Master Thesis Defense:
Bayesian Modeling in Neurological Diseases

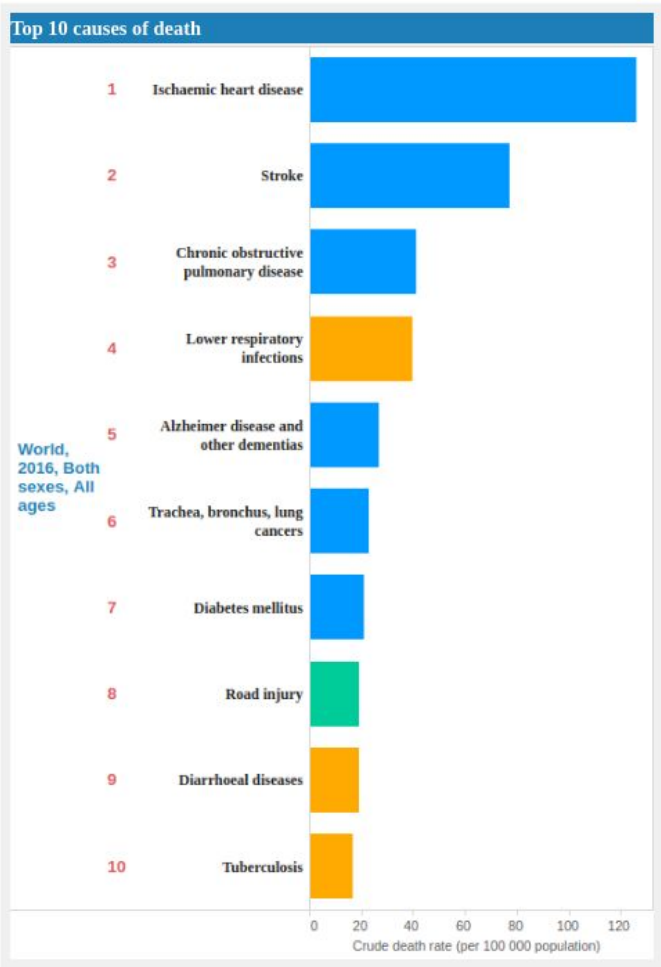
Hüseyin Camalan

Supervisors:
Prof.Dr. Kerstin Ritter
Dr. Michelle Livne

Neurological Diseases

- are difficult to cure
- damage to nerve tissue is largely irreversible

- Neurological diseases associated with aging present the worst portrait:

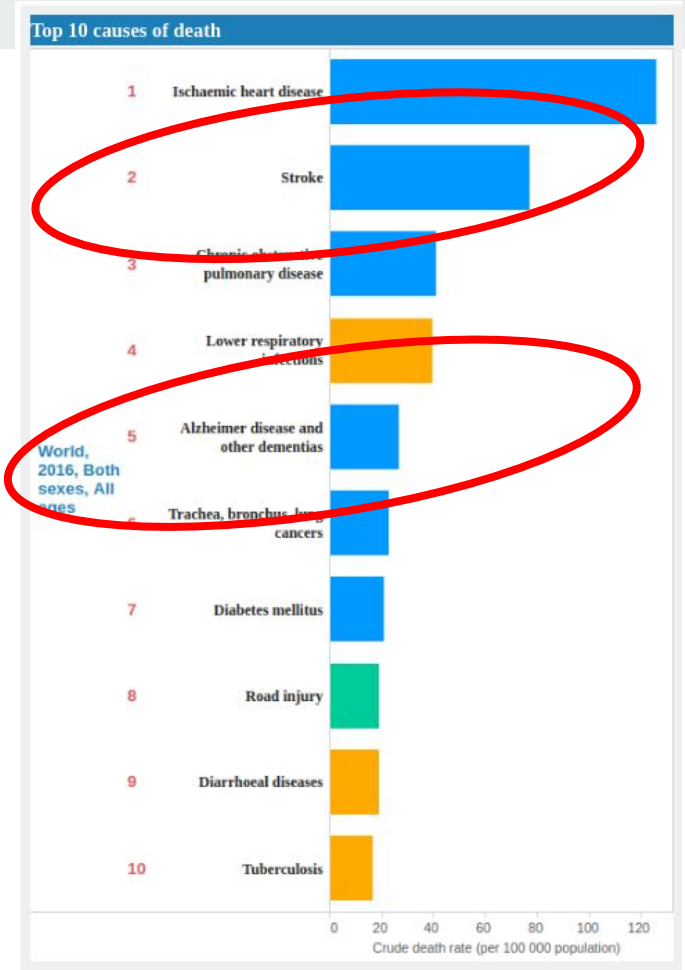


https://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/

Neurological Diseases

- are difficult to cure
- damage to nerve tissue is largely irreversible

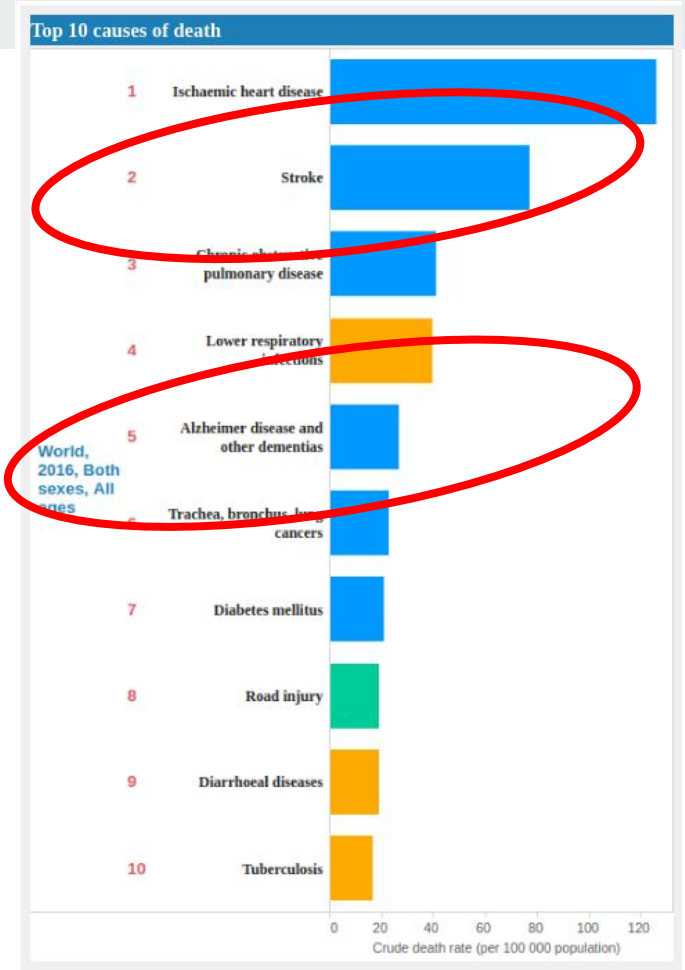
- Neurological diseases associated with aging present the worst portrait:



https://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/

Neurological Diseases

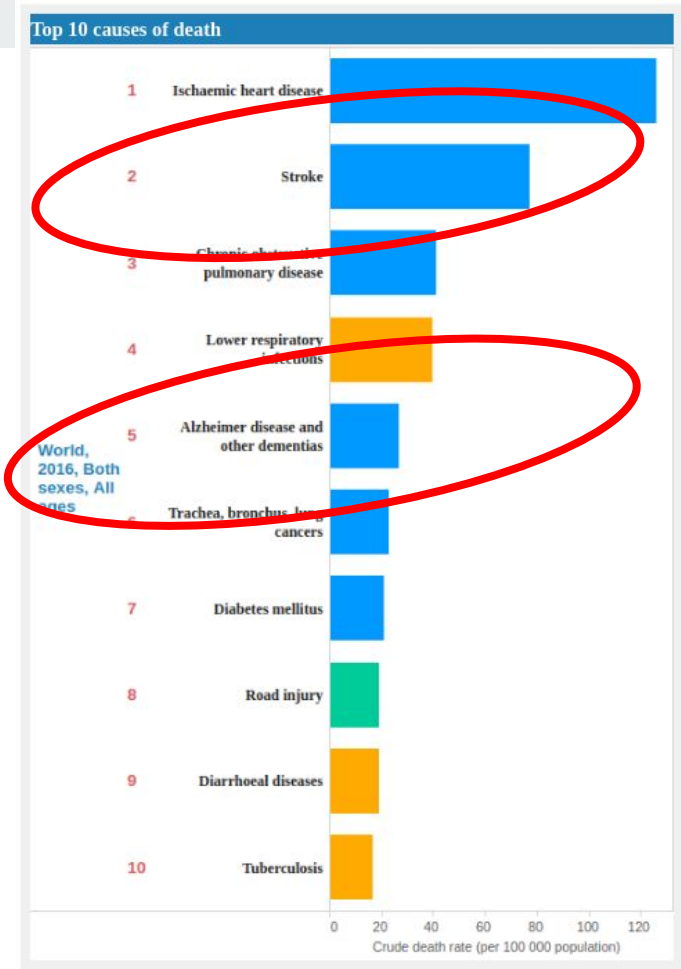
- Both stroke and dementia increased twofold in the last two decades and are projected to increase twofold again
- Population distributions skewing towards older ages is a likely culprit in this increase



https://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/

Neurological Diseases

- From here onwards, focus will be on dementia/Alzheimer's disease
- However, most statements are mutually applicable and the separation between the two diseases is also not completely clear
    - e.g. vascular dementia



https://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/

Dementia

- Umbrella term for diseases that cause cognitive decline through degradation of brain tissue
- Various disease mechanisms at play
  - vascular, oxidative, genetic, environmental (among others)
- Roughly %75 Alzheimer's disease (AD), followed by %20 vascular dementia
  - AD: characterized by memory loss and damage to corresponding brain regions (especially hippocampus)

Dementia

- Prevention efforts currently constitute the best option
  - mainly based around a healthy lifestyle and mental stimulation
  - prevention delays disease onset, not a definitive cure
- Early identification of at-risk-individuals could be valuable

Total potentially modifiable risk factors for dementia

Risk Factor	Relative risk for dementia	Prevalence (%)	Communality (%)	PAF (%)	Weighted PAF* (%)
Early life					
Less education (none or primary school only)	1.6	40.0	64.6	19.1	**7.5**
Midlife (age 45-65)					
Hypertension	1.6	8.9	57.3	5.1	**2.0**
Obesity	1.6	3.4	60.4	2.0	**0.8**
Hearing loss	1.9	31.7	46.1	23.0	**9.1**
Later life (age >65)					
Smoking	1.6	27.4	51.1	13.9	**5.5**
Depression	1.9	13.2	58.6	10.1	**4.0**
Physical inactivity	1.4	17.7	26.6	6.5	**2.6**
Low social contact	1.6	11.0	45.9	5.9	**2.3**
Diabetes	1.5	6.4	70.3	3.2	**1.2**
Total adjusted for communality					**35.0**

Table 1. Risk factors for dementia; relative risk for dementia, prevalence, communality and Population Attributable Fraction (PAF) for dementia

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., ... & Cooper, C. (2017). Dementia prevention, intervention, and care. *The Lancet*, *390*(10113), 2673-2734.

Goal:

- Build a model that can predict the risk of developing dementia/AD in the future

How to determine “risk”?

Goal:

- Build a model that can predict the risk of developing dementia/AD in the future

How to determine “risk”?

- Utilize probability theory
  - Use Bayes’ rule to derive probabilities for AD given variables

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayesian Networks on Dementia/AD

1. Specify the problem using Bayes rule
2. Factorize joint probability distributions
3. Extract individual probability terms
4. Evaluate the Bayes rule

   Test model performance on real-world data

- Alzheimer's Disease Neuroimaging Initiative (ADNI) Database

Detour: ADNI Database

- Alzheimer's Disease Neuroimaging Initiative (ADNI)
- Large scale longitudinal study of Alzheimer's disease (AD) patients, age-matched controls and patients with in-between statuses: mild cognitive impairment (MCI), significant memory concern (SMC)
- Contains demographic information, medical history, neuropsychological tests, genetic analyses, MRI scans, biological marker analyses, etc.

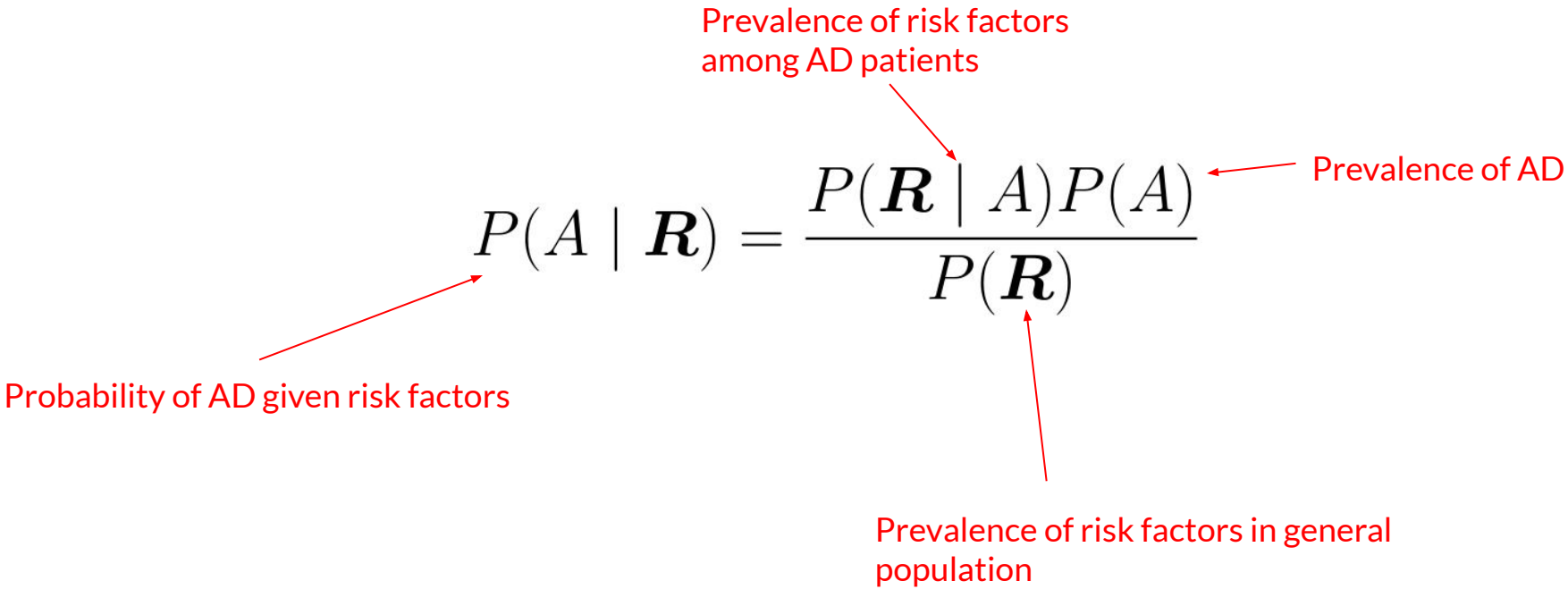
More info at https://adni.loni.usc.edu/

1. Specify the problem using Bayes rule

$$P(A \mid \boldsymbol{R}) = \frac{P(\boldsymbol{R} \mid A)P(A)}{P(\boldsymbol{R})}$$

- A: Alzheimer's disease, **R**: relevant variables/risk factors
- **R** is bold because it represents multiple variables, i.e. a joint probability distribution

1. Specify the problem using Bayes rule



2. Factorize joint probability distributions

- The equation is unsolvable until we break down the joint probability distributions P(**R**) and P(**R**|A)
- Two possible options (not exhaustive):
  1. Assume conditional independence between all variables (naive Bayes)
     - Simply multiply all individual probability terms
  2. Construct a Bayesian network of **R**, which can then be used to break the term down to its individual parts (ideally based on evidence)
     - Use scientific evidence to build the network

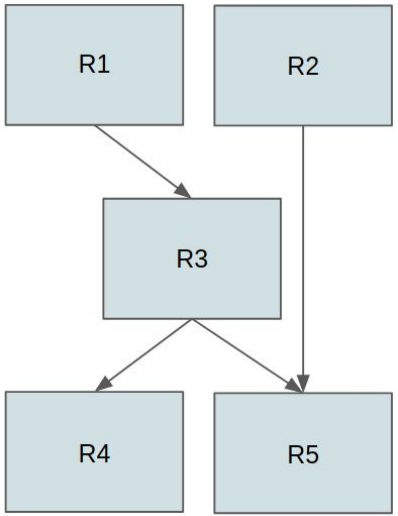
Naive Bayes

$$P(\boldsymbol{R} \mid A)P(A) =$$

$$P(R_1, ...R_N, A) = P(A) \prod_{i=1...N} P(R_i \mid A)$$

Bayesian Network

- Assume R_1...R_5 are variables that make up **R**



$$P(R_1, R_2, R_3, R_4, R_5) =$$

$$P(R_4 \mid R_3)P(R_5 \mid R_3, R_2)P(R_3|R_1)P(R_2)P(R_1)$$

Bayesian Network of ADNI Variables

- AGE and APOE4 are highly connected to brain regions and test scores
- PTEDUCAT is connected to brain size and test scores
- PTMARRY is connected to test scores
- Brain areas are assumed to be conditionally independent to each other

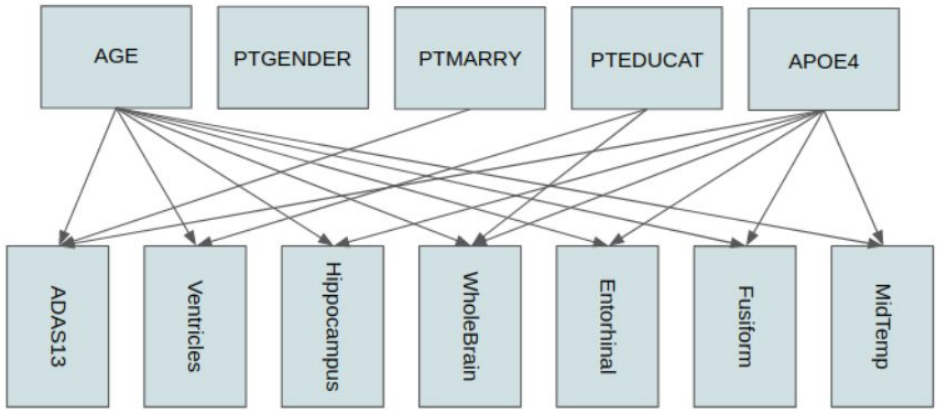


Figure 2: DAG for the selected ADNI database variables based on medical evidence. Vertical arrows indicate the direction of causality.

Extract individual probability terms

- The equation can now be solved once probabilities are plugged in
- How to look up probabilities (in the case of AD)?

Extract individual probability terms

- The Bayes rule equation can now be solved once probabilities are plugged in
- How to look up probabilities (in the case of AD)?

  Options:

1. Look up scientific literature to find descriptive statistics about relevant groups
2. **Extract probabilities from data**

Extract individual probability terms

Four ways based on **(i)** type of variable and **(ii)** its position in the Bayesian network:

1. Variable with parent nodes in the network: use regression
    a. Discrete: class probability estimates of a Logistic Regression model (not used in practice)
    b. Continuous: PDF calculated from a Bayesian Ridge Regression model
2. Variable without parent(s): use histograms
    a. Discrete
    b. Continuous: smoothed histogram values

Testing on ADNI Database

Specific research questions:

1. Can we separate patients who were initially diagnosed as AD or clinically healthy? (AD-CN)
    - Extract a subset of the data that contain only these two groups
2. Can we predict the future status of patients who are in-between AD and clinically healthy? (MCI-D)
    - a.k.a. Mild Cognitive Impairment (MCI)
    - Extract a subset of the data that contains only MCI patients
    - Predict which MCI patients develop dementia in 2 years

Testing on ADNI Database

- Train, validate and test models using 5-fold nested crossvalidation
    - 5 validation folds, i.e. roughly %65-%15-%20 training-validation-test split
    - Validation parameter is the decision threshold to classify a patient ($P_{threshold}$)
    - Test Naive Bayes (henceforth “NB”) and manually built Bayesian networks (henceforth “Manual”)
    - Test models with different normalization procedures: no normalization, z-score normalization, min-max feature scaling

Testing on ADNI Database

- Use generic machine learning models as benchmarks
  - Multilayer Perceptron: one hidden layer with 100 neurons, 200 iterations, with ReLu activation function and momentum term
    - L2 regularization as validation parameter
  - C-SVM : linear kernel
    - regularization term C as validation parameter

Results

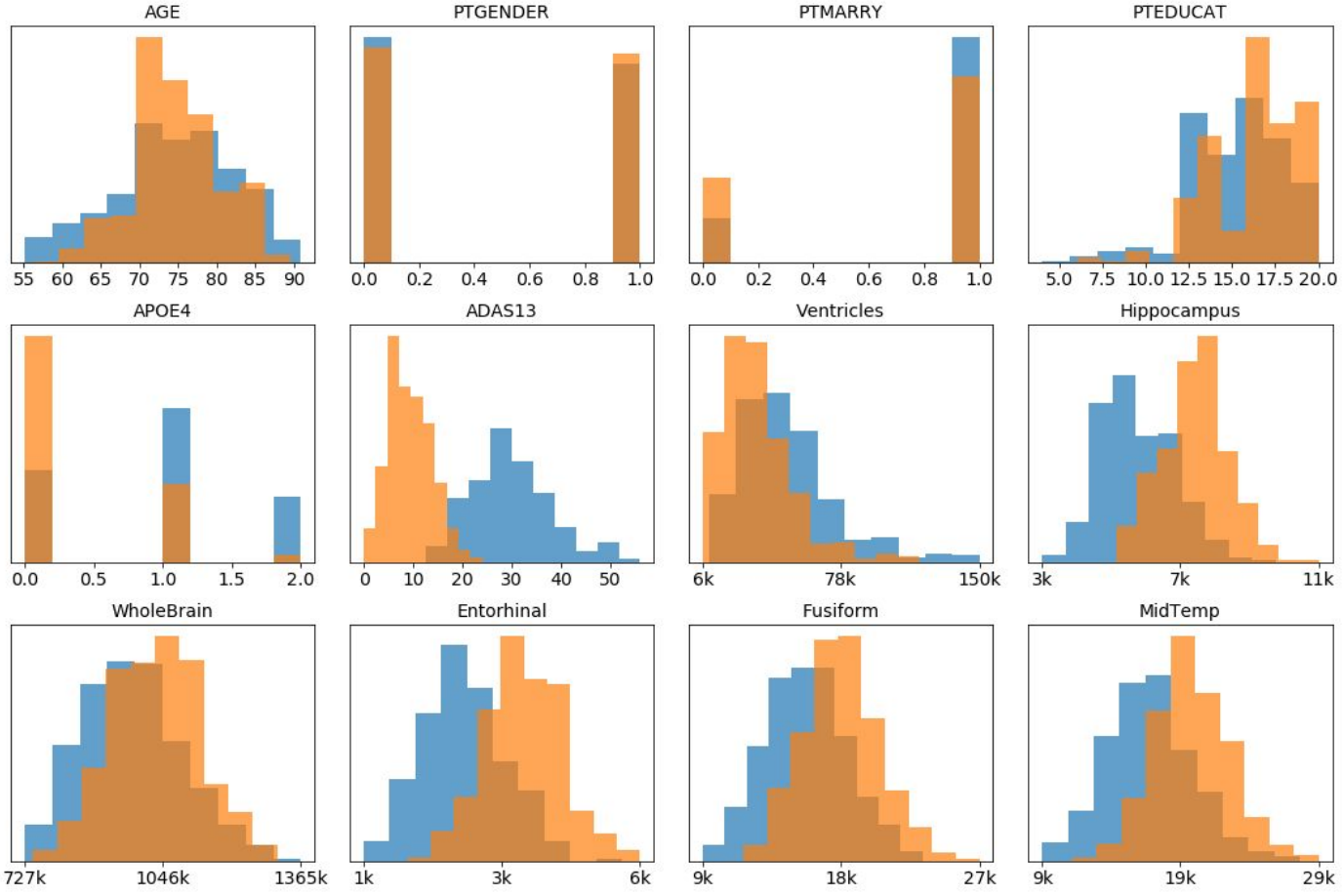
1. Variable distributions
2. Classification performance
3. Visualizations of model prediction behavior

1. Variable distributions

AD-CN Dataset

Orange - initially diagnosed as clinically normal

Blue - initially diagnosed as AD patient

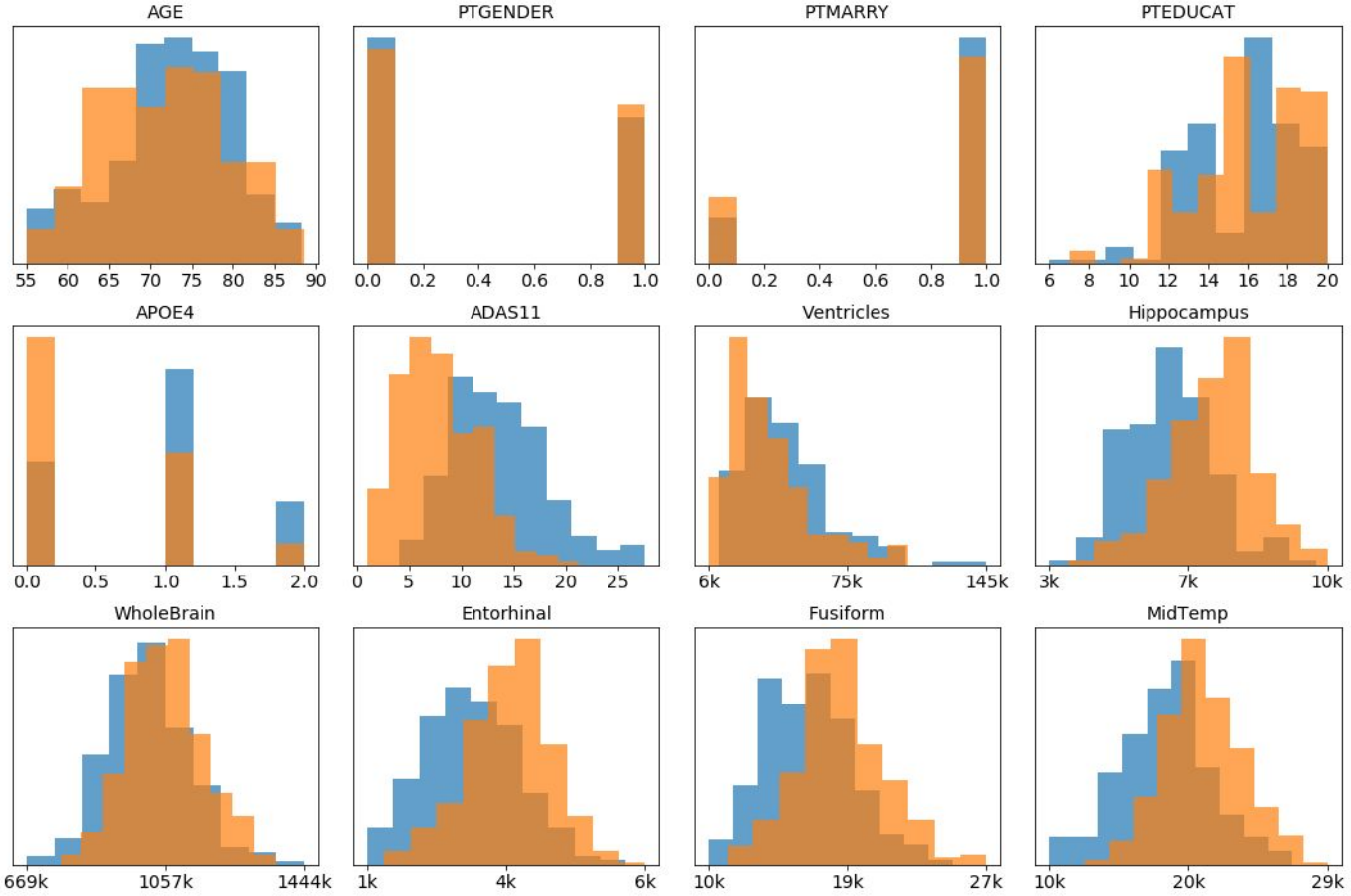


1. Variable distributions

MCI-D Dataset

Orange - does not develop dementia in two years

Blue - develops dementia in two years



2. Classification performance

AD-CN Dataset

- unnormalized > min-max > z-score
- Manual ~ NB
- MLP=C-SVM > Bayesian models
- Optimal $P_{threshold}$ is consistently quite low

Model name	Balanced accuracy	ROC-AUC score	Best parameter
NB - Unnormalized	.914	.901	$P_{threshold} = 0.1$
NB - Normalized (z-score)	.667	.638	$P_{threshold} = 0.5$
NB - Normalized (min-max)	.829	.834	$P_{threshold} = 0.1$
Manual - Unnormalized	.884	.892	$P_{threshold} = 0.1$
Manual - Normalized (z-score)	.630	.610	$P_{threshold} = 0.1$
Manual - Normalized (min-max)	.813	.796	$P_{threshold} = 0.1$
C-SVM	.972	.996	$C = 0.4$
MLP	.971	.995	$\alpha = 0.001$

2. Classification performance

MCI-D Dataset

- unnormalized ~ min-max > z-score
- Worse prediction than AD-CN
- Same as above

Model name	Balanced accuracy	ROC-AUC score	Best parameter
NB - Unnormalized	.709	.719	$P_{threshold} = 0.1$
NB - Normalized (z-score)	.558	.544	$P_{threshold} = 0.4$
NB - Normalized (min-max)	.714	.722	$P_{threshold} = 0.1$
Manual - Unnormalized	.718	.719	$P_{threshold} = 0.1$
Manual - Normalized (z-score)	.532	.539	$P_{threshold} = 0.4$
Manual - Normalized (min-max)	.713	.705	$P_{threshold} = 0.1$
C-SVM	.809	.892	$C = 0.2$
MLP	.797	.891	$\alpha = 1$

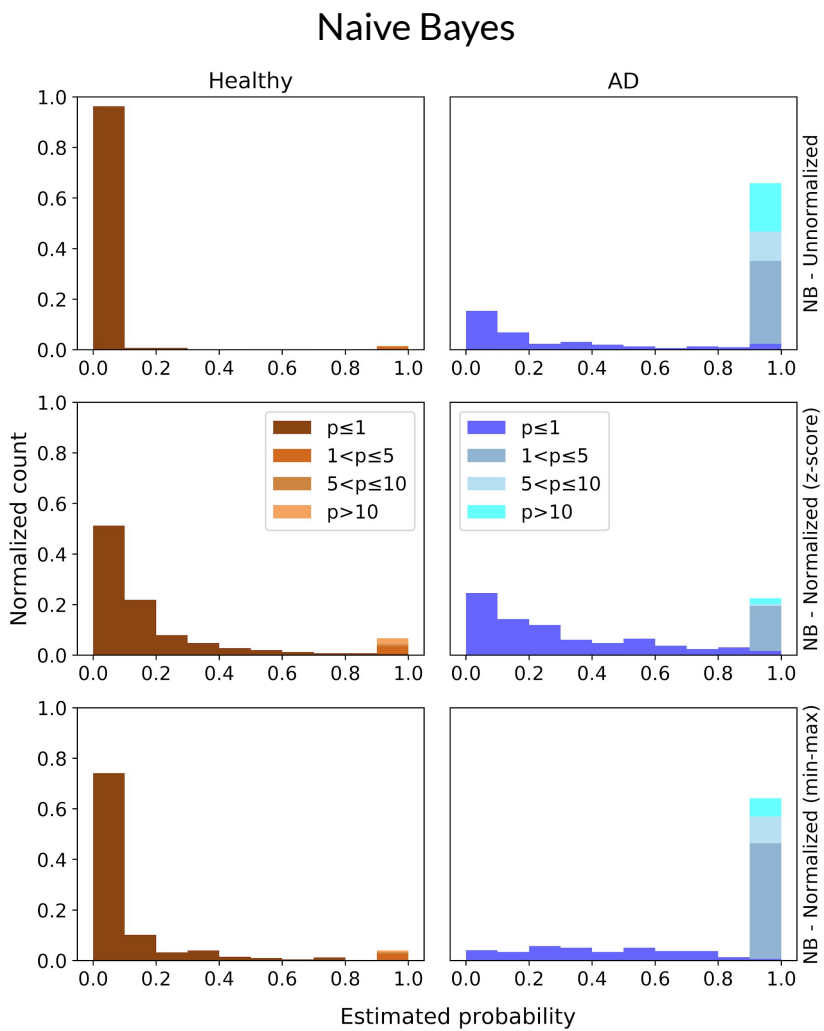
3. Visualizations of model prediction behavior

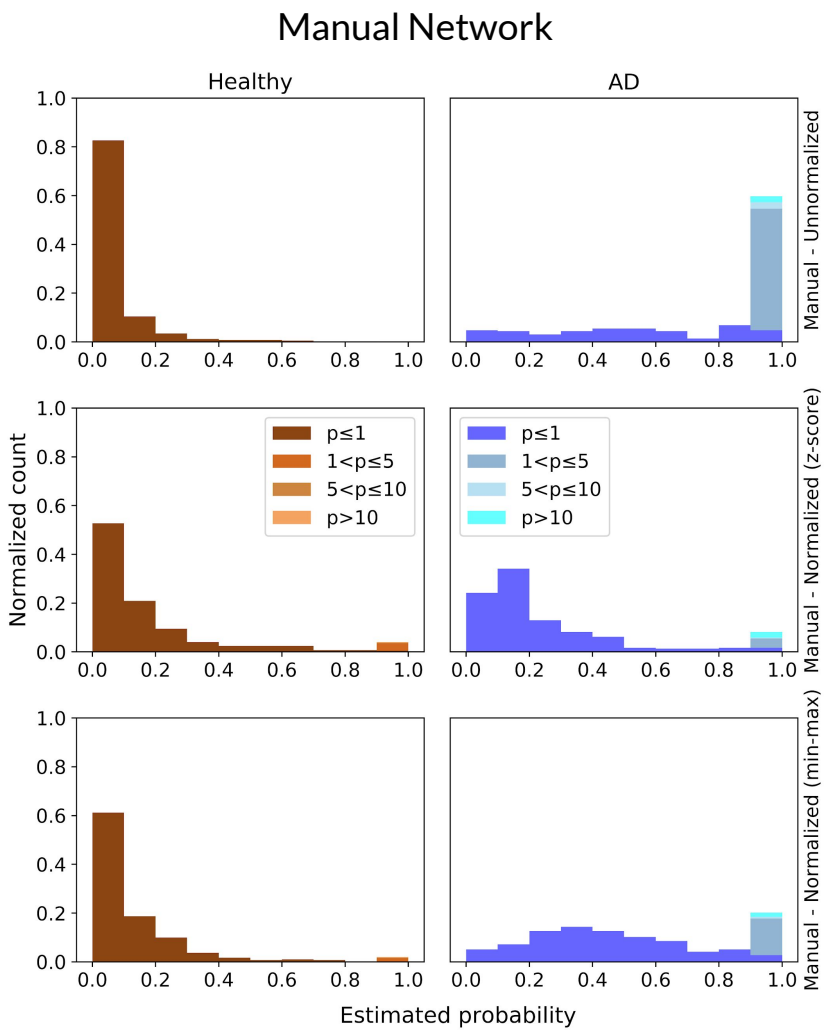
- What is causing the predictions to be so high?
- Are both class predictions affected by this behavior?

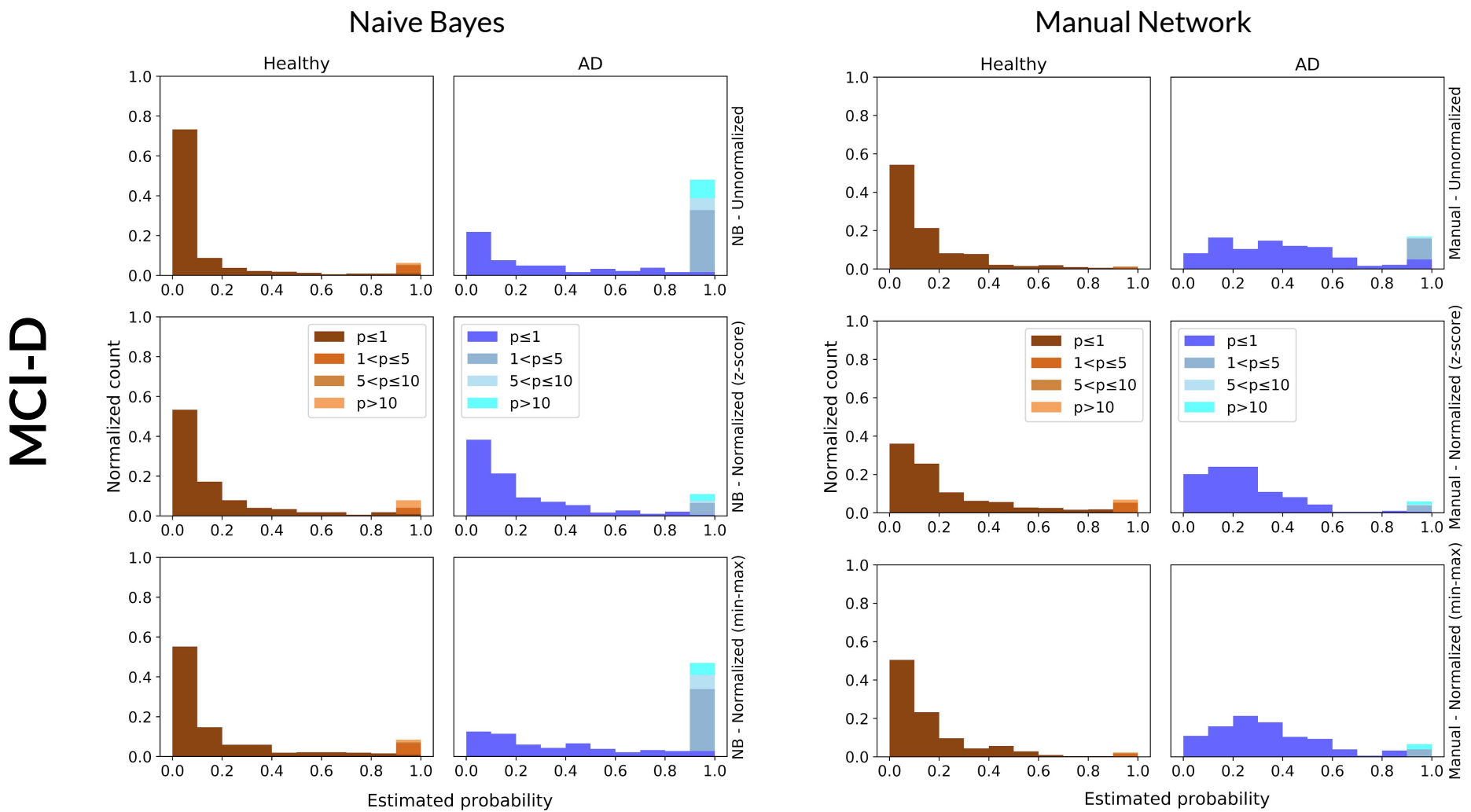
	1	2	5	10	100
NB - Unnormalized	0.724	0.778	0.869	0.918	0.988
NB - Normalized (z-score)	0.876	0.938	0.968	0.976	0.983
NB - Normalized (min-max)	0.711	0.814	0.918	0.965	0.997
Manual - Unnormalized	0.768	0.898	0.978	0.99	1.
Manual - Normalized (z-score)	0.953	0.977	0.986	0.988	0.991
Manual - Normalized (min-max)	0.917	0.976	0.987	0.991	0.997

Proportion of model predictions that are below specified values above
(ideally, all predictions would be below 1)

AD-CN


Naive Bayes
Healthy
AD
NB - Unnormalized
NB - Normalized (z-score)
NB - Normalized (min-max)
Normalized count
p≤1
1<p≤5
5<p≤10
p>10
Estimated probability

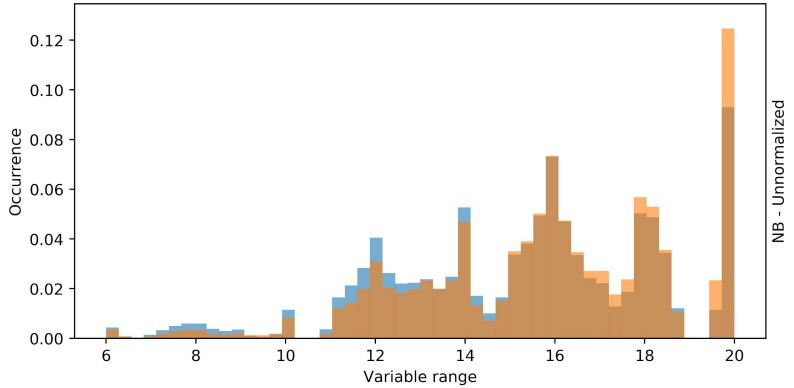

Manual Network
Healthy
AD
Manual - Unnormalized
Manual - Normalized (z-score)
Manual - Normalized (min-max)
Normalized count
p≤1
1<p≤5
5<p≤10
p>10
Estimated probability


MCI-D
Naive Bayes
Manual Network
Healthy
AD
NB - Unnormalized
NB - Normalized (z-score)
NB - Normalized (min-max)
Manual - Unnormalized
Manual - Normalized (z-score)
Manual - Normalized (min-max)
Normalized count
Estimated probability
p≤1
1<p≤5
5<p≤10
p>10

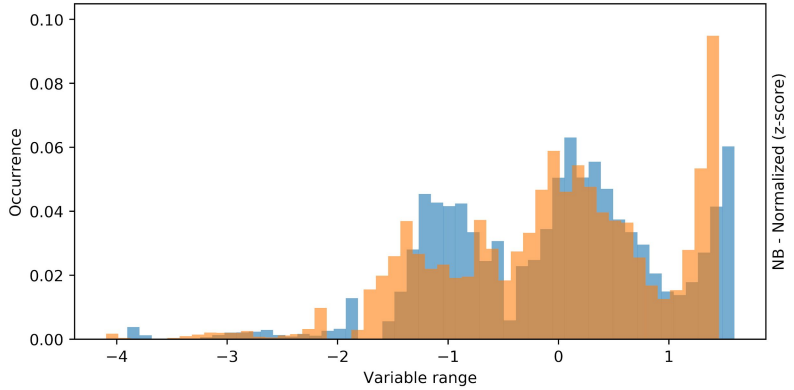
	AGE	PTGENDER	PTMARRY	PTEDUCAT	APOE4	ADAS13	Ventricles	Hippocampus	WholeBrain	Entorhinal	Fusiform	MidTemp
NB - Unnormalized	1.1	1.0	1.0	1.1	1.3	19.9	68.6	10.3	1.1	24.2	1.6	3.2
NB - Normalized (z-score)	1.2	1.0	1.1	877.1	1.3	1.1	3.1	3.7	3.1	1.0	4.2	1.8
NB - Normalized (min-max)	1.1	1.0	1.1	1.1	1.3	28.6	11.4	1.5	1.1	14.4	1.1	1.1
Manual - Unnormalized	1.1	1.0	1.1	1.1	1.3	2.6	1.1	1.3	1.2	1.3	1.1	1.1
Manual - Normalized (z-score)	1.2	1.0	1.1	879.9	1.3	1.1	1.0	1.0	1.0	1.1	1.0	1.0
Manual - Normalized (min-max)	1.1	1.0	1.1	1.1	1.3	1.1	1.1	1.2	1.1	1.2	1.2	1.1

Ratio of corresponding probability terms on the upper and lower hand of the Bayes rule (average value)

Good model



Bad model



Discussion

- Large performance differences depending on type of normalization
  - Normalization should typically not affect results because the probability extraction methods are supposed to be scale invariant
    - Smoothing of empirical variable distributions may have introduced unintended artifacts
  - There is likely an undetermined numerical problem in the z-score normalization
- Models classify AD-CN better than MCI-D
- NB performed similarly to manual models
- Bayesian models slightly worse than generic ML models
  - nature of the task may not be favoring Bayesian models

Discussion

- Advantages of Bayesian models
  - Designed to deal with uncertainties
    - May prove to be more useful depending on the problem
  - Possible to extend models without data
  - No regularization necessary
- Disadvantages of Bayesian models
  - (Apparently) not optimal for typical classification tasks
  - Probability estimations can be unrealistic

Future Directions

- Automatized Bayesian Networks
  - Building a Bayesian network on medical literature is an arduous process
    - Work scales up exponentially as variables increase
    - Conflicting evidence
    - Literature does not necessarily match data
  - Finding optimal Bayesian networks is an NP-complete problem
  - Algorithms exist, which tend to perform better than NB
    - e.g. Tree-augmented Naive Bayes

Future Directions

- How do model predictions relate to time?
- Consider the two questions:
  - “What is the probability for a patient to develop dementia?”
  - “What is the probability for a patient to develop dementia in X years?”
- Extract a new dataset with only MCI patients getting dementia and correlate it with the probability outputs of the Bayesian model

Future Directions

- Expand model with new variables
  - More biological markers, medical and family history
- Large amount of missing data
  - Requires imputation methods

Thank you for listening