Report on results for Ingrid's AD study

Luca <pgl@portamana.org>

17 December 2021; updated 23 December 2021

Here I report the results for the features in Ingrid's study. See previous report for technical details and further explanations.

All material is available at https://github.com/pglpm/ADBayes.

1 Main results

Our general problem: to prognose the future onset of AD as opposed to stable MCI – predictand binary variate `Subgroup_num_` – given the set of twelve features: `AGE`, `ANARTERR_neuro`, `AVDEL30MIN_neuro`, `AVDELTOT_neuro`, `Apoe4_`, `CATANIMSC_neuro`, `GDTOTAL_gds`, `Gender_num_`, `LRHHC_n_long`, `RAVLT_immediate`, `TRAASCOR_neuro`, `TRABSCOR_neuro`. In the following I use shorter names for them.

The results are reported for three different calculation set-ups, denoted `narrow`, `broad`, `all`, explained in the previous report. The first two use only the training dataset and have different prior smoothness preferences for the inference: `narrow` and `broad`. The third uses `all` datapoints, with a broad prior smoothness preference.

1.1 Results for individual features

This is the mutual information for each feature if it were used *individually* to make the prognosis, ranked from highest to lowest according to the `all` setup. One standard deviation in the numerical computation error is also given:

feature	mutual information/bit		
	narrow	broad	all
RAVLT	0.11 ± 0.02	0.08 ± 0.01	0.08 ± 0.01
AVDEL30MIN	0.11 ± 0.02	0.08 ± 0.01	0.08 ± 0.01
AVDELTOT	0.076 ± 0.01	0.057 ± 0.01	0.06 ± 0.01
TRABSCOR	0.04 ± 0.01	0.04 ± 0.01	0.030 ± 0.008
LRHHC	0.04 ± 0.01	0.025 ± 0.008	0.020 ± 0.008
CATANIMSC	0.030 ± 0.009	0.023 ± 0.008	0.019 ± 0.007
TRAASCOR	0.031 ± 0.009	0.023 ± 0.007	0.014 ± 0.006
Apoe4	0.008 ± 0.005	0.004 ± 0.004	0.006 ± 0.004
AGE	0.004 ± 0.004	0.005 ± 0.004	0.003 ± 0.003
Gender	0.004 ± 0.003	0.003 ± 0.003	0.002 ± 0.002
ANARTERR	0.003 ± 0.003	0.001 ± 0.003	0.001 ± 0.002
GDTOTAL	0.001 ± 0.002	0.000 ± 0.001	0.0002 ± 0.0008

mutual info cannot be negative, so the '±' intervals must be understood to have a lower bound of at least zero

The rank is the same in all computation set-ups; the values agree within their numerical errors. These values give us not only a ranking, but also an estimate of the predictive power. See the previous report for what these numbers actually mean.

1.2 Results for joint feature set

This is the mutual information for the set of features, used *jointly*:

narrow	broad	all
(0.27 ± 0.02) bit	(0.15 ± 0.02) bit	(0.15 ± 0.02) bit

As explained in the previous report, with a mutual information of 0.15 bit, in 100 new prognoses we can expect that between 58 ± 9 and 72 ± 9 will be correct. Analogously, a mutual information of 0.27 bit means from 64 ± 9 to 79 ± 8 correct prognoses in 100 new cases.

1.3 'Importance' of individual features when used jointly

Here is the 'importance' percent of the individual features, as defined in the previous report, ranked from highest to lowest according to the `all` setup:

feature	$\Delta I_{\setminus \text{feature}}/\%$		
	narrow	broad	all
TRABSCOR	13.9 ± 0.8	10.2 ± 0.6	11.5 ± 0.9
ANARTERR	6.9 ± 0.4	6.5 ± 0.5	6.0 ± 0.5
AVDEL30MIN	3.0 ± 0.2	2.4 ± 0.1	3.7 ± 0.3
AVDELTOT	1.32 ± 0.06	1.7 ± 0.1	2.0 ± 0.2
RAVLT	2.4 ± 0.2	3.0 ± 0.3	2.0 ± 0.1
TRAASCOR	3.4 ± 0.2	3.2 ± 0.2	1.68 ± 0.08
AGE	0.110 ± 0.001	0.55 ± 0.06	0.71 ± 0.08
GDTOTAL	0.58 ± 0.04	0.203 ± 0.004	0.47 ± 0.04
Apoe4	0.35 ± 0.04	0.55 ± 0.07	0.33 ± 0.04
LRHHC	0.50 ± 0.04	0.65 ± 0.07	0.30 ± 0.02
CATANIMSC	0.212 ± 0.004	0.71 ± 0.06	0.2541 ± 0.0005
Gender	0.33 ± 0.03	0.32 ± 0.04	0.16 ± 0.01

this percentage cannot be negative, so the '±' intervals must be understood to have a lower bound of at least zero

Note that the percentages do not add up to 100%, nor should they, owing to the reasons given in the previous report.

We see that omitting TRABSCOR_neuro or ANARTERR from the twelve features would reduce the mutual information by roughly 10% or 6%; that is, from a value of 0.15 bit to roughly 0.14 bit. This also means that in 100 new prognoses we would drop from a best case of 72 ± 9 correct ones to a best case of around 71 ± 9. In clinical-importance terms this difference is not small: it means on average roughly 10 000 additional incorrect predictions (more exactly around 7 300) every million prognoses.

1.4 Further remarks

• The mutual informations estimated above are calculated assuming that *we will not receive further training data*. It is also possible to calculate an estimate of what the mutual informations would be if we had a very large number of training data. I'm currently setting up the code to calculate this estimate.

• As shown in the enclosed plots (all coming from the all setup), many features have two clearly different population distributions for the future AD or MCI patients. This means that something is already at work affecting those features. However, the two population distributions have very large overlaps, so the feature does not help very much in

making predictions for a single individual, as clear also from the mutual informations for the individual features.

• A further analysis of direct and inverse conditional frequencies in the population suggest that the dataset might have some biases for some features. These biases, however, could be overcome by appropriately combining inverse and direct predictions from the features. This point and topic would require a very extensive discussion, so I won't write any further details here.

• As explained in the previous report, the calculation also allows us to count the number of peaks in the joint distribution of predictand and features, possibly hinting at the presence of subpopulations. Let me know if a rough estimate of the peaks is needed.

See previous report for further technical remarks.