**Review of DeltaAccuracy pseudocode (Algorithm 1), v.01, 31/12/23**

Line 3, expand ‘sqrt’ to ‘sqrt(m)’, where m (or another symbol) is the number of features in D\_a and D\_b (I presume, assuming D\_a and D\_b have the same number of features)

Line 7, replace ‘b’ by ‘D\_b’

The method name ‘fi’ is not good because ‘feature importance’ is too generic and technically ‘ms’ is also a feature importance method (it measures the importance of a feature in terms of how often it is included in the set of minimally sufficient features).

To avoid confusion, I suggest to rename ‘fi’ to ‘dfvp’ (difference of feature value probabilities) in lines 13, 15; and keep ‘ms’. (Or, if you prefer acronyms with just two letter, you could use ‘dp’ for ‘difference of probabilities’, less precise but shorter, and still more precise than the very vague ‘fi’.) This also helps to avoid some overloading of notation, because in the current code, ‘fi’ is used both as the value of the ‘method’ variable and as a variable by itself in lines 18-20. Keep using ‘fi’ as a variable in lines 18-20.

Also, in line 13, I think ‘dfvp\_b’ is better (more precise) than just ‘dfvp’; and in line 21, ‘ms\_b’ is more precise than just ‘ms’. Hence, the set of possible values for ‘method’ will be {dfvp\_b, dfvp\_b-a, ms\_b, ms\_b-a}.

In line 18, it seems you need to take the absolute value of the difference. Otherwise you can get negative values for ‘fi’, which seems problematic when computing ‘fi’ in line 18 and ‘sum(fi)’ in line 20, since positive and negative values could cancel themselves out. Does it make sense to have a negative value of ‘fi’? An analogous problem seems to occur in lines 26 and 28. But this also depends on line 19 (and 27), which I don’t quite understand.

In line 19, what exactly the ‘MinMax’ function is doing? Put that into comments in the pseudocode, e.g. after some comment symbol like ‘//’ after the instruction (in the same line).

In the first, second and fourth lines without number after the line 30, these two lines seem useless, each is just assigning a variable to itself, no change. (By the way, those lines without a number should have a line number, like the other lines.)

Finally, the pseudocode does not explicitly refer to individual features, it implicitly assumes that when variables like fa, fb, fi, etc. are computed, they are computed for each feature separately. An alternative would be to introduce a loop over the features, indexing each feature by k (since ‘i’ is already used to denote ‘importance’ in ‘fi’). But this would make the pseudocode a bit more complicated, not sure if it is worth. For now, if you keep the pseudocode as it is, without the loop over features, add a note at the top of Algorithm 1 mentioning that the variables fa, fb, fi, ma, mb, ms, featureBias are vectors of feature importance scores (one score for each feature), and as vectors they should all be in boldface. It should also be explicitly mentioned that the ‘sum’ in lines 20 and 28 is computed over all feature importance scores (over all features), right?