1) a) i) Low learning rate prevents explosion and high variance. Similarly, continuously changing m prevents vanishing gradient and high variance. Thus, this low variance helps in preventing exploding and vanishing gradients resulting in learning and preventing overshooting. ii) Stable and previously low gradients get higher updates. This handles sparse gradients.

b) Dropout removes high codependency among neurons. Each neuron also gets strengthened further. We need all neurons for good prediction.