Justified Counterfactual Explanations: An Untapped Way to Increase Trust (Supplementary Material)

Alejandro Kuratomi* Zed Lee* Panagiotis Papapetrou* Tony Lindgren*

1 r values evaluated

Assume n features, n_{cont} ontinuous features, and the distance matrix DM of $Data$. The 18 r values considered are:

1. $r_1 = \sqrt{n}/5$
2. $r_2 = \sqrt{n}/10$
3. $r_3 = \sqrt{n}/50$
4. $r_4 = \sqrt{n}/100$
5. $r_5 = \sqrt{n_{cont}}/5$
6. $r_6 = \sqrt{n_{cont}}/10$
7. $r_7 = \sqrt{n_{cont}}/50$
8. $r_8 = \sqrt{n_{cont}}/100$
9. $r_9 = max([minPerRow(DM)])$
10. $r_{10} = r_9/5$
11. $r_{11} = r_9/10$
12. $r_{12} = r_9/50$
13. $r_{13} = r_9/100$
14. $r_{14} = min(DM[0,:])$
15. $r_{15} = r_{14} + (r_{14} + max(DM[0,:]))/5$
16. $r_{16} = r_{14} + (r_{14} + max(DM[0,:]))/10$
17. $r_{17} = r_{14} + (r_{14} + +max(DM[0,:]))/50$
18. $r_{18} = r_{14} + (r_{14} + +max(DM[0,:]))/100$

r_{14} is the vector of minimum values per row of matrix DM. r_5 was chosen as the best r value.

*Department of Computer and Systems Sciences - DSV, Stockholm University {alejandro.kuratomi, zed.lee, panagiotis, tony}@dsv.su.se

2 Synthetic datasets generation functions

The following functions are used to generate the synthetic datasets.

$$S1 = \begin{cases} (N(2,0.5), N(2,0.5)), class = blue \\ (N(6,0.5), N(6,0.5)), class = orange \end{cases}$$

$$S2 = \begin{cases} (N(2,0.5), N(2,0.5)), class = blue \\ (N(2,0.5), N(6,0.5)), class = orange \\ (N(6,0.5), N(2,0.5)), class = orange \\ (N(6,0.5), N(6,0.5)), class = orange \end{cases}$$

$$S3 = (N(4,0.5), N(4,0.5)), both$$

$$S4 = \begin{cases} (0, N(3,0.5)), class = blue \\ (1, N(5,0.5)), class = orange \end{cases}$$

$$S5 = \begin{cases} (0, N(2,0.5)), class = blue \\ (0, N(6,0.5)), class = blue \\ (1, N(2,0.5)), class = orange \\ (1, N(6,0.5)), class = orange \end{cases}$$

$$S6 = (random(0,1), N(4,0.5)), both$$

3 NAJ algorithm run time ($1/100^{th}s$)

	MO	FT	NT	RT	DJN	IJN	CIJN
S1	0.11	2.41	0.00	0.00	0.14	0.82	0.84
S2	0.17	1.67	0.00	0.00	0.13	0.74	0.77
S3	0.24	1.79	0.00	0.00	7.80	1.62	1.70
S4	0.00	0.97	0.01	0.00	0.15	0.26	0.25
S5	0.02	1.34	0.01	0.00	0.11	0.38	0.39
S6	0.01	1.24	0.00	0.00	0.08	0.17	0.18
U1	9.67	54.3	0.00	0.00	124	103	68.6
U2	1.00	1.4e3	0.01	0.01	169	171	175
U3	0.01	157	0.00	0.01	0.24	1.40	0.78
U4	1.05	43.9	0.01	0.00	21.2	37.3	23.6
U5	1.16	1.6	0.01	0.01	0.19	1.25	1.45
U6	0.01	1.2e6	0.01	0.01	0.18	0.09	0.09
Avg. S	0.09	1.57	0.00	0.00	1.40	0.66	0.69
Avg. U	2.15	1.9e5	0.00	0.01	52.5	52.4	44.8