FAQs – Model Performance Measures, ML Pipeline and Hyperparameter Tuning

1. What is the difference between fit, fit_transform, and transform?

The way we use fit and predict in regression, similarly for functions that transform the data - we have fit and transform

fit - is used to fit parameters of the function

transform - transforming the data using parameters fitted with the fit function

fit_transform - to first fit the parameters of the function and then transform the data also

2. How to tune a model using train, test and validation split?

1. Pick a combination of hyperparameter
2. Train a model using those hyperparameters
3. Find the model's performance on the validation test
4. Repeat this process for all combinations available
5. Choose the model with the best validation score, and find out the final(generalized) score on the test set

3. Tuning the model using Grid search is taking a long time to run. How to proceed?

Tuning a model using grid search usually takes a long time, you can try the following to get more insights

grid_cv = GridSearchCV(estimator=pipe, param_grid=param_grid, scoring=scorer, cv=5, **n_jobs = -1, verbose = 2**)

1. *n_jobs = -1* can speed up the tuning process by utilizing all the CPU cores.
2. *verbose = 2* will give you the number of times the model has to be fit so that you will get an idea of how much time will it take

4. I am getting the same performance with both GridSearchCV and RandomizedSearchCV. How can I change this as this doesn't look practical to me?

Getting the same results is not incorrect, you might get the same results from both grid and random search.

However few things that can be checked in such cases are:

1. If the value of *n_iter* is greater than the possible number of combinations of hyperparameters then you will get the same results from both.
2. Check if you have passed the obtained value of hyperparameters while building the model.