Classification with pyAgrum for prediction in nursing homes

Clara CHARON, Pierre-Henri WUILLEMIN, Joël BELMIN

March 18, 2022

Summary

1. Probabilistic Classification and Graphical Models
2. Implementation of Bayesian Network Classifiers
3. Context of the Application in Nursing Homes
4. Bayesian Network Classifier for Pressure Ulcers Classification
5. Conclusion and Future Works

Probabilistic Classification and Graphical Models

- Let the features set $X_1, \ldots, X_n$ and its class Y

Definition (Probabilistic Classification)

We are searching for the classifier $\hat{C}$ such as :

$$\boldsymbol{Y} \approx \hat{\boldsymbol{Y}} = \hat{\boldsymbol{C}}\left(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\right)$$

Definition (Maximum Likelihood)

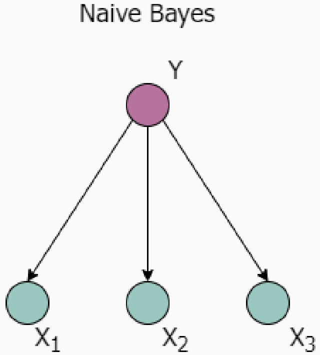
$$\hat{\boldsymbol{Y}}_{\boldsymbol{ML}} = \arg\max_{\boldsymbol{Y}} \boldsymbol{P}\left(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \middle| \boldsymbol{Y}\right)$$

Definition (Maximum A Posteriori)

$$\hat{\boldsymbol{Y}}_{\boldsymbol{MAP}} = \arg\max_{\boldsymbol{Y}} \boldsymbol{P}\left(\boldsymbol{Y} \middle| \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\right)$$

Probabilistic Classification and Graphical Models

Fixed structure

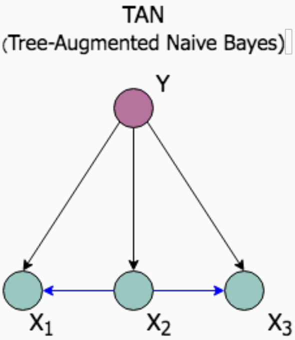


$\rightarrow$ We suppose that $\forall k \neq i,\ X_k \perp\!\!\!\perp X_i | Y$

$\rightarrow$ The MAP calculation becomes :

$$\hat{y} = \arg\max_{y} \left(P(y) \cdot \prod_{k=1}^{d} P(x^k | y) \right)$$

Simple Structure



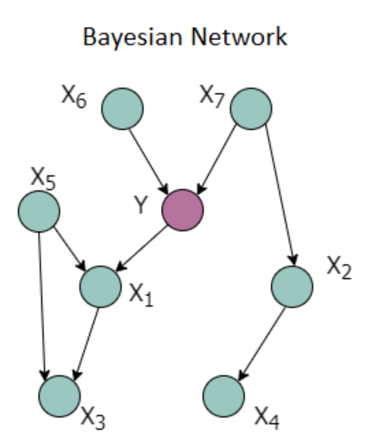
$\rightarrow$ Any X_i feature can have one other parent than Y

$\rightarrow$ Here the MAP becomes :

$$\hat{y} = \arg\max_y \left(P(y) \cdot \prod_{k=1}^{d} P(x^k | y, (x_{Parent^k})\right)$$

Probabilistic Classification and Graphical Models

Any structure



→ Y class is processed as any feature

→ Learning to find the structure and inference for the prediction

Classification of classification methods

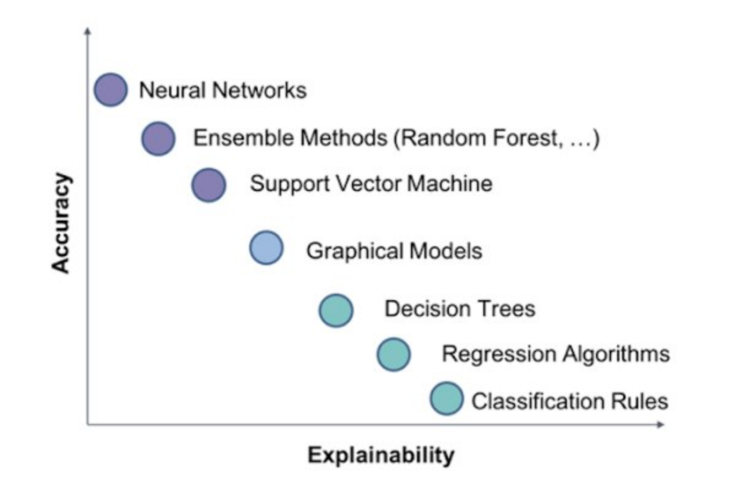


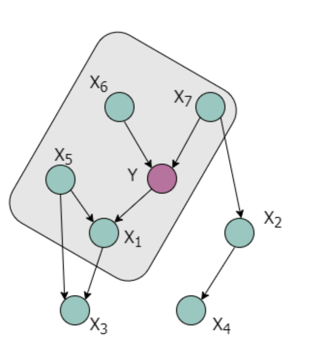
Figure 1 – Classification of Machine Learning models in function of predictive power and their explainability (Dam, Tran, and Ghose 2018)

Markov Blanket

Definition

A **Markov Blanket** of a random variable Y in a random variable set $\mathcal{S} = \{X_1, \dots, X_n\}$ is the minimal subset $\mathcal{S}_1$ of $\mathcal{S}$, conditioned on which other variables are independent with Y :

$$Y \perp\!\!\!\perp \mathcal{S} \backslash \mathcal{S}_1 \mid \mathcal{S}_1.$$



→ Markov Blanket of Y

→ contains its parents, its children and every others parents of its children

→ **Feature selection**

Implementation of Bayesian Network Classifiers

BNClassifier in pyAgrum

- `skbn` : Machine learning module of pyAgrum which builds a probabilistic classifier compatible with `scikit-learn`
  $\rightarrow$ `BNClassifier`

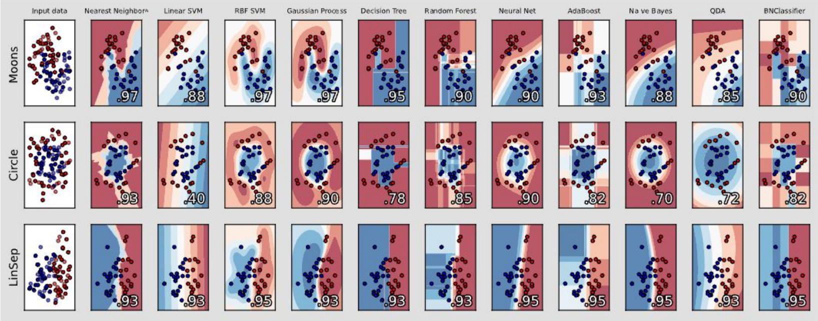
```python
# Create different classifiers.
classifiers = {
    'L1 logistic': LogisticRegression(C=C, penalty='l1',
                                      solver='saga',
                                      multi_class='multinomial',
                                      max_iter=10000),
    'L2 logistic (Multinomial)': LogisticRegression(C=C, penalty='l2',
                                                    solver='saga',
                                                    multi_class='multinomial',
                                                    max_iter=10000),
    'L2 logistic (OvR)': LogisticRegression(C=C, penalty='l2',
                                            solver='saga',
                                            multi_class='ovr',
                                           max_iter=10000),
    'Linear SVC': SVC(kernel='linear', C=C, probability=True,
                      random_state=0),
    'GPC': GaussianProcessClassifier(kernel),
    'BN' : BNClassifier(learningMethod='MIIC',
                        aPriori='Smoothing', aPrioriWeight=1,
                        discretizationNbBins=6,
                        discretizationStrategy="kmeans",
                        discretizationThreshold=10)
}
```

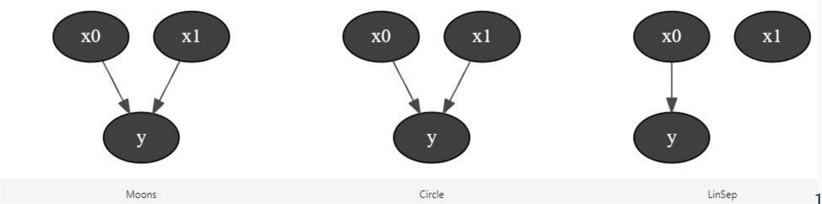
$\rightarrow$ `fit()`

$\rightarrow$ `predict()`

BNClassifier in pyAgrum

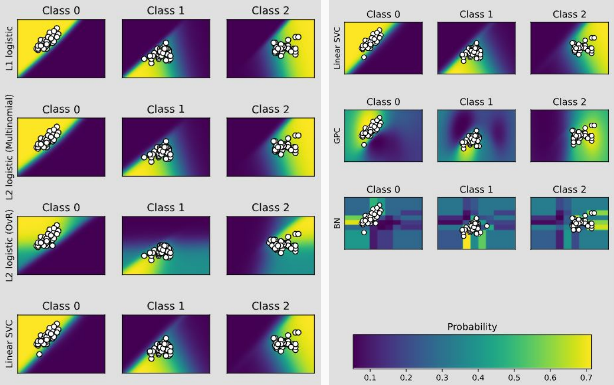
- 1st example : Comparing several binary classifiers





BNClassifier in pyAgrum

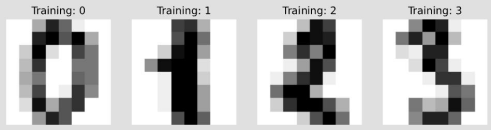
- 2nd example : Comparing several n-ary classifiers (on IRIS dataset)

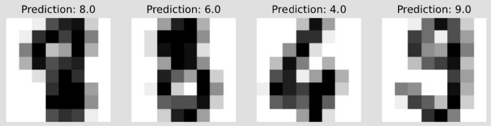


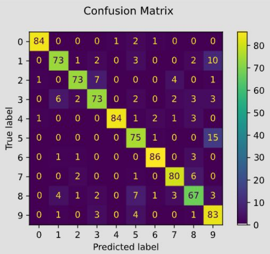

```
Accuracy (train) for L1 logistic: 82.7%
Accuracy (train) for L2 logistic (Multinomial): 82.7%
Accuracy (train) for L2 logistic (OvR): 79.3%
Accuracy (train) for Linear SVC: 82.0%
Accuracy (train) for GPC: 82.7%
Accuracy (train) for BN: 83.3%
```

BNClassifier in pyAgrum

- 3rd example : Recognizing hand-written digits



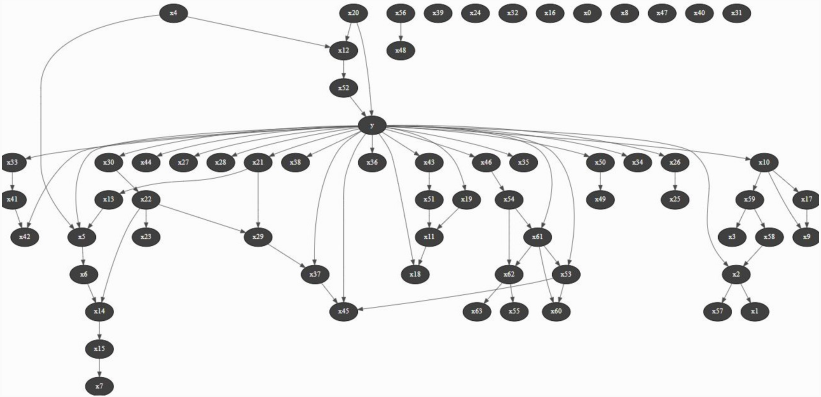




$\rightarrow$ Accuracy of 0.87
(0.97 for SVC classifier)

BNClassifier in pyAgrum

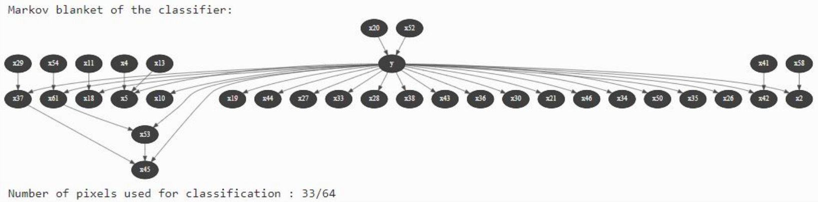
- 3rd example : Recognizing hand-written digits

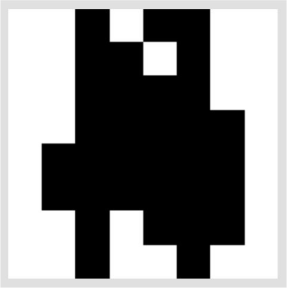


→ Bayesian Network

BNClassifier in pyAgrum

- 3rd example : 3rd example : Recognizing hand-written digits





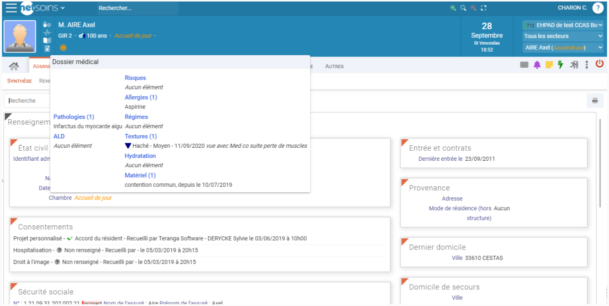
→ Visualization of the only pixels used in the classification

Examples from https://webia.lip6.fr/~phw//aGrUM/docs/last/notebooks/CompareClassifiersWithSklearn.ipynb.html

Context of the Application in Nursing Homes

NETSoins

- leading competitor in the healthcare software market for french nursing homes
- a file for each resident used and filled in by all types of staff : care assistants, nurses, doctors, animators, pharmacists, administrative and paramedical staff...
- Databases with 550,000 residents over 3 years on average



- First unfavorable health event we are trying to predict
- Skin lesion when the pressure is too high → immobility
- Prevalence of 7.2% in french nursing homes
- Demeaning, long, painful and costly to treat condition
- BUT highly avoidable with a specific and multidisciplinary approach

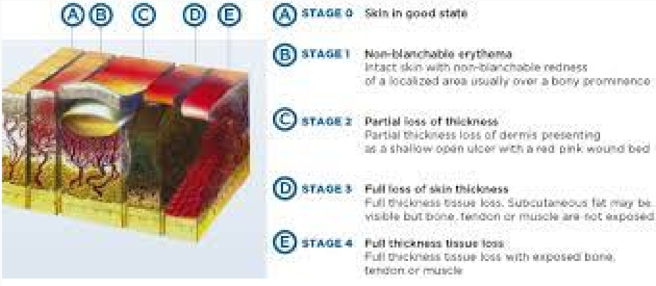


Figure 2 – The different stages of pressure ulcers (Belmin et al, 2016)

Actual medical methods

Example of Braden

Sensory Perception		Moisture		Activity		Mobility		Nutrition		Friction and Shear	
No Impairment	4	Rarely Moist	4	Walks Frequently	4	No Limitations	4	Excellent	4		
Slightly Limited	3	Occasionally Moist	3	Walks Occasionaly	3	Slightly Limited	3	Adequate	3	No Apparent Problem	3
Very Limited	2	Very Moist	2	Chair bound	2	Very Limited	2	Probably Inadequate	2	Potential Problem	2
Completely Limited	1	Constantly Moist	1	Bedbound	1	Completely Immobile	1	Very Poor	1	Problem	1

- Norton and Braden scales : risk detecting methods for pressure ulcers used in nursing homes
- Problem : simple to use but not very effective, therefore not widely used

How can we use NETSoins' data to improve pressure ulcers risk detection ?

Bayesian Network Classifier for Pressure Ulcers Classification

Methodology and Pretreatment

- Issues with the access to health data
- Transformation of an event database into a tabular database suitable for learning while keeping medical meaning
  $\rightarrow$ research and creation of 30 features
- Base separation : 75 % of residents for the training dataset and 25 % for the validation dataset
- 3 datasets with 3 different timeframes objectives for about 100 000 residents each
- Missing values completion with `KNNImputer` method
- Automatic discretization into 10 categories when necessary
- `BNClassifier` parameters :
  - Use of the `MIIC` algorithm

Results

	Predicted Class	
Actual Class		Class = 1	Class = 0
	Class = 1	True Positive	False Negative
	Class = 0	False Positive	True Negative

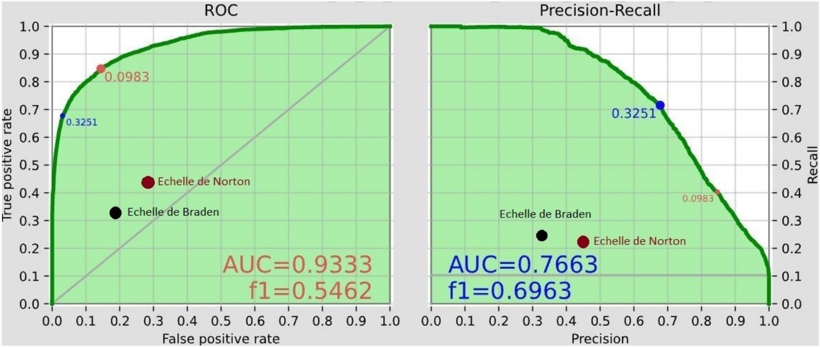
Definition (F-Score)

$$Fscore = 2 \times \frac{precision \times sensibility}{precision + sensibility}$$
$$= 2 \times \frac{\frac{TP}{TP+FP} \times \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}}$$

- $Sensibility = Recall$
- $Accuracy = \frac{TP+TN}{Total}$

Results

- ROC and Precision-Recall Curve :



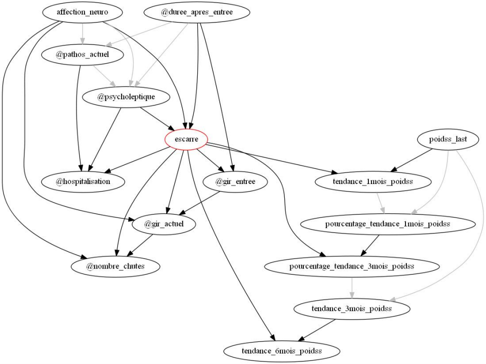
- Threshold choice

Results

- Summary of the different F-scores according to the methods and prediction timeframe :

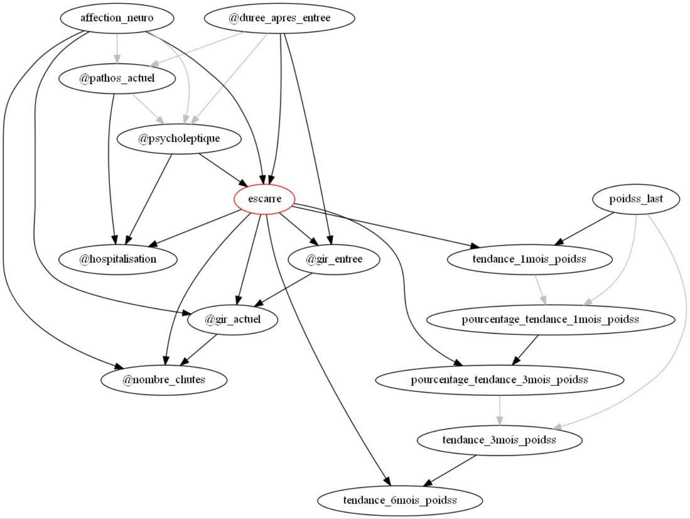
F-Score	**1-month**	**2-months**	**3-months**
BN Classifier	0,70	0,69	0,67
Random Forest	0,72	0,69	0,70
AdaBoost	0,69	0,67	0,69
Nearest Neighbors	0,55	0,55	0,56
Logistic Regression	0,32	0,36	0,42
Braden	0,32	-	-
Norton	0,29	-	-

- The obtained Bayesian Network and its Markov Blanket :



Graphical Results

- Markov Blanket :



Conclusion and Future Works

Conclusion and Future Works

- Classifier for pressure ulcers prediction efficient and relevant thanks to Bayesian Networks, better than the scores currently used in nursing homes
- Many possibilities of improvement :
  - integration of other features from NETSoins into aggregators
  - more complete exploitation of time series
  - better management of missing data
  - "expert" discretization
  - Improve the explainability (SHAP Values, ...)
- Application to other adverse health events
- Integration of the classifier in NETSoins, with alerts to bring high-risk situations to the attention of physicians and caregivers