**Week 12: R to Python**

|  |  |  |  |
| --- | --- | --- | --- |
| **R** | **Python** | **Use** | **Difference** |
| randomForest() | RandomForestClassifier() | Create a random forest classifier model with specified parameters and training data | The python function in sklearn does not retain information from subsequent trees leading up to ntrees like R does. Must create a for loop to view error based on number of trees.  Also, the output of the R function gives oob error rate and a confusion matrix where in Python you must extract these on your own  [Python](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html)  [R](https://www.rdocumentation.org/packages/randomForest/versions/4.7-1.1/topics/randomForest) |
| pregnancy\_RF$err.rate | (1-pregnancy\_RF.oob\_score\_) | Extract oob error rate from model | In Python you must first get oob score from model (pregnancy\_RF) then subtract it from 1 to get error. Similar problem as above in terms of not retaining subsequent tree information. As a result, in python there is only 1 error rate per randomforest.  [Python](https://scikit-learn.org/stable/auto_examples/ensemble/plot_ensemble_oob.html)  [R](https://www.rdocumentation.org/packages/randomForest/versions/4.7-1.1/topics/randomForest) |
| pregnancy\_RF$votes | pregnancy\_RF.oob\_decision\_function\_ | View the percentage of trees that voted for each data point to be in each class. | Similar, extract from randomforest model (pregnancy\_RF)  [Python](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html)  [R](https://www.rdocumentation.org/packages/randomForest/versions/4.7-1.1/topics/randomForest) |
| pregnancy\_RF$predicted | np.argmax(pregnancy\_RF.oob\_decision\_function\_,axis=1) | Receive a vector of predictions for each data point. | No direct function in python, easily extracted by using np.argmax on oob\_decision\_function, as seen.  [Python](https://numpy.org/doc/stable/reference/generated/numpy.argmax.html)  [R](https://www.rdocumentation.org/packages/randomForest/versions/4.7-1.1/topics/randomForest) |
| pregnancy\_RF$importance | permutation\_importance() | Receive the mean decrease in accuracy of features in a model | The R function will also give it by individual class and the mean decrease in Gini which there is a separate function in Python for receiving.  [Python](https://scikit-learn.org/stable/modules/permutation_importance.html)  [R](https://www.rdocumentation.org/packages/randomForest/versions/4.7-1.1/topics/randomForest) |
| NA | explainer = lime\_tabular.LimeTabularExplainer | Create a lime explainer object, label it explainer for reference in the next box | Pass training data, feature names, and class names.  Training data must be in an array format.  [Python](https://lime-ml.readthedocs.io/en/latest/lime.html) |
| NA | exp = explainer.explain\_instance() | Retrieve visual containing prediction and feature importance for singular selected row of test data. | Pass a singular row of test data and prediction probability of row from model  [Python](https://lime-ml.readthedocs.io/en/latest/lime.html) |
| NA | exp.show\_in\_notebook() | Display above visual in your notebook for analysis | [Python](https://lime-ml.readthedocs.io/en/latest/lime.html) |
| treesize(  pregnancy\_RF,  terminal = FALSE) | [tree.tree\_.node\_count for tree in pregnancy\_RF\_tune.estimators\_] | Returns how many nodes each tree has in a random forest | In python you must first get the estimators from the randomforest model (pregnancy\_RF) and then loop through them and get their node counts using .tree\_.node\_count  [Python](https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html)  [R](https://www.rdocumentation.org/packages/randomForest/versions/4.7-1.1/topics/treesize) |
|  |  |  |  |