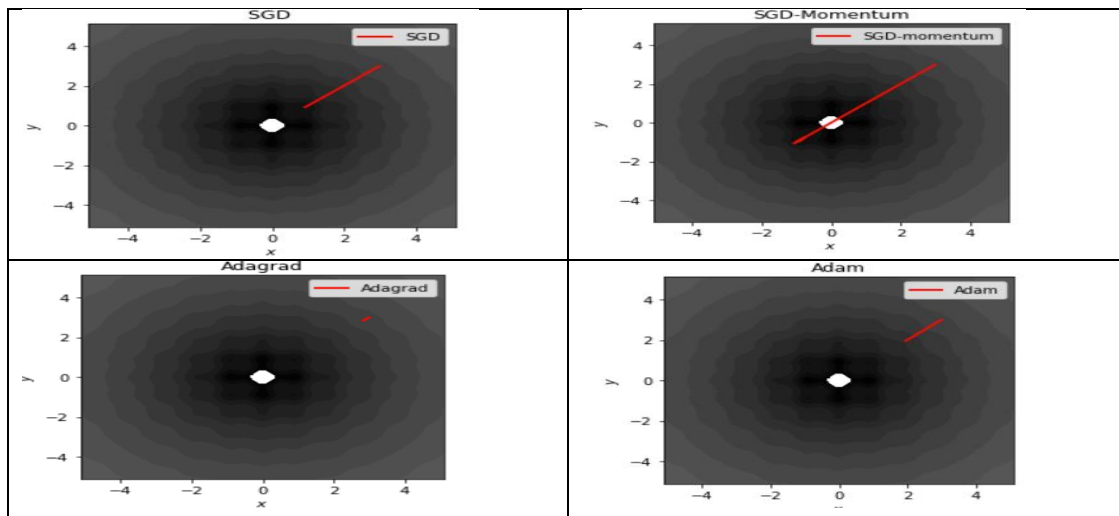
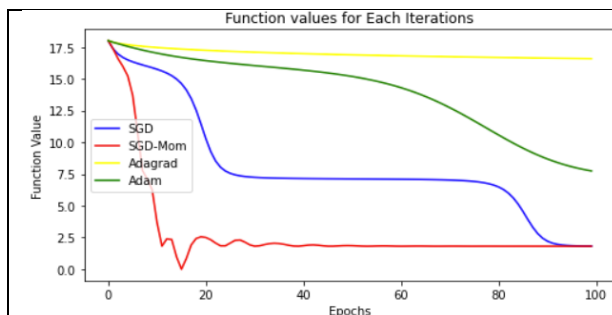


## Question 1

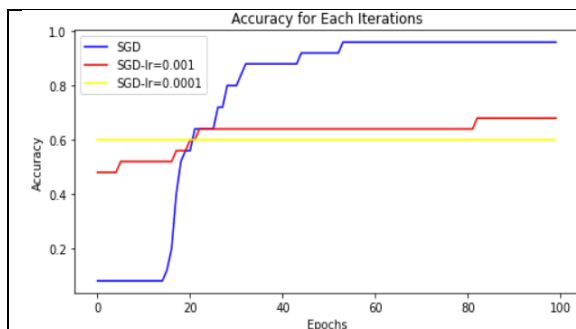


Contour plot of trajectories of different optimizers

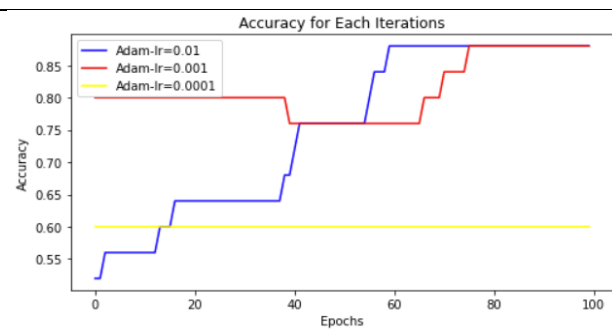


Based on the loss plot, SGD optimizer with momentum produces the best result in terms of the lowest function value. Although the training is unstable in the beginning, the function value converges to a constant value after round 40 epochs when SVM was trained with SGD optimizer with momentum.

## Question 2



	Mean	Variance
<b>SGD 0.01</b>	0.9028	0.001736
<b>SGD 0.001</b>	0.5912	0.060755
<b>SGD 0.0001</b>	0.5592	0.067935



	Mean	Variance
<b>Adam 0.01</b>	0.8920	0.000624
<b>Adam 0.001</b>	0.6488	0.061555
<b>Adam 0.0001</b>	0.4828	0.074312

The SVM was first trained once over 100 epochs for different optimizers with different learning rates and the accuracy was plotted. This was repeated 100 times for each optimizer with different learning rates and the mean and variance of accuracy were calculated. From the above results, SGD and Adam with learning rate of 0.01 produces the best result in terms of accuracy among all 3 variants of SGD and Adam optimizers with different learning rates respectively. The lower the value of the learning rate for each optimizer, the higher the variance and the lower the accuracy. This may be because lower learning rate causes the gradient descent algorithm to take small steps towards global minima and the model may get stuck at local minima. This causes high variance in accuracy because the model is not close or has not reached the global minima.