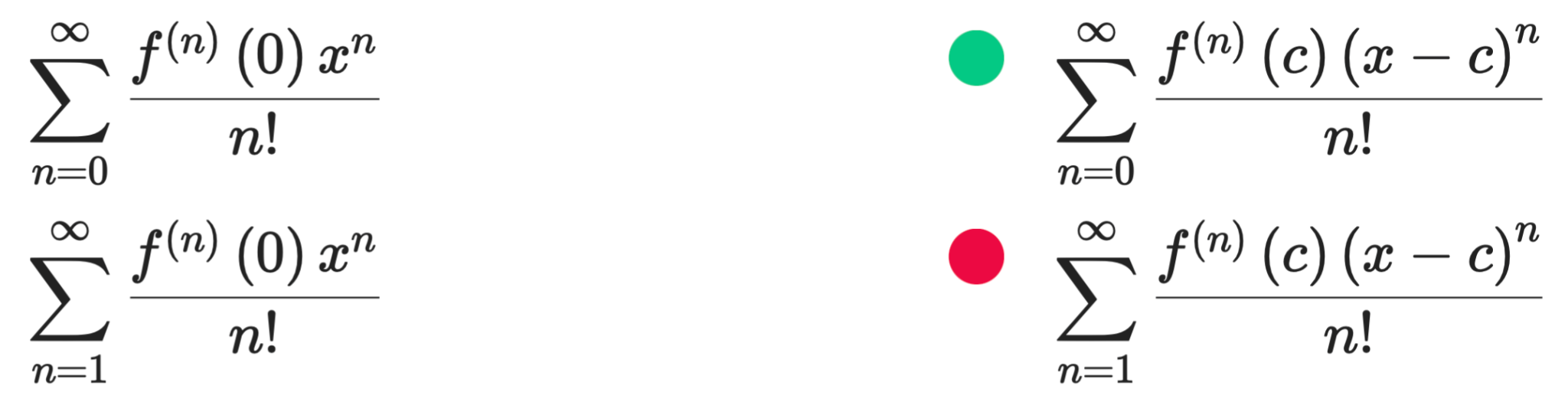
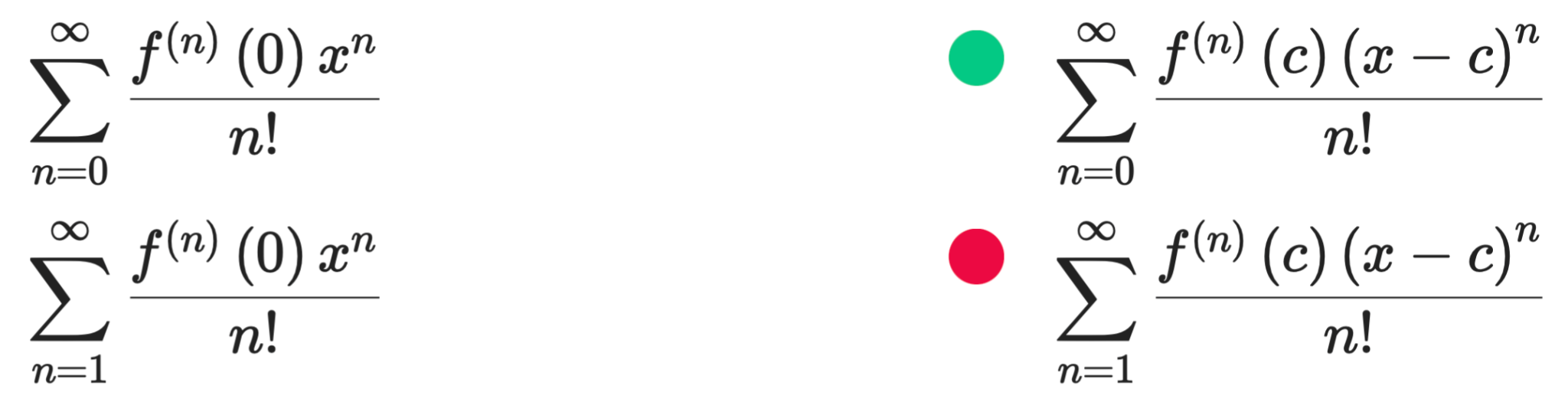
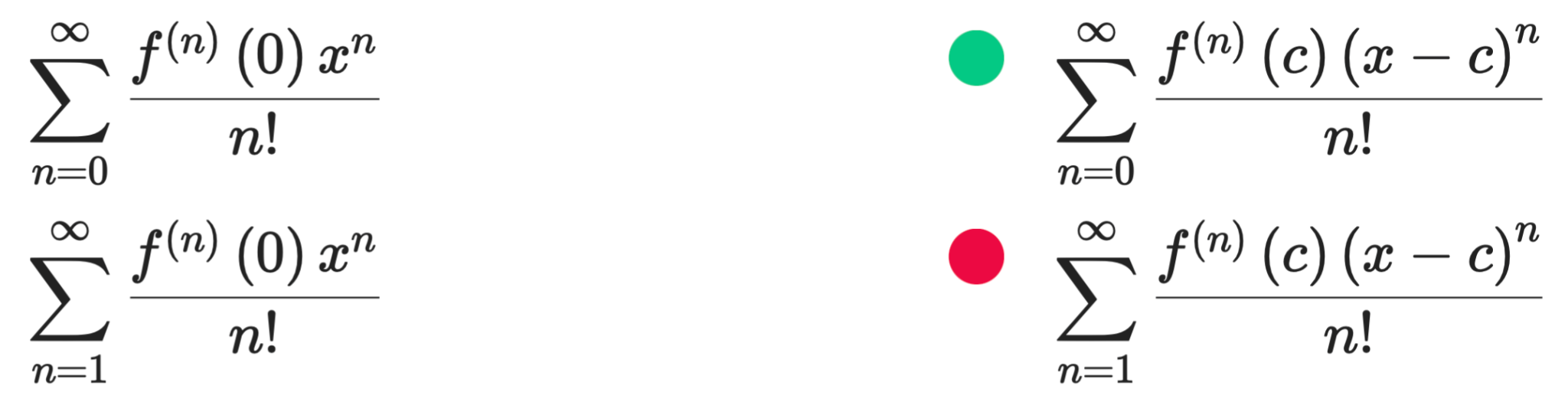
**hw\_08**

1. **What are the commonly used gradient descent optimization functions?(a,b,c,d)**
2. stochastic gradient descent
3. Adagrad
4. momentum
5. RMSProp
6. **Which of the following statements about Adam is/are correct? (a,b,d)**
7. We usually use default values for the bets=(0.9, 0.999) in Adam
8. The learning rate hyperparameter in Adam usually needs to be tuned
9. Adam should be used with batch gradient computations, not with mini-batches
10. Adam combines the advantages of RMSProp and momentum
11. **Which of the following is/are Per-parameter adaptive learning rate method(s)? (b,c,d)**
12. Nesterov Momentum
13. Adagrad
14. RMSprop
15. Adam
16. **How can we calculate Accumulated gradient(momentum) for weights update? Select the appropriate momentum parameter(♈) and inequalities for a and b.(c)**



1. ♈ =0.9 ,a=(-), b=(+)
2. ♈ =0.99 ,a=(+), b=(-)
3. ♈ =0.9 ,a=(+), b=(-)
4. ♈ =0.99 ,a=(-), b=(+)
5. ♈ =0.9 ,a=(-), b=(-)
6. **Which is the formula of the Taylor series? (a)**

a. b. 

c. d. 