Since the output variable is not a binary variable, I use the Loss function as:

Where

Our sample size:

Where the matrix has 1s and 0s of binary variables.

Suggest we should use regularization for the above problem to avoid overfitting.

IDEA of Cross Validation:

Where

* corresponds to the th subfold of the Cross validation split.
* corresponds to the number of subfolds
* corresponds to the number of elements in each subfold

I have used random shuffle and according to initial seed, so the shuffle everytime is constant.

To get a more accurate and single number, I have

* Run the Cross Validation times and averaged it for accuracy
* Averaged the Cross Validation Error across all Tenors to get a single number

|  |  |  |  |
| --- | --- | --- | --- |
| Cross Validation Results | CV Error | CV Error | CV Error |
| Standard Least Squares | 1.24 | 5.3 | 2.75 |
| -regularisation | 0.163099319098 | 0.16428284993 | 0.164255387577 |
| -regularisation | 2.22525066464 | 2.22404213856 | 2.22271368142 |
| Bayesian Ridge Regression | 0.162810689975 | 0.164109536249 | 0.164117869174 |

The error should be the lower the better.

The regularization is more stable than the OLS method.

<https://stats.stackexchange.com/questions/118712/why-does-ridge-estimate-become-better-than-ols-by-adding-a-constant-to-the-diago>