

1. Tại sao không dùng 2 hàm loss như trong tính huống a và b mà lại dùng công thức c
 - Đối với công thức a) $L(\theta) = \tilde{y} - y$ vì kết quả có thể sẽ ra một số âm, khi đó ta nói L nhỏ nhất sẽ không đúng vì khi $L(\theta) = -\infty$ là rất nhỏ nhưng sự sai lệch rất lớn => không thỏa mãn tính chất của một loss function
 - Đối với công thức b) $L(\theta) = |\tilde{y} - y|$, hàm này là hàm không âm, nhưng lại không thuận tiện cho việc cực tiểu hóa bởi vì đạo hàm của nó không liên tục (bị đứt quãng tại $\theta = 0$), thường các loss function phải có tính đạo hàm
 - Đối với công thức c) $L(\theta) = (\tilde{y} - y)^2$, hàm này thỏa mãn tính chất của hàm Loss (không âm, có tính đạo hàm) nên thích hợp làm loss function