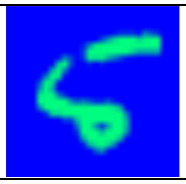
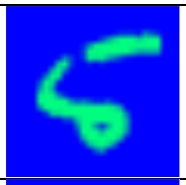
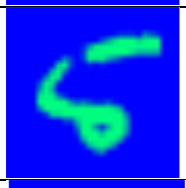
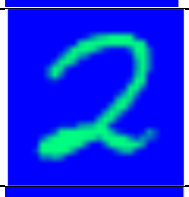
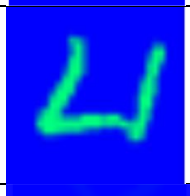
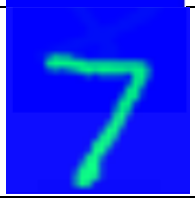
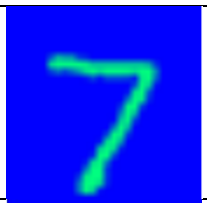
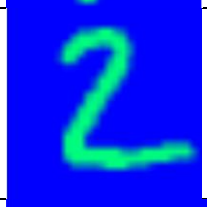
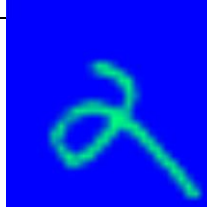
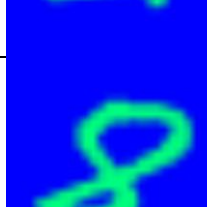
Hidden Layer Sizes	Solver	Random State	Early Stop	Train/Test accuracy	First Wrong	Predicted
(200,)	lbfgs	2	FALSE	1.0/0.9659		0

The image is distorted and I am not clearly sure as to whether the number is 5 or 6 but the computer predicted it as 0.
The Ridge regression with p=5000 has a better accuracy than the model and at p=2000, it is comparatively similar to the Neural network model. As we only use 200 nodes, in general, the MLP has actually done very well.

Hidden Layer Sizes	Solver	Random State	Early Stop	Train/Test accuracy	First Wrong	Predicted
(100,)	Lbfgs	2	FALSE	0.99867/0.9617		0
(100,)	Lbfgs	2	TRUE	0.99867/0.9617		0
(200,)	Adam	2	FALSE	1.0/0.9704		1
(100,)	Lbfgs	7	FALSE	1.0/0.963		0
(100,2)	Lbfgs	2	FALSE	0.21808/0.2155		0

(100,2)	SGD	2	FALSE	0.54755/0.5412		0
(100,2)	Adam	2	FALSE	0.844633/0.7909		0
(200,)	Adam	2	FALSE	1.0/0.9704		1
(200,)	Adam	2	TRUE	0.9893/0.965		9
(200,)	Adam	7	TRUE	0.99203/0.9685		6
(300,)	Adam	2	FALSE	0.99967/0.9697		0
(300,)	Adam	2	TRUE	0.98695/0.9681		2

The wrongly predicted image in Adam (200,) is actually more convincing to be predicted as a 9 due to the closeness of the top to the curve and it does appear like 9 from an angle.

Similarly, with Adam (200,) and Random state 7, the number would be predicted by me to be 6 too, make it in the range of human error.

I would have predicted the 8 as an 8 and not a 2 in the case of Adam (300,) early stop.

Observations:-

- Hidden layers with sizes (200,1) and (100,2) have drastically different accuracies for the same number of nodes.
- Random state change affects performance.
- When we stop early, the accuracy of the model reduces, but is close to the one without early stops.