Loss function and accuracy, Momentum SGD w/500 epochs - no gradient optimizer