Estimation of parameter importance and interaction with Random Forests

Leslie Pérez Cáceres

In this document we are testing how to use Random Forest to assess importance and interactions of parameters based on the data gathered by irace.

For the analysis below we use as example ACOTSP:

- 20 secs cut off time
- 11 parameters
- 200 instances of size 2000
- 5000 experiments for configuration

1. Importance by iterations

Iteration 1:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
9	localsearch	1.923333	2358	1035.36554	351078.33	300	53	1.0000000
7	elitistants	2.196289	1629	302.27262	117013.84	298	33	1.0000000
8	instance	2.260000	14335	331.74027	197803.00	300	0	0.0000000
2	alpha	2.306667	7374	454.94489	152878.81	300	15	0.0000000
1	algorithm	2.426667	3242	440.17483	141545.59	300	21	1.0000000
10	nnls	2.543333	5219	834.75266	375722.79	300	87	0.1006209
13	rho	2.786667	7104	260.83199	94260.02	300	2	0.0000000
4	beta	2.843333	7498	153.02516	64521.13	300	1	0.0000000
3	ants	3.160000	6832	122.81817	50581.12	300	2	0.0000000
6	dummy	3.403333	5457	85.17861	31098.23	300	1	0.0000013
5	dlb	3.434811	1689	631.98070	268926.78	299	70	1.0000000
11	q0	4.030000	1825	116.15152	41169.44	300	14	1.0000000
12	rasrank	5.050000	2134	32.22108	13438.87	300	1	1.0000000

Iteration 2:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
9	localsearch	1.653333	1959	856.46190	272351.30	300	75	1
8	instance	2.130000	14347	527.89902	284162.71	300	0	0
1	algorithm	2.173333	3080	326.89585	100954.69	300	41	1
10	nnls	2.456667	5704	517.21857	198419.81	300	59	0
4	beta	2.486667	7444	126.92801	60572.11	300	14	0
13	rho	2.756667	7714	263.67639	102695.19	300	0	0
2	alpha	2.826667	7513	191.15959	77079.34	300	3	0
3	ants	2.896667	7278	163.44388	63050.53	300	10	0
7	elitistants	3.310700	1441	133.82985	45545.36	297	6	1
11	q0	3.360000	2043	120.51072	48412.48	300	24	1
6	dummy	3.433333	5926	74.73502	34987.67	300	0	0
5	dlb	3.516667	1622	459.31923	167487.58	300	68	1
12	rasrank	5.250000	1833	22.13809	10443.61	300	0	1

Iteration 3:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
9	localsearch	1.633333	1599	424.18018	166858.077	300	87	1.0e+00
5	dlb	1.736667	1152	354.01060	142080.693	300	73	1.0e+00
11	q0	2.236667	4300	341.34645	111054.290	300	36	1.0e+00
8	instance	2.293333	16403	254.48687	198306.752	300	0	0.0e+00
10	nnls	2.326667	6016	228.40351	103649.692	300	40	0.0e+00
1	algorithm	2.483333	1900	276.77600	85229.335	300	34	1.0e+00
3	ants	2.860000	6987	142.49097	59962.167	300	1	0.0e+00
13	rho	2.893333	6931	195.64315	76319.855	300	0	0.0e+00
2	alpha	2.923333	6938	147.32172	72367.588	300	7	0.0e+00
4	beta	3.386667	7121	138.93112	58067.791	300	3	0.0e+00
7	elitistants	3.564678	825	49.75965	21678.069	289	16	1.0e+00
6	dummy	3.620000	5447	61.77168	30361.173	300	0	6.5e-06
12	rasrank	5.117156	1247	11.81127	6150.341	296	3	1.0e+00

Iteration 4:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
11	q0	2.053333	5856	456.7959507	204757.551	300	87	0
8	instance	2.073333	19142	109.9454705	175248.182	300	0	0
5	dlb	2.248589	682	198.3924444	88471.534	299	55	1
10	nnls	2.326667	6954	448.4743255	199731.383	300	17	0
2	alpha	2.523333	7644	274.5870132	132111.369	300	10	0
3	ants	2.713333	7117	296.1679254	115488.913	300	4	0
1	algorithm	2.751267	759	297.5547136	121250.866	282	79	1
13	rho	2.876667	6881	216.1708660	104769.426	300	3	0
4	beta	2.893333	7446	258.7484972	133414.168	300	2	0
9	localsearch	3.254356	648	127.6228942	63932.308	296	27	1
6	dummy	3.510000	5839	115.5181506	51393.628	300	0	0
12	rasrank	4.753311	708	16.3466477	10052.173	274	15	1
7	elitistants	5.405944	294	0.8956393	2033.879	205	1	1

Iteration 5:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
11	q0	1.730000	8907	631.799954	405570.317	300	86	0
10	nnls	1.793333	7726	652.216438	327392.263	300	60	0
2	alpha	2.076667	10032	570.609297	422427.239	300	4	0
4	beta	2.256667	9377	514.240922	286393.678	300	47	0
8	instance	2.570000	26592	25.293135	442174.130	300	0	0
3	ants	2.820000	8565	460.217172	269344.470	300	10	0
1	algorithm	3.151333	484	111.714494	78657.781	264	48	1
6	dummy	3.156667	7345	272.837655	155610.822	300	13	0
13	rho	3.170000	8740	297.850763	201881.612	300	0	0
9	localsearch	3.792667	613	85.791078	78460.031	288	19	1
12	rasrank	4.898722	366	15.710155	14068.459	227	11	1
5	dlb	5.754167	324	23.482112	25171.625	225	1	1
7	elitistants	6.489111	161	2.634057	3957.041	124	1	1

Iteration 6:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
9	rho	1.510000	2940	56.5464952	15375.0485	300	91	1.0000000
7	nnls	1.533333	2613	67.1819365	16854.5277	300	79	1.0000000
3	beta	1.910000	3303	57.7862766	15414.2988	300	46	0.9954526
1	alpha	2.156667	3198	49.1196373	12024.0450	300	37	0.9999972
6	instance	2.323333	10383	3.5939575	22868.4203	300	2	0.0000000
8	q0	2.510000	2939	35.3343544	9017.0381	300	26	1.0000000
2	ants	2.740000	2949	27.1146465	8127.0565	300	12	1.0000000
5	dummy	2.936667	2586	22.6621042	6569.6271	300	4	1.0000000
4	dlb	6.300144	104	0.3378027	307.4434	97	3	1.0000000

Iteration 7:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
8	rho	1.829067	1512	21.577855	2659.569	298	72	1.0000000
1	alpha	2.026667	1612	13.571266	2159.330	300	40	0.9999979
3	beta	2.122867	1616	16.259031	2652.925	299	24	0.9999965
7	q0	2.174800	1427	15.985143	2139.691	296	59	1.0000000
6	nnls	2.422933	1167	10.839451	1424.963	292	58	1.0000000
5	instance	2.456667	4282	-2.259904	2787.891	300	5	0.0000000
2	ants	2.562867	1495	9.754629	1543.823	299	22	1.0000000
4	dummy	2.873333	1217	10.871563	1492.114	300	20	1.0000000

Iteration 8:

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
5	instance	1.518395	2752	-0.2808745	652.48853	299	61	0.0000000
1	alpha	2.064348	887	0.4243390	90.76926	283	63	0.9999994
3	beta	2.184348	944	0.2598003	76.74807	291	63	0.9980643
7	q0	2.667291	838	0.2098439	51.60506	281	31	1.0000000
4	dummy	2.766622	850	0.0193551	45.77732	276	22	1.0000000
8	rho	3.152174	793	0.1191928	43.50123	274	13	1.0000000
2	ants	3.305017	656	0.0647227	34.93565	269	14	1.0000000
6	nnls	3.583344	514	0.0351768	31.83332	246	33	1.0000000

The 5 most important parameters per iteration are :

```
## [[1]]
## [1] "localsearch" "instance"    "elitistants" "alpha"       "algorithm"
##
## [[2]]
## [1] "localsearch" "instance"    "algorithm"   "nnls"        "beta"
##
## [[3]]
## [1] "localsearch" "instance"
##
## [[4]]
## [1] "instance" "nnls"     "alpha"
##
## [[5]]
## [1] "q0"       "nnls"     "alpha"    "beta"     "instance"
##
## [[6]]
## [1] "rho"      "nnls"     "beta"     "alpha"    "instance"
##
## [[7]]
## [1] "alpha"    "rho"      "beta"     "instance" "q0"
##
## [[8]]
## [1] "instance" "alpha"    "beta"     "q0"
```

Parameter relevant interactions:

Iteration 1:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth

Iteration 2:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth
beta	nnls	1.057047	489	nnls:beta	1.800000
nnls	localsearch	1.600705	548	localsearch:nnls	1.563333

Iteration 3:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth

Iteration 4:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth
nnls	alpha	1.04625	563	alpha:nnls	1.366667

Iteration 5:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth
beta	q0	1.279362	562	q0:beta	1.656667

Iteration 6:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth
beta	rho	1.916484	517	rho:beta	1.666667
alpha	rho	2.223607	498	rho:alpha	1.773333

Iteration 7:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth

Iteration 8:

variable	root_variable	mean_min_depth	occurrences	interaction	uncond_mean_min_depth