1.3.2

Flatness v.s. Generalization - Part 2 :

1. Describe the settings of the experiments (e.g. which task, what training approaches) (0.5%)

(1) 使用CNN在MNIST上，結構如下：

Conv2D

ReLU

MaxPool2dㄉ2ㄉ2ㄉ

Conv2D

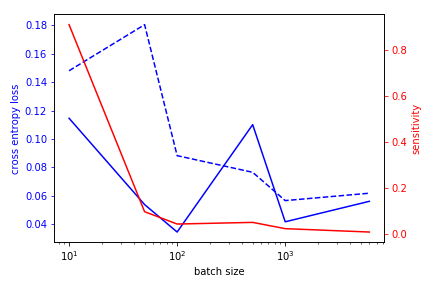
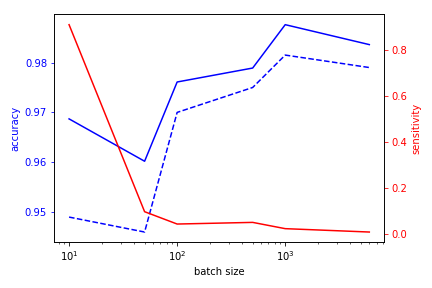
ReLU

MaxPool2d

Linear

(2) 我們在batch size上分別選擇10/ 50/ 100/ 500/ 1000/ 6000，並使用Adam作為optimizer

1. Plot the figures of both training and testing, loss and accuracy, sensitivity to your chosen variable. (1%)

圖中的實線為train，虛線為test

(sensitivity = Frobenius norm of gradient of loss to input x)

1. Comment your result. (1%)

(1) 簡化後的Sensitivity，與batch size之間有負相關：當batch size為10時，sensitivity趨近1，而在batch size調大到50時，Sensitivity急劇下降，介於0到1之間，之後則緩慢下降

(2) 我在MNIST上使用pytorch建CNN，在結果上，train及test之間的差異不大。Sensitivity與loss之間存在某種正相關。