

# Hazardous asteroids forecast via Markov random fields

Project for the exam: Probabilistic Modelling (DSE)

Marzio De Corato

December 11, 2021

# Outline

- 1 Intro
- 2 Celestial mechanics
- 3 Dataset
- 4 Prelim. analysis
- 5 Probabilistic modelling
- 6 Other ML algorithms
- 7 Conclusions
- 8 Supporting Info

# Introduction

- **Final Goal** Assessment of forecasts and interpretability for different machine learning algorithms, including the probabilistic models
- **Method** Use a dataset for which the laws that interconnect the different features are known from general principles
- **Dataset** CNEOS asteroids dataset for more than 3500 asteroids
- **Theoretical laws** Celestial mechanics
- **Algorithms involved - probabilistic models** GLASSO, mgm, minforest, mmod
- **Algorithms involved - others** Random forest, Support Vector Machines, Quadratic Discriminant Analysis, Logistic Regression

# Celestial mechanics

## Celestial mechanics [14]: equations of motion

Lets Consider the interaction between a planet of mass  $m_1$  at the position  $r_1$  (inertial frame) and an asteroid of mass  $m_2$  at the position  $r_2$

$$\mathbf{F}_1 = \mathcal{G} \cdot \frac{m_1 m_2}{r^3} \mathbf{r} = m_1 \ddot{\mathbf{r}}_1 \quad \mathbf{F}_2 = -\mathcal{G} \cdot \frac{m_1 m_2}{r^3} \mathbf{r} = m_1 \ddot{\mathbf{r}}_2 \quad (1)$$

If we consider the motion of the second item with respect to the first one

$$\ddot{\mathbf{r}} = \ddot{\mathbf{r}}_2 - \ddot{\mathbf{r}}_1 \quad \mu = \mathcal{G}(m_1 + m_2) \quad (2)$$

$$\frac{d^2 \mathbf{r}}{dt^2} + \mu \frac{\mathbf{r}}{r^3} = 0 \quad (3)$$

$\mathbf{r} \times \ddot{\mathbf{r}} = 0 \implies \mathbf{r}$  and  $\dot{\mathbf{r}}$  lies in the same plane

# Celestial mechanics [14]: equations of motion

With polar coordinates  $\hat{\mathbf{r}}$  and  $\hat{\boldsymbol{\theta}}$

$$\mathbf{r} = r\hat{\mathbf{r}} \quad (4)$$

$$\dot{\mathbf{r}} = \dot{r}\hat{\mathbf{r}} + r\dot{\theta}\hat{\boldsymbol{\theta}} \quad (5)$$

$$\ddot{\mathbf{r}} = \left(\ddot{r} - r\dot{\theta}^2\right)\hat{\mathbf{r}} + \left[\frac{1}{r}\frac{d}{dt}\left(r^2\dot{\theta}\right)\right]\hat{\boldsymbol{\theta}} \quad (6)$$

$$\mathbf{h} = r^2\dot{\theta}\hat{\mathbf{z}} \quad (7)$$

$$h = r^2\dot{\theta} \quad (8)$$

## Celestial mechanics [14]: 2<sup>th</sup> Kepler law

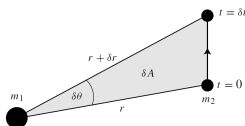


Figure 1: [14]

$$\delta A \approx \frac{1}{2}r(r + dr) \sin(\delta\theta) \approx \frac{1}{2}r^2\delta\theta \quad (9)$$

$$\frac{dA}{dt} = \frac{1}{2}r^2 \frac{d\theta}{dt} = \frac{1}{2}h \quad (10)$$

$h$  is constant  $\implies$  2<sup>th</sup> Kepler law

## Celestial mechanics [14]: 1<sup>th</sup> Kepler law

Using the substitution  $u = \frac{1}{r}$   $h = r^2 \dot{\theta}$

$$\dot{r} = -\frac{1}{u} \frac{du}{d\theta} \dot{\theta} = -h \frac{du}{d\theta} \quad (11)$$

$$\ddot{r} = -h \frac{d^2 u}{d\theta^2} \dot{\theta} = -h^2 u^2 \frac{d^2 u}{d\theta^2} \quad (12)$$

$$\frac{d^2 u}{d\theta^2} + u = \frac{\mu}{h^2} \quad (13)$$

$$u = \frac{\mu}{h^2} [1 + e \cos(\theta - \phi)] \quad (14)$$



# Celestial mechanics [14]: 1<sup>th</sup> Kepler law

$$r = \frac{p}{1 + e \cos(\theta - \phi)} \quad (15)$$

$e$  is **eccentricity**

- circle:  $e = 0$      $p = a$
- ellipse:  $0 < e < 1$   
 $p = a(1 - e^2)$
- parabola:  $e = 1$      $p = 2q$
- hyperbola:  $e > 1$   
 $p = a(e^2 - 1)$

$a$  is the **semi-major axis** of the conic

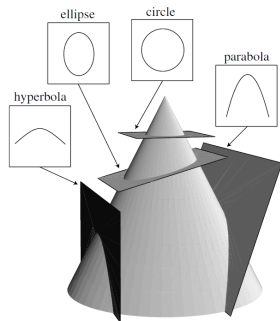


Figure 2: [14]

## Celestial mechanics [14]: 3<sup>th</sup> Kepler law

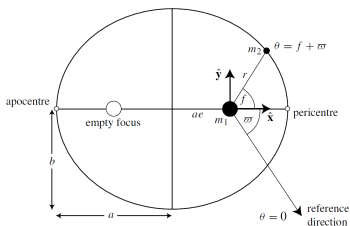


Figure 3: [14]

$$b^2 = a^2(1 - e^2) \quad (16)$$

$$r = \frac{a(1 - e^2)}{1 + e \cdot \cos(\theta - \phi)} \quad (17)$$

Area swept in one **orbital period**  $T$

$$A = \pi ab$$

We know that:  $hT/2 \quad h^2 = \mu a(1 - e^2)$

Therefore

$$T^2 = \frac{4\pi^2}{\mu} a^3 \quad (18)$$

## Celestial mechanics [14]: 3<sup>th</sup> Kepler law

$$\frac{m_c + m}{m_c + m'} = \left(\frac{a}{a'}\right)^3 \left(\frac{T'}{T}\right)^2 \quad (19)$$

But since  $m, m' \ll m_c$

$$\left(\frac{a}{a'}\right)^3 \approx \left(\frac{T'}{T}\right)^2 \quad (20)$$

And therefore

$$T' \approx a'^{3/2} \quad (21)$$

**Remark:** The mass of the asteroid is **not** involved

# Celestial mechanics [14]: Orbital parameters

Mean motion  $n = \frac{2\pi}{T}$

$$v_{perihelion} = na\sqrt{\frac{1+e}{1-e}} \quad (22)$$

$$v_{aphelion} = na\sqrt{\frac{1-e}{1+e}} \quad (23)$$

**Remark:** The mean motion of an asteroid is different with respect to the the asteroid relative velocity (measured from Earth), since the latter is different at the perihelion an at the aphelion

# Celestial mechanics [14]: Orbital parameters

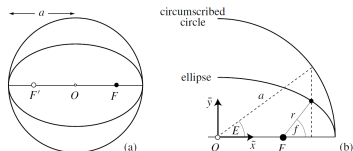


Figure 4: [14]

## Mean anomaly

$$M = n(t - \tau) \quad (24)$$

- $M = f = 0 \quad t = \tau$  Perihelion
- $M = f = \pi \quad t = \tau + T/2$  Aphelion

$$M = E - e \sin E \quad (25)$$

## Jupiter Tisserard invariant

$$T_P = \frac{a_p}{a} + 2 \cos I \sqrt{\frac{a}{a_p} (1 - e^2)} \quad (26)$$

# Celestial mechanics [14]: Orbital parameters

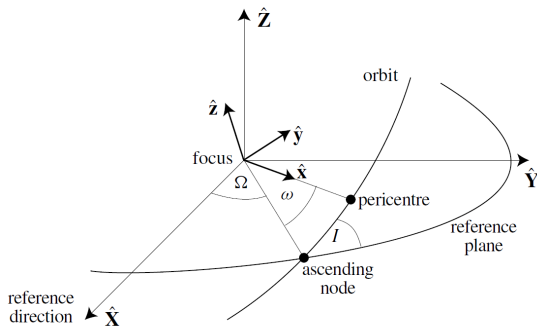


Figure 5: [14]

$I$ : inclination of the orbit

$\Omega$ : longitude of the ascending node

## Celestial mechanics [14]: Magnitude

$$\Phi = \frac{L}{4\pi r^2} \quad (27)$$

$$m = -2.5 \log_{10} \Phi + C \quad (28)$$

$$m_1 - m_2 = -2.5 \log_{10} \frac{\Phi_1}{\Phi_2} \quad (29)$$

$$M - m = -2.5 \log_{10} \frac{\Phi \cdot d^2}{\Phi \cdot 10^2} \quad (30)$$

$$M = m + 5 - 5 \log_{10} d \quad (31)$$

Where  $\Phi$  is the flux for a sphere of radius  $r$ ,  $m$  the relative magnitude and  $M$  the **Absolute magnitude**

## Celestial mechanics [14]: Magnitude

$$\Phi = \frac{L}{4\pi r^2} \quad (32)$$

$$m = -2.5 \log_{10} \Phi + C \quad (33)$$

$$m_1 - m_2 = -2.5 \log_{10} \frac{\Phi_1}{\Phi_2} \quad (34)$$

$$M - m = -2.5 \log_{10} \frac{\Phi \cdot d^2}{\Phi \cdot 10^2} \quad (35)$$

$$M = m + 5 - 5 \log_{10} d \quad (36)$$

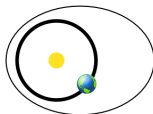
Where  $\Phi$  is the flux for a sphere of radius  $r$ ,  $m$  the relative magnitude and  $M$  the **Absolute magnitude**



# Celestial mechanics [1]: Classification

## Amors

Earth-approaching NEAs with orbits exterior to Earth's but interior to Mars' (named after asteroid (1221) Amor)



$$a > 1.0 \text{ AU} \\ 1.017 \text{ AU} < q < 1.3 \text{ AU}$$

## Apollos

**Earth-crossing** NEAs with semi-major axes larger than Earth's (named after asteroid (1862) Apollo)



$$a > 1.0 \text{ AU} \\ q < 1.017 \text{ AU}$$

## Atens

**Earth-crossing** NEAs with semi-major axes smaller than Earth's (named after asteroid (2062) Aten)



$$a < 1.0 \text{ AU} \\ Q > 0.983 \text{ AU}$$

## Atiras

NEAs whose orbits are contained entirely within the orbit of the Earth (named after asteroid (163693) Atira)



$$a < 1.0 \text{ AU} \\ Q < 0.983 \text{ AU}$$

( $q$  = perihelion distance,  $Q$  = aphelion distance,  $a$  = semi-major axis)

- **Potentially Hazardous Asteroids:**  $\text{MOID} \leq 0.05 \text{ au}$   $M \leq 22.0$   
*NEAs whose Minimum Orbit Intersection Distance (MOID) with the Earth is 0.05 au or less and whose absolute magnitude (M) is 22.0 or brighter*

# Dataset

- The asteroid dataset was retrieved from Kaggle [2], which reports into a more machine readable form the dataset of The Center for Near-Earth Object Studies (CNEOS) [3], a NASA research centre.
- 3552 Asteroids
- Among the 40 the features, the ones connected only to the other name of the asteroid, or connected only to the name of the orbit and the one connected with the orbiting planet ( since for all it was the Earth) were discarded
- The proportion hazardous/not hazardous was set 1:5
- The continuous measures were standardised and demeaned

# Features

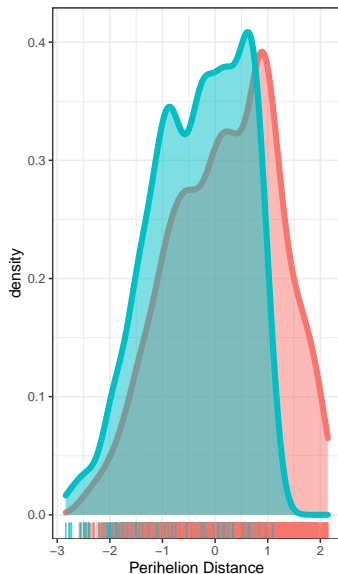
Features	Type
Neo Reference ID	not used
Absolute Magnitude	Continuous
Est Dia in KM (min)	Continuous
Est Dia in KM (max)	Continuous
Close Approach Date	Continuous
Epoch Date Close Approach	Continuous
Relative_Velocity	Continuous
Miss_Dist	Continuous
Min_Orbit_Intersection	Continuous
Jupiter_Tisserand_Invariant	Continuous
Epoch_Osculation	Continuous
Eccentricity	Continuous

# Features

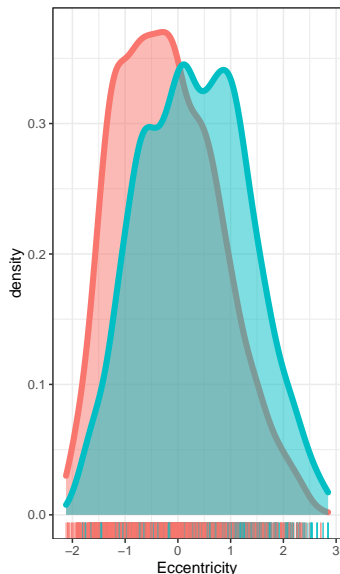
Features	Type
Semi Major Axis	Continuous
Inclination	Continuous
Asc Node Longitude	Continuous
Orbital Period	Continuous
Perihelion Distance	Continuous
Perihelion Arg	Continuous
Perihelion Time	Continuous
Mean_Anomaly	Continuous
Mean_Motion	Continuous
Hazardous	Categorical (Binary)

# Prelim. analysis

# Density Plot



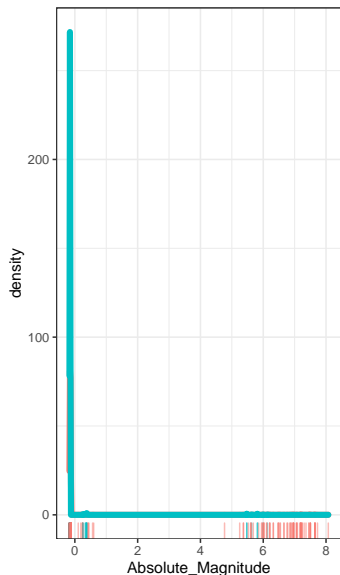
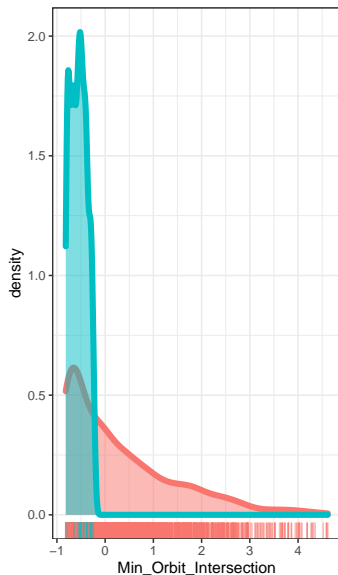
Hazardous  
FALSE  
TRUE



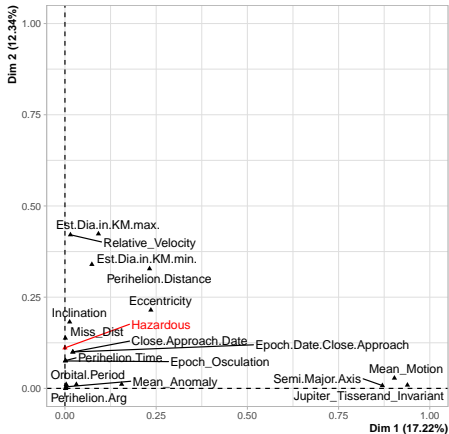
Hazardous  
FALSE  
TRUE



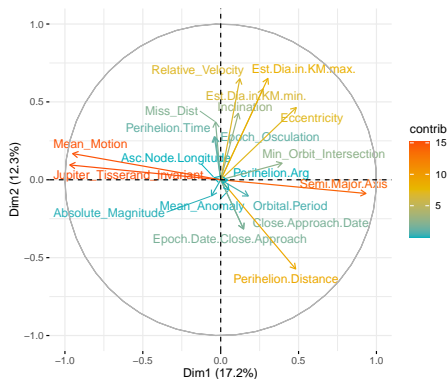
# Density Plot



Graph of the variables

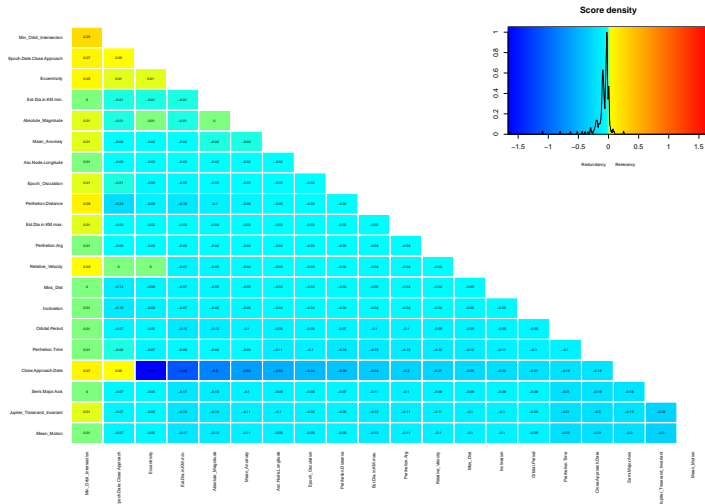


Quantitative variables – FAMD



Performed with the FactoMineR package [12]

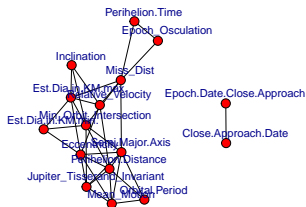
# Mutual information analysis



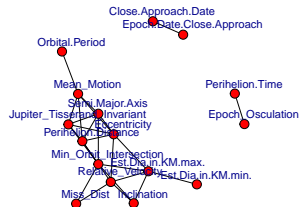
Performed with the varrank package [11]

# Probabilistic modelling

# GLASSO



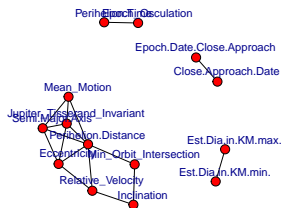
$\rho=0.1$



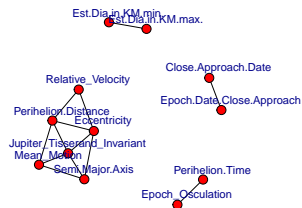
$\rho=0.2$

Performed with the GLASSO package [4]

# GLASSO



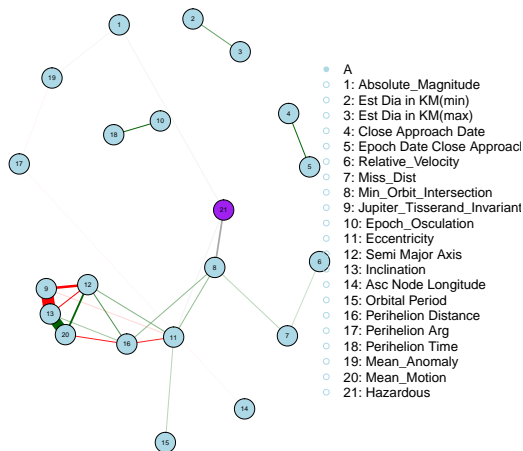
$$\rho=0.3$$



$$\rho=0.4$$

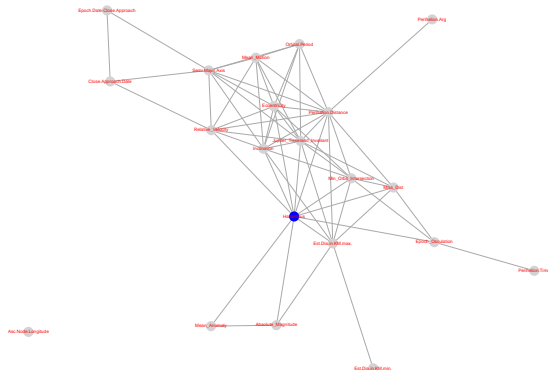
Performed with the GLASSO package [4]

# Mixed interactions: mgm



Performed with the mgm package [9]

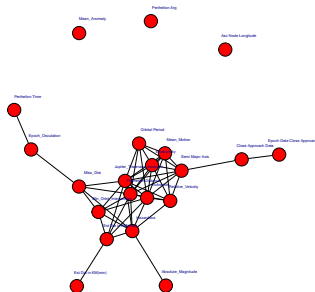
## Mixed interactions: minforest



Performed with the gRapHD package [7]



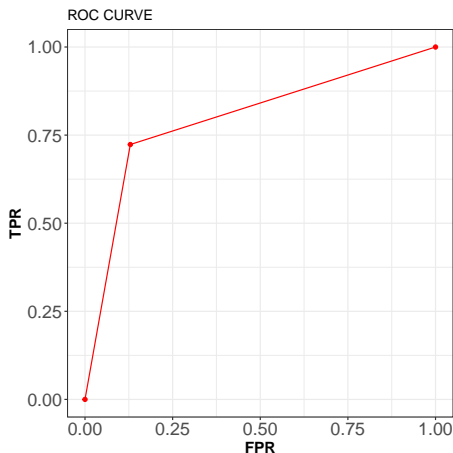
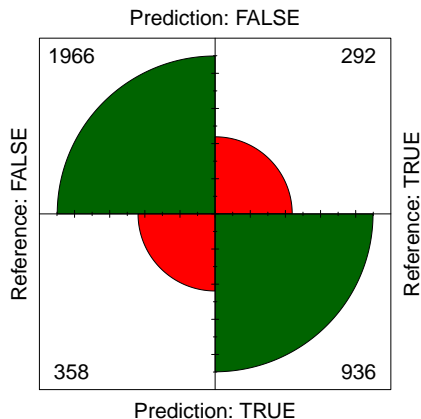
## Mixed interactions: mmod



Performed with the gRim package [10]

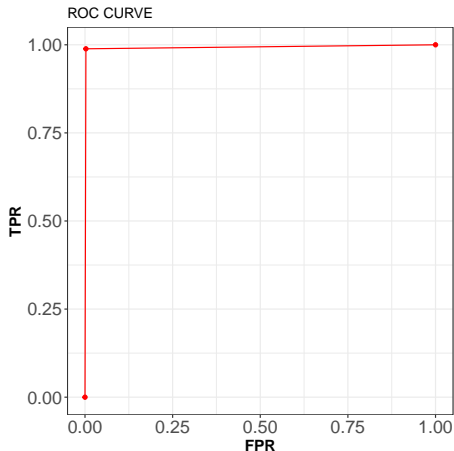
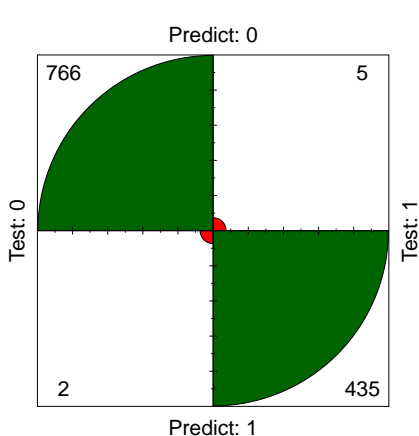
# Mixed interactions

The mgm model is the one that has the list of connection more coherent with the celestial mechanics laws.



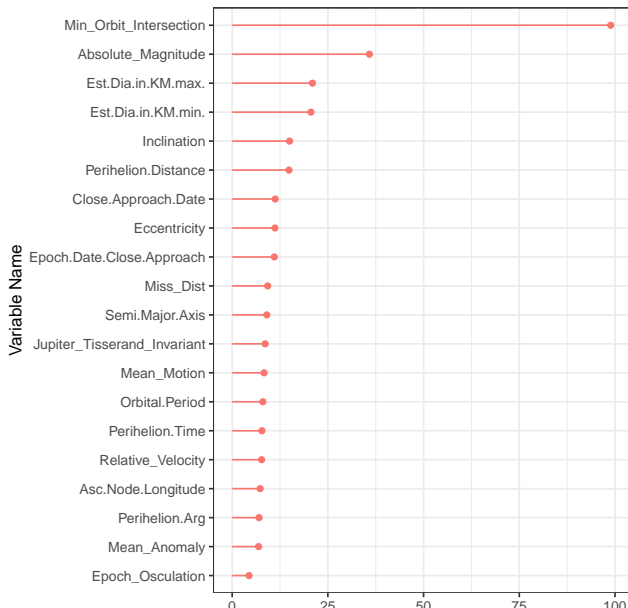
# Other ML algorithms

# Random Forest

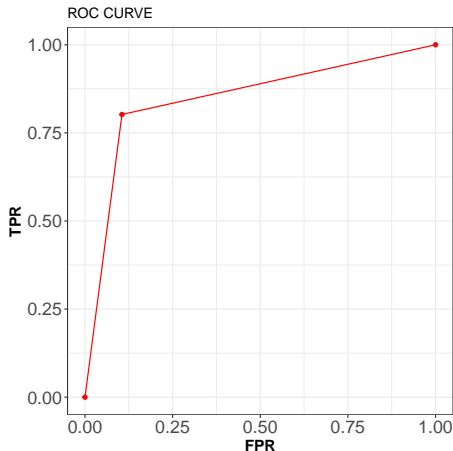
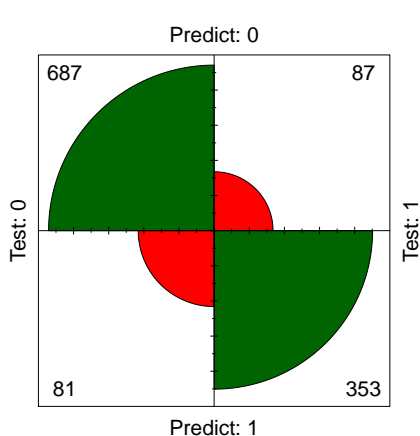


Performed with the rfor package [13]

# Random Forest

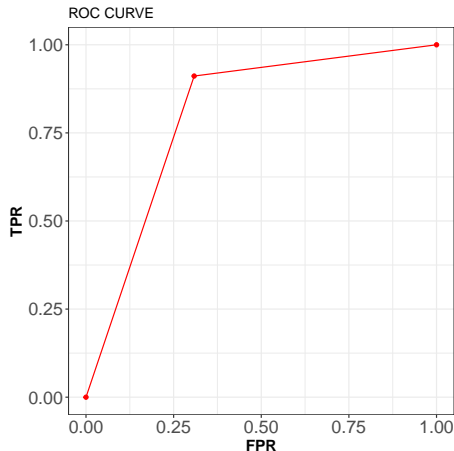
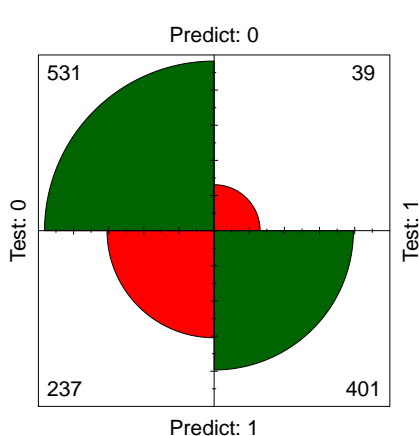


# Support Vector Machines



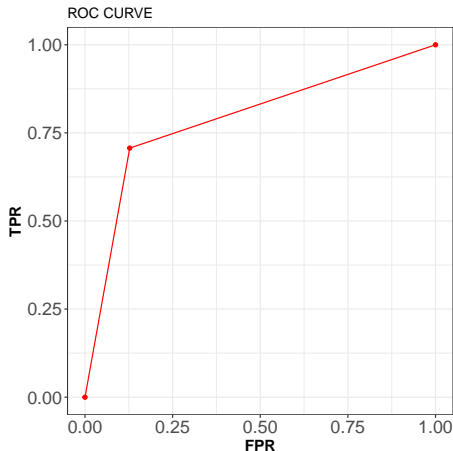
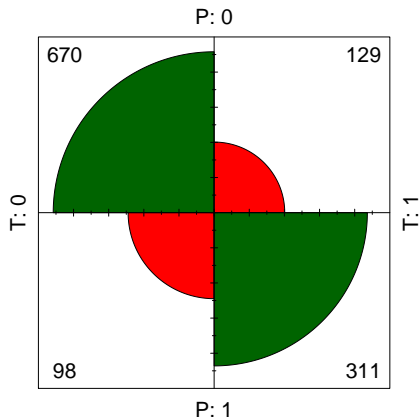
Performed with the e1071 package [8]

# Quadratic Discriminant Analysis (QDA)



Performed with the MASS package [17]

# Logistic regression



Performed with the stats package [15]



Table 1:  $\phi$  coefficient (also known as Matthews correlation coefficient )

Algorithm	$\phi$
RF	0.9876
SVM	0.7111
logistic	0.6173
mgm	0.5997
QDA	0.5562

# Conclusions

## Remark (Interpretability - Tarski definition)

*The formal theory  $T$  can be translated into  $S$  if and only if  $S$  can prove the theorem of  $T$  in its language [16]*

## Remark (Scientific method - Einstein definition)

*Science uses the totality of the primary concepts, i.e., concepts directly connected with sense experiences, and propositions connecting them. Such a state of affairs cannot, however, satisfy a spirit which is really scientifically minded; because the totality of concepts and relations obtained in this manner is utterly lacking in logical unity. In order to supplement this deficiency, one invents a system poorer in concepts and relations, a system retaining the primary concepts and relations of the first layer as logically derived concepts and relations. This new secondary system pays for its higher logical unity by having elementary concepts (concepts of the second layer), which are no longer directly connected with complexes of sense experiences [5]*

## Remark (Scientific method - Einstein definition )

*The essential thing is the aim to represent the multitude of concepts and propositions, close to experience, as propositions, logically deduced from a basis, as narrow as possible, of fundamental concepts and fundamental relations which themselves can be chosen freely (axioms) [5]*

## Conclusions: forecast performances vs interpretability

- The mgm algorithm is not the best one in term of performances, but it provides the connections between the features. On the other side, except for the variable importance in RF, the other are black box one
- The mgm model, as the other graphical model is open to a true scientific validation, the other not.
- The probabilistic models lack in the forecast is definitely compensated by their interpretability
- This is meaningful since this two features are in conflict
- The probabilistic models provide a good trade-off between interpretability and forecast performances, as long as one is interest to produce a really scientific result (e.g if the only aim is the forecast the RF is definitely better. However how long one can trust to the RF result ?)

*In its efforts to learn as much as possible about nature, modern physics has found that certain things can never be “known” with certainty. Much of our knowledge must always remain uncertain. The most we can know is in terms of probabilities. Richard P. Feynman (1918-1988)*

# Bibliography I

- [1] [https://cneos.jpl.nasa.gov/about/neo\\_groups.html](https://cneos.jpl.nasa.gov/about/neo_groups.html).
- [2] <https://www.kaggle.com/shrutimehta/nasa-asteroids-classification>.
- [3] <https://cneos.jpl.nasa.gov/>.
- [4] <https://cran.r-project.org/web/packages/glasso/glasso.pdf>.
- [5] <https://www.amacad.org/publication/physics-reality>.
- [6] <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>.
- [7] Gabriel CG de Abreu, Rodrigo Labouriau, and David Edwards.  
“High-dimensional graphical model search with graphd R package”.  
In: *arXiv preprint arXiv:0909.1234* (2009).



# Bibliography II

- [8] Evgenia Dimitriadou et al. “Misc functions of the Department of Statistics (e1071), TU Wien”. In: *R package 1* (2008), pp. 5–24.
- [9] Jonas Haslbeck and Lourens J Waldorp. “mgm: Estimating time-varying mixed graphical models in high-dimensional data”. In: *arXiv preprint arXiv:1510.06871* (2015).
- [10] Søren Højsgaard, David Edwards, and Steffen Lauritzen. *Graphical Models with R*. ISBN 978-1-4614-2298-3. New York: Springer, 2012. DOI: 10.1007/978-1-4614-2299-0.
- [11] Gilles Kratzer and Reinhard Furrer. “varrank: an R package for variable ranking based on mutual information with applications to observed systemic datasets”. In: *arXiv preprint arXiv:1804.07134* (2018).

# Bibliography III

- [12] Sébastien Lê, Julie Josse, and François Husson. “FactoMineR: an R package for multivariate analysis”. In: *Journal of statistical software* 25.1 (2008), pp. 1–18.
- [13] Andy Liaw and Matthew Wiener. “Classification and Regression by randomForest”. In: *R News* 2.3 (2002), pp. 18–22. URL: <https://CRAN.R-project.org/doc/Rnews/>.
- [14] Carl D Murray and Stanley F Dermott. *Solar system dynamics*. Cambridge university press, 1999.
- [15] R Core Team. *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria, 2013. URL: <http://www.R-project.org/>.
- [16] Alfred Tarski, Andrzej Mostowski, and Raphael Mitchel Robinson. *Undecidable theories*. Vol. 13. Elsevier, 1953.

## Bibliography IV

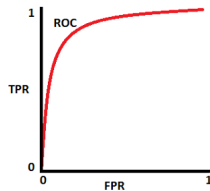
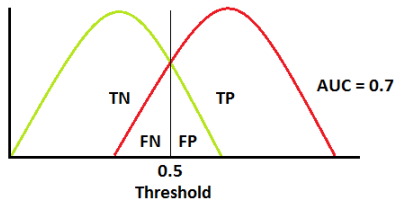
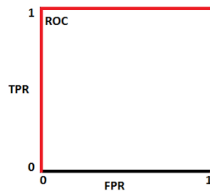
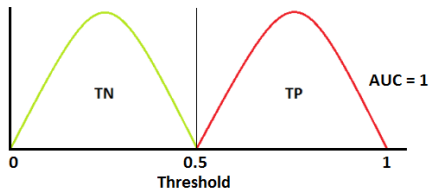
- [17] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Fourth. ISBN 0-387-95457-0. New York: Springer, 2002. URL: <https://www.stats.ox.ac.uk/pub/MASS4/>.

# ROC and $\phi$ factor

	Actual - N	Actual - P
Predicted - N	#TP	#FN
Predicted - P	#FP	#TP

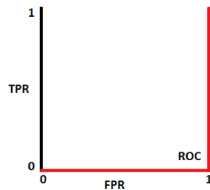
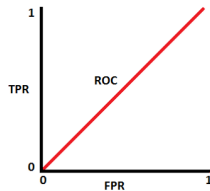
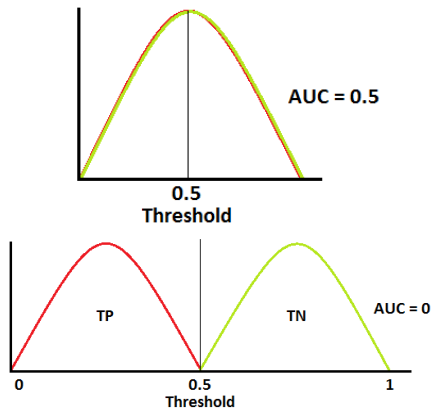
$$\phi = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

# Receiver operating characteristic



Images taken from [6]

# Receiver operating characteristic



Images taken from [6]

# FAMD



# Factor analysis of mixed data - FAMD

$r(z,k)$  correlation coefficient (z and k quantitative)

$\eta^2(z,q)$  correlation ratio (z quantitative and q qualitative)

$$\text{PCA} \rightarrow \max \sum_k r^2(z, k)$$

$$\text{MCA} \rightarrow \max \sum_q \eta^2(z, q)$$

$$\text{FAMD} \rightarrow \max \sum_k r^2(z, k) + \max \sum_q \eta^2(z, q)$$

# Graphical models

$$p(x_1, x_2, \dots, x_n) \quad (37)$$

$$p(x_{1:V}) = p(x_1)p(x_2|x_1)p(x_3|x_2, x_1)\dots p(x_V|x_{1:V-1}) \quad (38)$$

$$X \perp Y|Z \iff p(X, Y|Z) = p(X|Z)p(Y|Z) \quad (39)$$

$$p(\mathbf{x}_{1:V}) = p(x_1) \prod_{t=1}^V p(x_t|x_{t-1}) \quad (40)$$

## Theorem (Hammersley-Clifford)

*A positive distribution  $p(\mathbf{y}) > 0$  satisfies the CI properties of an indirect graph  $G$  iff  $p$  can be represented as a product of factor, one per maximal clique, i.e.*

$$p(\mathbf{y}|\theta) = \frac{1}{Z(\theta)} \prod_{c \in C} \psi_c(\mathbf{y}_c|\theta_c) \quad (41)$$

*where  $C$  is the set of all the (maximal) cliques of  $G$ , and  $Z(\theta)$  is the partition function given by*

$$Z(\theta) := \sum_{\mathbf{y}} \prod_{c \in C} \psi_c(\mathbf{y}_c|\theta_c) \quad (42)$$

*Note that this partition function is what ensures the overall distribution sums to 1*

$$p(y|\theta) = \frac{1}{Z(\theta)} \exp \left( - \sum_c E(y_c|\theta_c) \right) \quad (43)$$

$$\psi_c(y_c|\theta_c) = \exp(-E(y_c|\theta_c)) \quad (44)$$

# Parameters tuning

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_c} = \frac{1}{N} \sum_i \left[ \phi_c(y_i) - \frac{\partial}{\partial \boldsymbol{\theta}_c} \log Z(\boldsymbol{\theta}) \right] \quad (45)$$

$$\frac{\partial \log Z(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbb{E}[\phi_c(\mathbf{y}) | \boldsymbol{\theta}] = \sum_{\mathbf{y}} \phi_c(\mathbf{y}) p(\mathbf{y} | \boldsymbol{\theta}) \quad (46)$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_c} = \left[ \frac{1}{N} \sum_i \phi_c(y_i) \right] - \mathbb{E}[\phi_c(\mathbf{y})] \quad (47)$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_c} = \mathbb{E}_{p_{emp}}[\phi_c(\mathbf{y})] - \mathbb{E}_{p_{(\cdot | \boldsymbol{\theta})}}[\phi_c(\mathbf{y})] \quad (48)$$

$$\mathbb{E}_{p_{emp}}[\phi_c(\mathbf{y})] = \mathbb{E}_{p_{(\cdot | \boldsymbol{\theta})}}[\phi_c(\mathbf{y})] \quad (49)$$

$$p(\mathbf{y}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp \left( \sum_c \boldsymbol{\theta}_c^T \phi_c(\mathbf{y}) \right) \quad (50)$$

$$\mathcal{L}(\boldsymbol{\theta}) := \frac{1}{N} \sum_i \log p(\mathbf{y}_i|\boldsymbol{\theta}) = \frac{1}{N} \sum_i \left[ \sum_c \boldsymbol{\theta}_c^T \phi_c(y_i) - \log Z(\boldsymbol{\theta}) \right] \quad (51)$$

$$L_{pen}(K, \hat{\mu}) = \log \det(K) - \text{tr}(K S) - \rho \|K\| \quad (52)$$

$$K = \Sigma^{-1}$$

S: empirical covariance matrix



$$P(X_s|X_{\setminus s}) = \exp \left\{ E_s(X_{\setminus s})\phi_s(X_s) + B_s(X_s) - \Phi(X_{\setminus s}) \right\} \quad (53)$$

$\phi_s$  function of sufficient statistics  $B_s$  base measure

$$\begin{aligned} P(X) = \exp & \left( \sum_{s \in V} \theta_s \phi_s(X_s) + \sum_{s \in V} \sum_{r \in N(s)} \theta_{s,r} \phi_s(X_s) \phi_r(X_r) \right. \\ & \left. + \dots + \sum_{r_1, \dots, r_k \in C} \theta_{r_1, \dots, r_k} \prod_{j=1}^k \phi_{r_j}(X_{r_j}) + \sum_{s \in V} B_s(X_s) - \Phi(\theta) \right) \end{aligned} \quad (54)$$

$$\hat{\theta} = \arg \min_{\theta} \{ -\mathcal{L}(\theta, X) + \lambda \|\theta\|_1 \} \quad \|\theta\|_1 = \sum_{j=1}^J |\theta_j| \quad (55)$$

## mmod - Homogeneous Mixed Interaction (HMI) models

$$f(\mathbf{i}, \mathbf{y}) = p(\mathbf{i})(2\pi)^{-q/2} \det(\Sigma)^{-1/2} \exp \left[ -\frac{1}{2} (\mathbf{y} - \mu(\mathbf{i}))^T \Sigma^{-1} (\mathbf{y} - \mu(\mathbf{i})) \right] \quad (56)$$

$$f(\mathbf{i}, \mathbf{y}) = \exp \left[ g(\mathbf{i}) + h(\mathbf{i})^T \mathbf{y} - \frac{1}{2} \mathbf{y}^T K \mathbf{y} \right] \quad (57)$$

where  $K$  is the concentration matrix,  $g(i)$  and  $h(i)$  are the log-linear expansion of the probability  $p(\mathbf{i})$  (canonical parameters)

Decomposable model

$$\hat{f}(x) = \prod_{j=1}^k \frac{\hat{f}_{C_j}(x_{C_j})}{\hat{f}_{S_j}(x_{S_j})} \quad (58)$$

Likelihood

$$\hat{\mathcal{L}}_s = \sum_i n(i) \log \left( \frac{n(i)}{N} \right) - Nq \frac{\log(2\pi i)}{2} - N \frac{\log(\det(S))}{2} - N \frac{q}{2} \quad (59)$$

# Information theory

$$H(X) = - \sum_{x \in X} p(x) \log p(x) \quad (60)$$

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y) \quad (61)$$

$$\begin{aligned} H(X|Y) &= \sum_{x \in \mathcal{X}} p(x) H(Y|X = x) \\ &= - \sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \\ &= - E \log p(Y|X) \end{aligned} \quad (62)$$

$$H(X, Y) = H(X) + H(Y|X) \quad (63)$$

$$D(p||q) = \sum p(x) \log \frac{p(x)}{q(x)} \quad D(p||q) \geq 0 \quad (64)$$

$$\begin{aligned} I(X; Y) &= \sum_{(x, y)} \log \frac{p(x, y)}{p(x)p(y)} = D(p(x, y) || p(x)p(y)) \\ &= H(X) - H(X|Y) = H(Y) - H(Y|X) \end{aligned} \quad (65)$$

$$I(X_1, X_2, \dots, X_n; Y) = \sum_{i=1}^n I(X_i; Y | X_{i-1}, X_{i-2}, \dots, X_1) \quad (66)$$

$$g(\alpha, \mathbf{C}, \mathbf{S}, f_i) = MI(f_i; \mathbf{C}) - \sum_{f_s \in S} \alpha(f_i, f_s, \mathbf{C}, \mathbf{S}) MI(f_i; f_s) \quad (67)$$