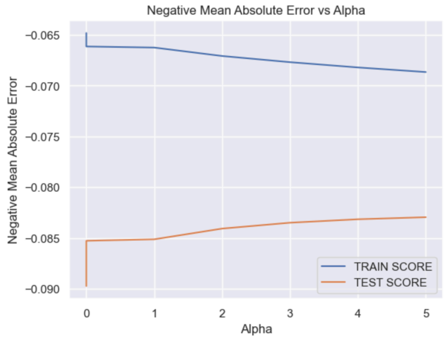
Question 1:

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer 1:

The optimal value of alpha for ridge regression is 2.

In [694]:
```
out['param_alpha'] = out['param_alpha'].astype('int32')
plt.plot(out['param_alpha'], out['mean_train_score'])
plt.plot(out['param_alpha'], out['mean_test_score'])
plt.title('Negative Mean Absolute Error vs Alpha')
plt.xlabel('Alpha')
plt.ylabel('Negative Mean Absolute Error')
plt.legend(['TRAIN SCORE', 'TEST SCORE'], loc='lower right')
plt.show()
```



Applying Ridge Regression

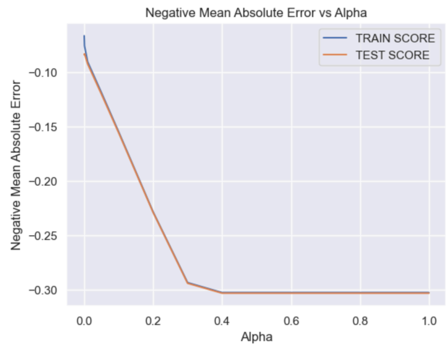
Negative Mean Absolute Error stabilises at Alpha = 2

In [695]:
```
alpha = 2
ridge = Ridge(alpha=alpha)

```

The optimal value of alpha for lasso regression is 0.01

In [681]:
```
out['param_alpha'] = out['param_alpha'].astype('float32')
plt.plot(out['param_alpha'], out['mean_train_score'])
plt.plot(out['param_alpha'], out['mean_test_score'])
plt.title('Negative Mean Absolute Error vs Alpha')
plt.xlabel('Alpha')
plt.ylabel('Negative Mean Absolute Error')
plt.legend(['TRAIN SCORE', 'TEST SCORE'], loc='upper right')
plt.show()
```



Negative Mean Absolute Error is low. Alpha = 0.4 A lower value of alpha has to be used to balance the trade-off b/w Bias-Variance

In [682]:
```
alpha = 0.01
lasso = Lasso(alpha=alpha)
```

When we double the value of alpha for our ridge regression no we will take the value of alpha equal to 10 the model will apply more penalty on the curve and try to make the model more generalized that is making model more simpler and no thinking to fit every data of the data set .from the graph we can see that when alpha is 10 we get more error for both test and train.
Similarly when we increase the value of alpha for lasso we try to penalize more our model and more coefficient of the variable will reduced to zero, when we increase the value of our r2 square also decreases.

If we choose to double the value of alpha for ridge regression, we take the value of alpha equal to 10. As a result, model will apply more penalty on the curve and make the model more generalized. From the graph we see that we get more error for both test and train when alpha is 10.
If we choose to increase the value of alpha for lasso regression, our model is penalised more and more coefficient of the variable are reduced to zero. When the value is increased r2 squared is also decreased.

The most important predictor variables after implementing changes for Ridge Regression:

- GrLivArea
- SaleCondition_Normal
- Neighborhood_Crawfor
- MSZoning_FV
- MSZoning_RL
- MSZoning_RH
- MSZoning_RM
- SaleCondition_Partial
- Neighborhood_StoneBr
- Exterior1st_BrkFace

The most important predictor variables after implementing changes for Lasso Regression:

- GrLivArea
- GarageArea
- OverallQual
- OverallCond
- TotalBsmtSF
- LotArea
- LotArea
- LotFrontage
- BsmtFinSF1
- Fireplaces

Question 2:
You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answer 2:

According to the optimal value of lambda, Lasso regression is the choice. It uses a tuning parameter called lambda as the penalty to get coefficients which is identified by cross validation. When the value for lambda increases Lasso shrinks the coefficient towards zero and it make the variables exactly equal to 0. Lasso also does variable selection. We perform simple linear regression when lambda value is small. When value for lambda increases, shrinkage takes place and variables with 0 value are neglected by the model.

In Ridge regression we use lambda as the tuning parameter. It uses the same as the penalty as it is square of magnitude of coefficients which is identified by cross validation. Residual sum of squares (RSS) should be small when using the penalty. The coefficients that have greater values gets penalised because of the penalty which is equal to lambda times sum of squares of the coefficients. When we increase the lambda value, the model variance is less and bias is constant.

Question 3:
After building the model, you realized that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

Answer 3:

Five most important predictor variables for exclusion:
1. GarageArea
2. GrLivArea
3. TotalBsmtSF
4. OverallQual
5. OverallCond

<u>Question 4:</u>

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

<u>Answer 4:</u>

To make sure that a model is robust and generalisable, the models test score should not be lesser than the train score. Outliers should be dealt with properly. The overall the model should be as simple as possible. In order to make it simple we might need to accept a decrease in accuracy, but it makes the model model and generalisable. The intention is to get an equally performing model whose accuracy does not change on both training and test data.