**Question-1:**

Rahul built a logistic regression model having a training accuracy of 97% while the test accuracy was 48%. What could be the reason for the seeming gulf between test and train accuracy and how can this problem be solved.

Answer:

Reason for this difference between train and test accuracy could be due to the problem of overfitting.

Overfitting happens when the model complexity increases to an extent that the model completely learns the train data rather than learning the underlying trends in the train data.

As a result, it provides very high accuracy on the train dataset, but when the model is used to predict the dependent variables values on test data, its accuracy comes down drastically.

Below analogy can be used to explain this behavior:

Suppose there are 2 students studying for the IIT entrance exams and their study styles are a bit different.

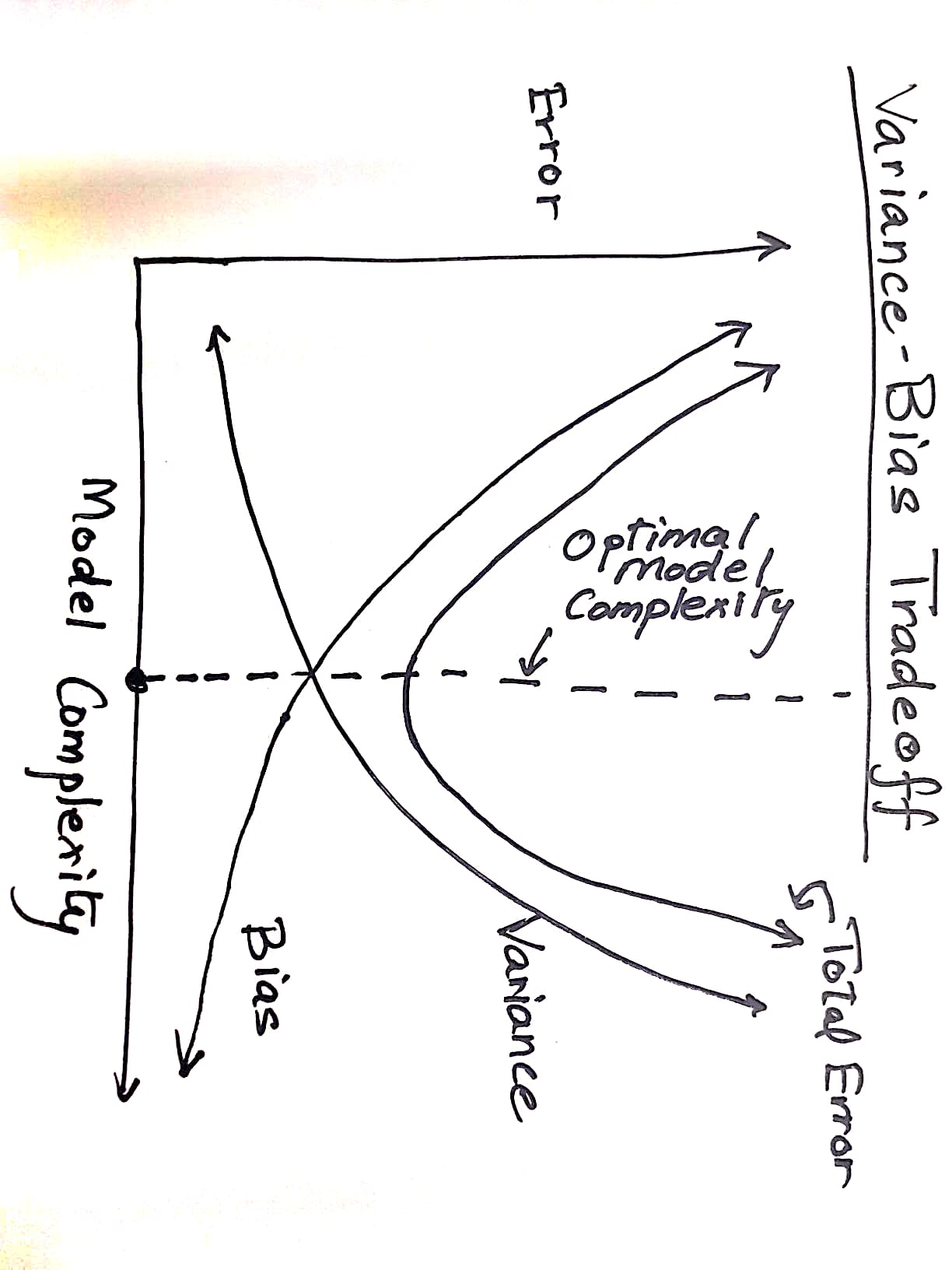
The first student memorizes the solutions to all types of problems from the mathematical guides etc. and he does not give much weightage to the underlying physics, mathematics concepts etc.

The second student, gives more weightage to the concepts of physics and mathematics and does less of practice of problems compared to the first student.

The first student, might not be able to solve some unseen pattern in the main exam, but the second student might be able to crack such unseen problems i.e. he performs well on train data but underperforms on the unseen test scenarios.

But it can also happen that if the second student does too less of problem pattern solving, the second student's understanding is very naïve and basic and as a result he is not able to crack the entrance exams, takes more time to solve the problems in the time bound exam.

This brings us to the concept of **variance-bias tradeoff**.



A more complex model would have high variance but less bias.

Vice versa a very basic naïve model will have low variance and high bias.

Variance measures the change in the model itself with the change in the training data. If the model has high variance, it will be an unstable one, i.e. its prediction on test data will vary heavily with slight changes in the train data.

Bias quantifies how well the model will perform on unseen test data. A model with very low bias, will be too generalized and thus not very successful in predicting the value of dependent variable from the test features with a good amount of accuracy.

So the purpose would be to find a tradeoff between the 2 such that the Total Error(Error added up due to both variance and bias) is kept to a minimum.

Some methods to keep the model complexity under check while not making it too generalized are:

1. Regularization – In this method, a regularization term is introduced along with the error term of regression.

The regression cost function is to minimize the error and it does not take into account the effect on model complexity, the regularization term introduced will try to keep the model complexity under check and try to achieve a balance between the variance and bias.

The 2 types of Regularization are Ridge and Lasso.

GridSearch Cross Validation can be used to find the optimal value of hyperparameter alpha.

2. GridSearch Cross Validation can also be used to find the optimal value of other examples of hyperparameters like "Optimal No of Feature Variables to be used while performing RFE" and to find "the degree of polynomial equation to fit a model for Generalized Linear Regression" to achieve the best score for R2, AIC, BIC or other such Model Evaluation Metrics.

**Question-2:**

List at least 4 differences in detail between L1 and L2 regularization in regression.

Answer:

1. In L2 or Ridge Regression, Penalty term that is added to the Error term to keep the model complexity under check is the Sum of squares of coefficients whereas in L1 or Lasso sum of Absolute values of coefficients is added to the Error Term.

2. L1 or Lasso has an additional advantage that it reduces values of coefficients of some of the insignificant variables to zero, thus also achieving feature elimination.

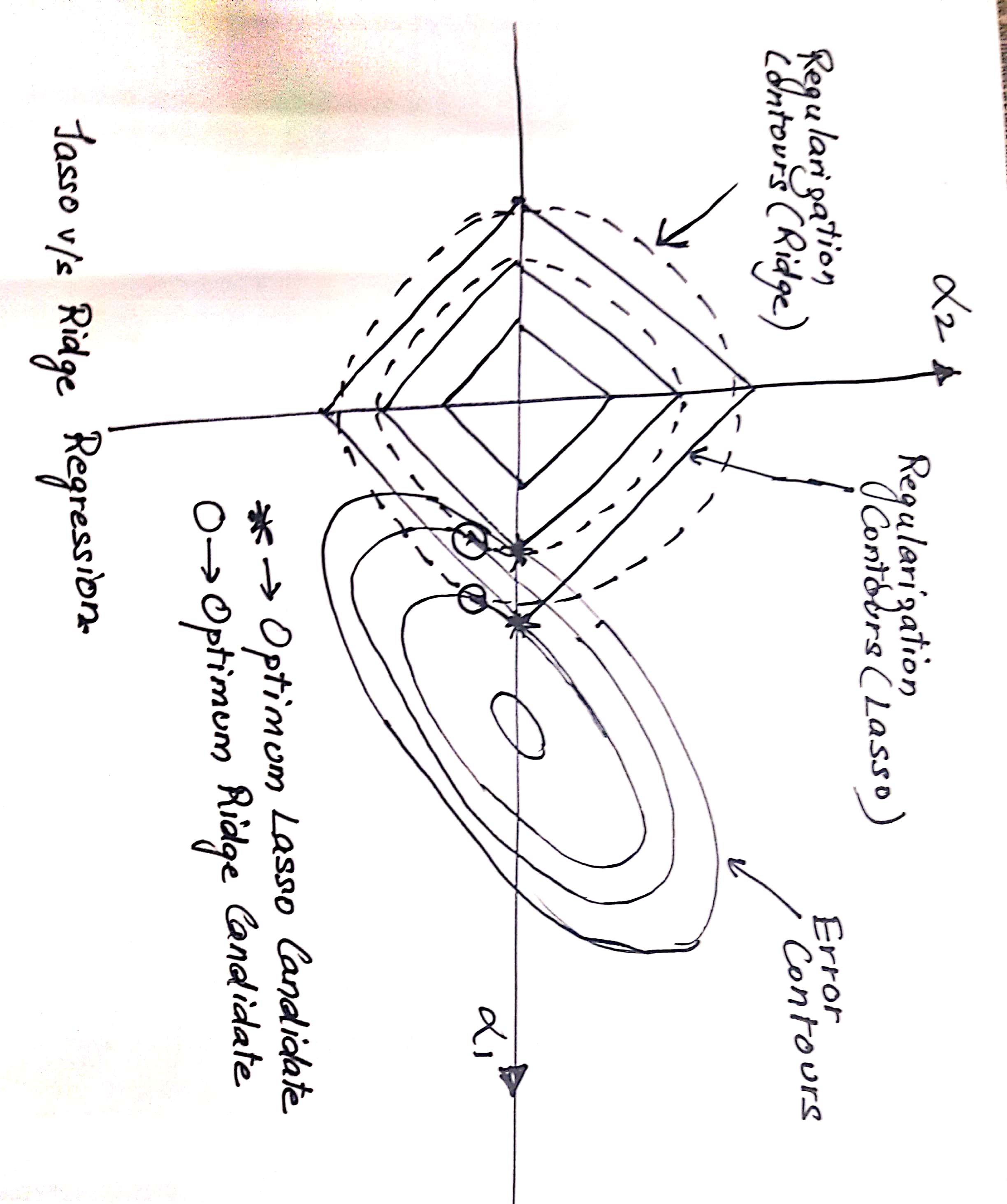
Thus, Lasso or L1 Regression also serves as variable shrinkage method whereas Ridge or L2 Regression does not.

3. The next difference could be in terms of the shape of the regularization contours.

Since Ridge Regularization term is the sum of squares of coefficients, its contours would be like circles around the origin and for Lasso wherein Regularization term is sum of absolute values of coefficients, the contours are diamond shaped around origin.

Due to this the touch points for the Error and Regularization contours for Lasso Regression are at the corners of the regularization contour diamonds which fall on the axis and as a result many Regression Coefficients post Lasso Regression are zero.

Below figure explains this:



4. In terms of usage, if the features are less but highly correlated, Ridge Regression could be used. In case the number of features is huge, we should go for Lasso Regularization due to the advantage of Variable Shrinkage.

5. Lasso is computationally intensive as compared to Ridge Regression.

6. Lasso solution cannot be determined using a simple matrix solution but Ridge can be.

**Question-3:**

Consider two linear models

*L1: y = 39.76x + 32.648628*

And

*L2: y = 43.2x + 19.8*

Given the fact that both the models perform equally well on the test dataset, which one would you prefer and why?

Answer:

If both the models perform equally well on the test dataset, I would pick the L2 model, as the coefficients have less digits after the decimal (less number of bits in the binary encoding of the model) and hence are more simpler compared to the first model.

This would be in accordance with the fundamental tenet of machine Learning known as the **Occam's Razor** which states that a predictive model should be as simple as possible, but no simpler.

**Question-4:**

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

**Answer:** A more Robust Generalizable model would be a stable one and less complex. This would in term reduce the accuracy of the model as it will have more bias.

A more complex model would in turn have more accuracy but would be less robust and less generalized and hence could become unstable with slight variation in train data.

This is in accordance with the variance- bias tradeoff concept. (Please refer answer1 figure)

**Question-5:**

As you have determined the optimal value of lambda for ridge and lasso regression during the assignment, which one would you choose to apply and why?

Both Lasso and Ridge Regression gave R2 score of approx. 0.93 on test data.

Even other metrics valued like mean\_squared\_error and mean\_absolute\_error are almost same for predictions using both Lasso and Ridge Regression.

But if we look at coefficients of Lasso, a lot of them are zero and hence less complex compared to Ridge, we achieve feature elimination as an added advantage due to Lasso.

Hence I would choose Lasso over Ridge Regression.