

Data Explorer 01



Using the dataset Hitters.csv to build a model to predict Salary.

Quote the following for your best model:

- the model you used (e.g. Ridge, Lasso, etc)
- what features you used
- the degree of the polynomial you used
- the regularization parameter used
- mse on the validation set
- a plot of the test set alongside a plot of your prediction for the test set
- a correlation between the test set and your prediction for the test set

Suggestions

- To get going quickly just drop the Nans
- You can convert the 'League', 'Division', and 'NewLeague' columns to numerical data using `pd.factorize()` (assuming you want to use these columns)
- Don't worry about an independent test set, meaning just use a training set and a validation set to build your model