As asked, we know that log likelihood function defines how well parameter values fit the training examples. If we set the alpha values between 10-7 and 107, then model gives us the varied joint log likelihood. The model seems to fit well for the training set for alpha values from 10-7 to 100 (checked by plotting graph) and from 104. We can say that substantial increase in alpha values from 10-7 to 100, would give us stable or constant values for the joint log likelihood while for alpha values from 101 to 103 or 104, joint log likelihood tends to drop for both the test and train sets and becomes stable for alpha values from 104. As the alpha values increases from 100 to 103, the model seems to show the drop in joint log likelihood values for the training set (all datasets except 1 and 4). Joint log likelihood computed for both training and test sets follow the same trend of growth and drop. However, joint log likelihood computed for test set is less negative than that of train set i.e. absolute sum of train set is greater than absolute sum of test sum.