

0.6 - 0.5 - 0.4 -
0.3 0.2 0.1 0.0 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0 CRITICS_POINTS
<pre>modify_a_col(combine.USER_POINTS,0.26) train_df["USER_POINTS"] = boxcox(train_df["USER_POINTS"], 0.26) test_df["USER_POINTS"] = boxcox(test_df["USER_POINTS"], 0.26) 0.024270050203997705</pre>
0.30 0.25 0.20 0.15 0.10 0.05 0.00 -4 -3 -2 -1 0 1
<pre>user_points train_df.RATING.unique() array(['E', 'E10+', 'M', 'T'], dtype=object) rating_dict = { "E" : 1, "E10+" : 5, "M" : 10, "T" : 15</pre>
<pre># train_df["RATING"] = train_df["RATING"].map(rating_dict) # test_df["RATING"] = test_df["RATING"].map(rating_dict) # combine["RATING"] = combine["RATING"].map(rating_dict) # train_df["Skills"] = inv_boxcox(train_df.USER_POINTS, 0.26) * train_df.RATING # test_df["Skills"] = inv_boxcox(test_df.USER_POINTS, 0.26) * test_df.RATING</pre>
<pre># test_df["Skills"] = test_df["Skills"].fillna(np.mean(train_df.Skills)) # combine["Skills"] = inv_boxcox(combine.USER_POINTS, 0.26) * combine.RATING # combine["Skills"] = combine["Skills"].fillna(np.mean(train_df.Skills)) # train_df["Skills"] = boxcox(train_df.Skills, 0.20) # test_df["Skills"] = boxcox(test_df.Skills, 0.20) # combine["Skills"] = boxcox(combine.Skills, 0.20)</pre>
<pre>train_df.skew() ID</pre>
<pre>print(test_df.CONSOLE.unique()) ['ps2' 'psp' 'x' '3ds' 'ds' 'pc' 'x360' 'xone' 'wii' 'gc' 'psv' 'gba' 'ps3' 'wiiu' 'ps4' 'ps' 'dc'] len(train_df.CONSOLE.unique()) 17 print(train_df.CONSOLE.unique())</pre>
<pre>['ds' 'wii' 'pc' 'ps2' 'ps3' 'psp' 'x' 'x360' 'psv' 'gba' 'gc' 'ps' 'wiiu' 'xone' 'ps4' '3ds' 'dc'] console_dict = { "x" : "x-type", "xone" : "x-type", "x360" : "x-type", "psp" : "p-type", "ps2" : "p-type",</pre>
<pre>"ps3" : "p-type", "ps4" : "p-type", "ps" : "p-type", "psv" : "p-type", "pc" : "pc", "wii" : "w-type", "wiiu" : "w-type", "ds" : "d-type", "3ds" : "d-type", "dc" : "d-type",</pre>
<pre>"gba" : "g-type", "gc" : "g-type", } print(len(console_dict.keys())) print(set(console_dict.keys()) == set(train_df.CONSOLE.unique())) # train_df["CONSOLE"] = train_df.CONSOLE.map(console_dict) # test_df["CONSOLE"] = test_df.CONSOLE.map(console_dict) # combine["CONSOLE"] = combine.CONSOLE.map(console_dict)</pre>
<pre>17 True # train_df["How_much_Old"] = train_df["YEAR"].apply(lambda x : 2020 - x) # test_df["How_much_Old"] = test_df["YEAR"].apply(lambda x : 2020 - x) # combine["How_much_Old"] = combine["YEAR"].apply(lambda x : 2020 - x) # train_df = train_df.drop("YEAR",axis = 1) # test_df = test_df.drop("YEAR",axis = 1) # combine = combine.drop("YEAR",axis = 1)</pre>
<pre># combine = combine.drop("YEAR", axis = 1) def move_target_to_last(dataset, col_name): x = dataset[col_name] dataset = dataset.drop(col_name, axis = 1) dataset[col_name] = x return dataset</pre>
<pre>train_df = move_target_to_last(train_df, "SalesInMillions") train_df.head() ID CONSOLE YEAR CATEGORY</pre>
1 731 wii 2012 simulation Konami Digital Entertainment E10+ 1.796218 0.527919 -1.257900 2 495 pc 2019 shooter Activision M 1.219271 -2.811454 -0.576469 3 2641 ps2 2002 sports Electronic Arts E 1.163286 -0.970196 0.339633 4 811 ps3 2013 action Activision M 0.723042 -2.267121 -1.814317 test_df.head() ID CONSOLE YEAR CATEGORY PUBLISHER RATING CRITICS_POINTS USER_POINTS
0 4310 ps2 2008 action Tecmo Koei T 1.122234 -0.664144 1 4011 psp 2007 strategy Atari E10+ 1.338705 -1.907006 2 2185 ps2 2004 shooter Electronic Arts T 0.944129 -1.979223 3 1644 x 2006 action Electronic Arts E 0.914674 -0.125732 4 188 3ds 2011 racing Ubisoft E10+ 0.146945 -1.371757
ID YEAR CRITICS_POINTS USER_POINTS SalesInMillions ID 1.000000 -0.052969 0.007047 0.061334 -0.032628 YEAR -0.052969 1.000000 -0.181344 -0.415385 0.004861 CRITICS_POINTS 0.007047 -0.181344 1.000000 0.521520 -0.119885 USER_POINTS 0.061334 -0.415385 0.521520 1.000000 -0.205847 SalesInMillions -0.032628 0.004861 -0.119885 -0.205847 1.000000
<pre>new_train = train_df.copy() new_test = test_df.copy() new_train = new_train.drop("CRITICS_POINTS", axis = 1) new_test = new_test.drop("CRITICS_POINTS", axis = 1) train_df.head()</pre>
ID CONSOLE YEAR CATEGORY PUBLISHER RATING CRITICS_POINTS USER_POINTS SalesInMillions 0 2860 ds 2008 role-playing Nintendo E 0.894459 -1.024760 0.623434 1 731 wii 2012 simulation Konami Digital Entertainment E10+ 1.796218 0.527919 -1.257900 2 495 pc 2019 shooter Activision M 1.219271 -2.811454 -0.576469 3 2641 ps2 2002 sports Electronic Arts E 1.163286 -0.970196 0.339633 4 811 ps3 2013 action Activision M 0.723042 -2.267121 -1.814317
<pre>class Jiten(object): def get_final_error(self, error, weight): return np.sqrt(boxcox(error,0.27) / (weight + 1e-38)) def is_max_optimal(self): return False def evaluate(self, approxes, target, weight): assert len(approxes) == 1 assert len(target) == len(approxes[0])</pre>
<pre>approx = approxes[0] error_sum = 0.0 weight_sum = 0.0 for i in range(len(approx)): w = 1.0 if weight is None else weight[i] weight_sum += w error_sum += inv_boxcox(w,0.27) * ((inv_boxcox(approx[i],0.27) - inv_boxcox(target[i],0.27))**2)</pre>
<pre>return error_sum, weight_sum class RmseObjective(object): def calc_ders_range(self, approxes, targets, weights): assert len(approxes) == len(targets) if weights is not None: assert len(weights) == len(approxes) targets = inv_boxcox(targets, 0.27)</pre>
<pre>approxes = inv_boxcox(approxes, 0.27) result = [] for index in range(len(targets)): der1 = targets[index] - approxes[index] der2 = -1 if weights is not None: der1 *= weights[index]</pre>
<pre>der2 *= weights[index]</pre>
<pre>cat_columns ['CONSOLE', 'CATEGORY', 'PUBLISHER', 'RATING'] model = CatBoostRegressor(use_best_model=True, learning_rate=0.01, n_estimators=10000, objective=Rms jective(), cat_features=cat_columns, eval_metric=Jiten())</pre>















































