

1 **Common comments.** Some reviewers raised the issue of the paper’s “*narrow interest*” given its “*focused contribu-*  
2 *tion*”, and due to linear classifiers (LCs) being “*interpretable*”. We cannot agree with the reviewers. First, LCs and  
3 NBCs are extensively used in different settings, with NBCs being deemed by some as one of the top algorithms in  
4 data mining. Second, the best-known heuristic explainers identify simple (local) linear models as a way of explaining  
5 complex ML models. Our work can be used jointly with these heuristic explainers for computing (local) PI-explanations.  
6 For instance, heuristic explainers can be used to produce more complex (and more rigorous) linear models, with more  
7 features, from which PI-explanations can then be computed. Clearly, one can envision other research directions relating  
8 heuristic and rigorous explanations. Third, interpretability depends on the number of features one needs to account for.

9 **Review #1. Q2:** We are very happy to read the reviewer’s comment : “*this will really shift the way I think about this*  
10 *problem, and ...*”. We believe our results will impact a broader community if the paper is accepted.

11 **Q3:** It should be underscored that linear models are used for explaining complex ML models. Until our work, there was  
12 no efficient solution for computing a PI-explanation in those cases, and now there is one.

13 **Q8:** We will include a reference to the KR’20 paper. Thanks for the reference. However, that paper was not available  
14 when our paper was submitted. There was an earlier CoRR report, but it is from April 2020. The following paragraph  
15 merits comments: “*For me, this paper provides an interesting result, ... For example, the above paper by Shih et al.*  
16 *applied this algorithm for compiling linear classifiers towards ...*”. (i) We see a direct connection with neural networks  
17 and other black-box ML models. Well-known heuristic explainers find local linear models given a complex ML model  
18 and an instance. In those cases, we can compute PI-explanations. An efficient algorithm for computing PI-explanations  
19 of linear models will likely motivate additional results. (ii) The work of Shih et al. started in 2018 (IJCAI 2018) and  
20 focused solely on NBCs and monotone BNCs, with experimental results only for NBCs. The work has since been  
21 extended, but it started with a worst-case exponential in time and space approach for explaining NBCs.

22 **Review #2. Q3:** We will cover the references mentioned by the reviewer. These do not affect the novelty of our  
23 results, but are of course important for completeness. Thanks for the references.

24 **Q4:** We reduce NBCs to XLCs so that we can develop a unified algorithmic solution. The results are for XLCs but,  
25 given the transformation, apply to NBCs as well. Regarding the experiments, the main issue was the available space.  
26 We will include further additional detail in the supplementary materials, which already include additional detail.

27 **Q5:** We will address the comments on notation.

28 **Q6:** The references will be added and we will relate those with our work.

29 **Review #3. Q3:** The following statement merits comments: “*On a similar note, ..., there will be many PI-explanations*  
30 *(thousands to even millions) per instance, which ...*”. First, we showed these results solely to demonstrate that our  
31 algorithms scale to large problems, with many features and with many explanations. Second, analysis by hand of  
32 millions of PI-explanations is unrealistic. However, such PI-explanations can be analyzed automatically, e.g. to gather  
33 statistics and test hypotheses. Also, we can enumerate PI-explanations that respect some additional properties. There  
34 are many possible scenarios, as long as enumeration of PI-explanations is easy, which it now is.

35 **Q5:** We opted for a restricted version of XLC to describe the algorithms to keep the notational overhead to a minimum.  
36 Also, all proofs are included in the supplementary materials.

37 **Q8:** The reviewer is correct that weights in linear classifiers are heuristic. However, that is orthogonal to our work.  
38 If one fixes the linear model we will compute rigorous PI-explanation in log-linear time. If a different (extended)  
39 linear model is picked, the same rationale applies. Heuristic explainers are significantly different than what we propose.  
40 Heuristic explainers either approximate a complex ML model with a linear model (in the case of LIME) or heuristically  
41 identify a set of literals as a tentative explanation (in the case of Anchor). As stated above, our approach can be used  
42 with LIME or SHAP, or other heuristic explainers that identify a linear model. And this extends significantly the reach  
43 of our work, but also the reach of earlier work. We will address all comments regarding readability.

44 **Review #4. Q3:** The review is not entirely correct since our algorithm is not based on “*greedily flip the least impactful*  
45 *variables*”. No variables are flipped; flipping is only used for implementing backtracking. Our algorithm works in such  
46 a way that entailment is guaranteed at each step, this without directly checking entailment with an NP oracle. This  
47 is novel and likely to change the way these problems are looked at (as noted by Reviewer #1). Before our work, the  
48 best only approach for computing PI-explanations of NBCs was given in [29], but this is worst-case exponential. The  
49 approach hinted by the comment “*keeping the most impactful features to guarantee*” is insufficient, and would require  
50 additional reasoning to ensure that entailment is preserved. This is what our algorithm does, and that reflects part of  
51 the novelty. Also, our algorithm enumerates explanations with log-linear delay. No heuristic approach for computing  
52 explanations is capable of deterministically enumerating explanations. We kindly ask the reviewer for references that  
53 might cast doubt on the novelty of our work. Otherwise, we would expect the reviewer to revise their scores.

54 **Q8:** Our results do not follow from the work of Nordh and Zanuttini since a linear classifier is a global function  
55 employing real coefficients whereas their work considers language-tractability of purely logical functions.