Bayesian Model Selection

Complexity of spin models

in collaboration with

Alberto Beretta, Claudia Battistin and Matteo Marsili

- Data with random errors: find the model that best captures the patterns hidden within the data

- Model:

  — not too simple: we want to be able to fit the data

  — not too complex: to capture the main patterns

  Simple models are preferred, unless the data calls for a more complex one.

- How can we characterise the complexity of a model?

Spin Models

- Consider a system of $\boxed{n \text{ spins}}$ that takes random values in $\{-1, +1\}$

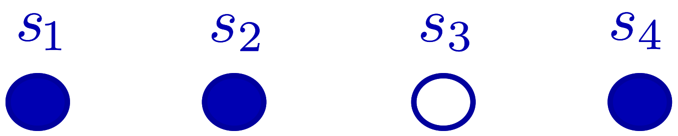
Ex.

$n = 4$

Spin Models

- Consider a system of $\boxed{n \text{ spins}}$ that takes random values in $\{-1, +1\}$

Ex.



$$\vec{s}^{(1)} = \{1, 1, -1, 1\}$$

$$n = 4$$

Spin Models

- Consider a system of $\boxed{n \text{ spins}}$ that takes random values in $\{-1,+1\}$

Ex.

$$n = 4$$

s_1	s_2	s_3	s_4
●	●	○	●
●	○	●	○
○	●	●	●
	$\cdots$		
●	●	○	○

$$\left.\begin{aligned} \vec{s}^{(1)} &= \{1, 1, -1, 1\} \\ \vec{s}^{(2)} &= \{1, -1, 1, -1\} \\ &\cdots \\ \vec{s}^{(N)} &= \{1, 1, -1, -1\} \end{aligned}\right\}$$

Data set:

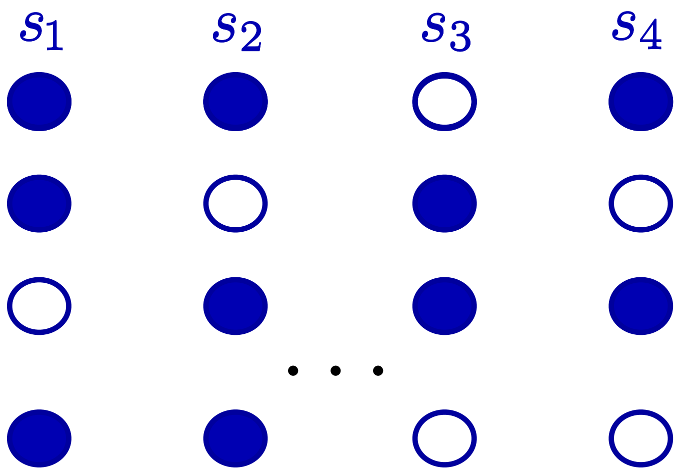
$\boxed{N \text{ samples} \quad \hat{s} = \{\vec{s}^{(i)}\}}$

Spin Models

- Consider a system of $\boxed{n \text{ spins}}$ that takes random values in $\{-1, +1\}$

Ex.

$n = 4$



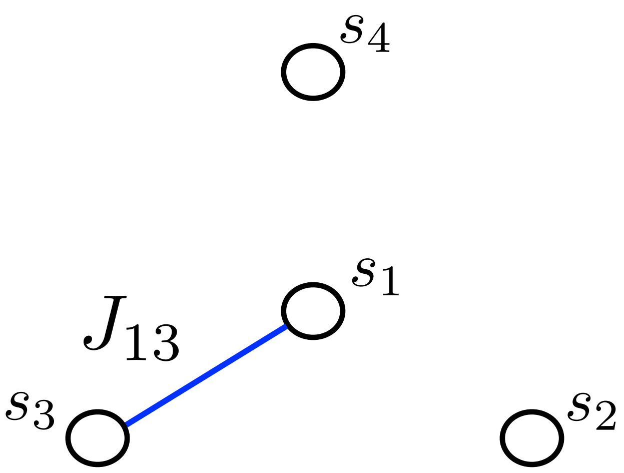
$$\left.\begin{array}{l} \vec{s}^{\,(1)} = \{1, 1, -1, 1\} \\ \vec{s}^{\,(2)} = \{1, -1, 1, -1\} \\ \quad \dots \\ \vec{s}^{\,(N)} = \{1, 1, -1, -1\} \end{array}\right\}$$

Data set:

N samples

$\hat{s} = \{\vec{s}^{\,(i)}\}$

- What would be the best model for the system, that could explain what we observe/reproduce similar data?



Parameters:

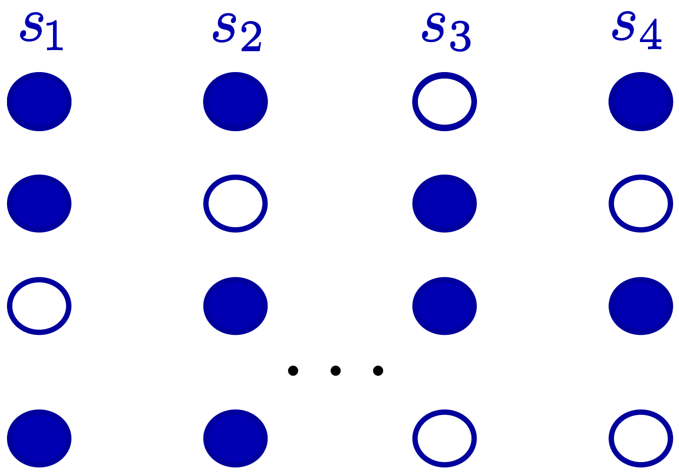
$\{J_{13}\}$

Spin Models

- Consider a system of n spins that takes random values in $\{-1, +1\}$

Ex.

$n = 4$



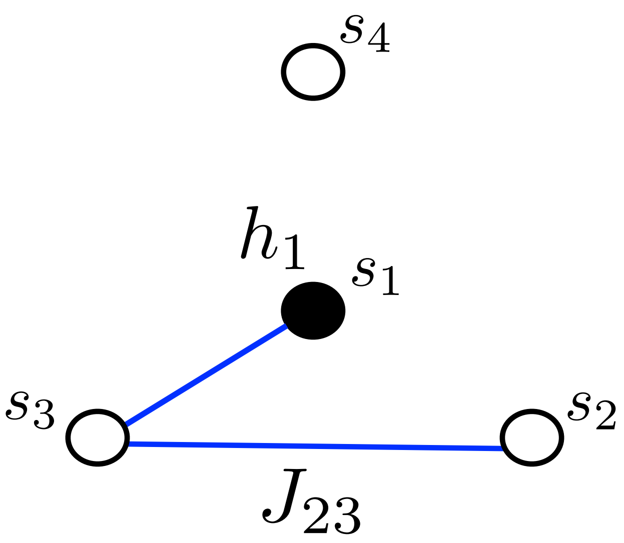
$$\vec{s}^{\,(1)} = \{1, 1, -1, 1\}$$
$$\vec{s}^{\,(2)} = \{1, -1, 1, -1\}$$
$$\dots$$
$$\vec{s}^{\,(N)} = \{1, 1, -1, -1\}$$

Data set:

N samples

$\hat{s} = \{\vec{s}^{\,(i)}\}$

- What would be the best model for the system, that could explain what we observe/reproduce similar data?



Parameters:

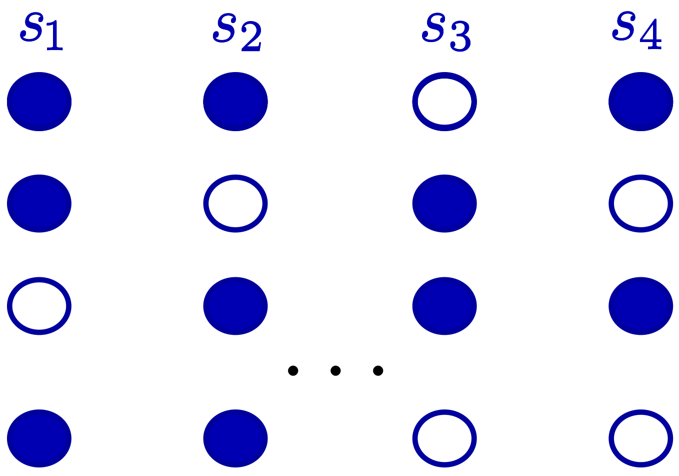
$$\vec{\theta} = \{h_1, J_{13}, J_{23}\}$$

Spin Models

- Consider a system of $\boxed{n \text{ spins}}$ that takes random values in $\{-1, +1\}$

Ex.

$n = 4$



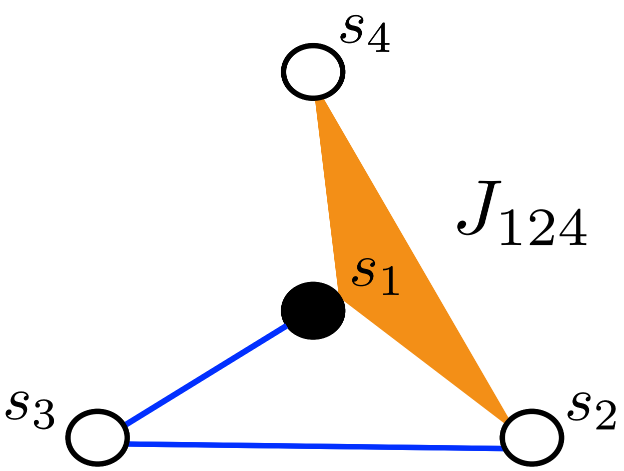
$$\left.\begin{aligned} \vec{s}^{(1)} &= \{1, 1, -1, 1\} \\ \vec{s}^{(2)} &= \{1, -1, 1, -1\} \\ &\dots \\ \vec{s}^{(N)} &= \{1, 1, -1, -1\} \end{aligned}\right\}$$

Data set:

N samples

$\hat{s} = \{\vec{s}^{(i)}\}$

- What would be the best model for the system, that could explain what we observe/reproduce similar data?



Parameters:

$$\vec{\theta} = \{h_1, J_{13}, J_{23}, J_{124}\}$$

Spin Models

- Consider a system of $\boxed{n \text{ spins}}$ that takes random values in $\{-1,+1\}$

Ex.

$n = 4$

s_1	s_2	s_3	s_4
●	●	○	●
●	○	●	○
○	●	●	●
	$\dots$		
●	●	○	○

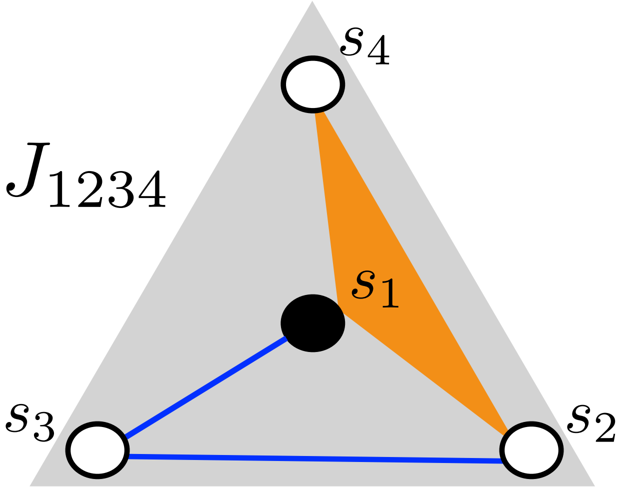
$$\left.\begin{aligned} \vec{s}^{\,(1)} &= \{1,1,-1,1\} \\ \vec{s}^{\,(2)} &= \{1,-1,1,-1\} \\ &\dots \\ \vec{s}^{\,(N)} &= \{1,1,-1,-1\} \end{aligned}\right\}$$

Data set:

N samples

$\hat{s} = \{\vec{s}^{\,(i)}\}$

- What would be the best model for the system, that could explain what we observe/reproduce similar data?



Parameters:

$$\vec{\theta} = \{h_1, J_{13}, J_{23}, J_{124}, J_{1234}\}$$

Bayesian Model Selection

- **The idea** Classifying all the possible models in term of their *posterior probability:*

$$P(\mathcal{M}_i \,|\, \hat{s})$$

in order to find the model $\mathcal{M}_i$ that has the highest probability to be a good model for the system given the data set $\hat{s}$

- **Difficulty** The huge number of models.

Ex. for spin models: — $2^n - 1$ possible type of interactions
— $2^{2^n - 1}$ possible models

Ex.

$n = 2:$	8 models	$n = 4:$	32768 models
$n = 3:$	128 models	$n = 5:$	2147483648 models

Bayesian Model Selection

- **The idea** Classifying all the possible models in term of their *posterior probability:*

$$P(\mathcal{M}_i \,|\, \hat{s})$$

in order to find the model $\mathcal{M}_i$ that has the highest probability to be a good model for the system given the data set $\hat{s}$

- **Difficulty** The huge number of models.

  Ex. for spin models: — $2^n - 1$ possible type of interactions

  — $2^{2^n - 1}$ possible models

- Simplify, using models with only fields and pairwise interactions:

  Number of possible model $\sim 2^{n^2}$

Ex.	$n = 2:$	8	models	$n = 4:$	1024	models
	$n = 3:$	64	models	$n = 5:$	32768	models

Only fields and pairwise interactions?

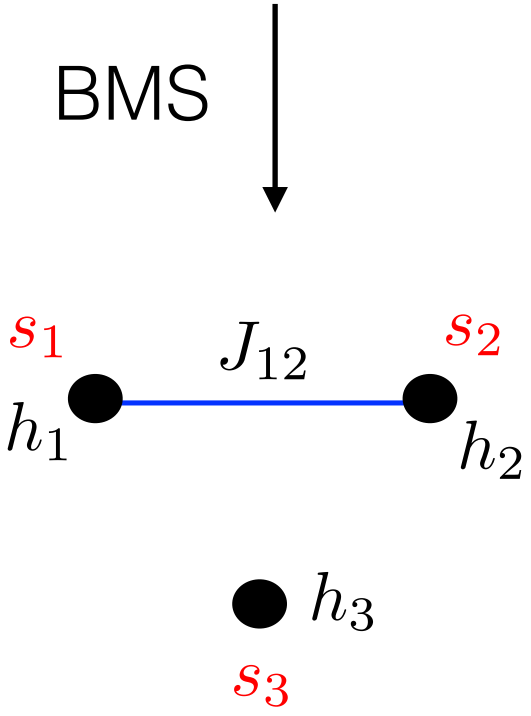
- Is it a good idea to restraint the choice to this type of models? Does pairwise/field interactions play a special role? Are these models *simpler?* (less *complex*)

- They don't seem to play a special role:

Ex.
with n=3

1rst data set

$$\hat{s} = \{\vec{s}^{(i)}\}$$



Only fields and pairwise interactions?

- Is it a good idea to restraint the choice to this type of models? Does pairwise/field interactions play a special role? Are these models *simpler?* (less *complex*)

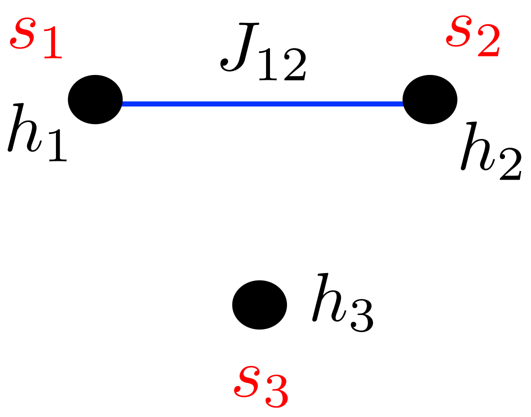
- They don't seem to play a special role:

Ex.
with n=3

1rst data set

2nd data set

$$\begin{tikzcd}
\hat{s} = \{\vec{s}^{\,(i)}\} \arrow[r, "\mathcal{T}", "\begin{pmatrix} \sigma_1 = s_1 s_2 s_3 \\ \sigma_2 = s_2 \\ \sigma_3 = s_3 \end{pmatrix}"'] \arrow[d, "\text{BMS}"'] & \hat{\sigma} = \{\vec{\sigma}^{\,(i)}\} \arrow[d, "\text{BMS}"'] \\
{} & {}
\end{tikzcd}$$

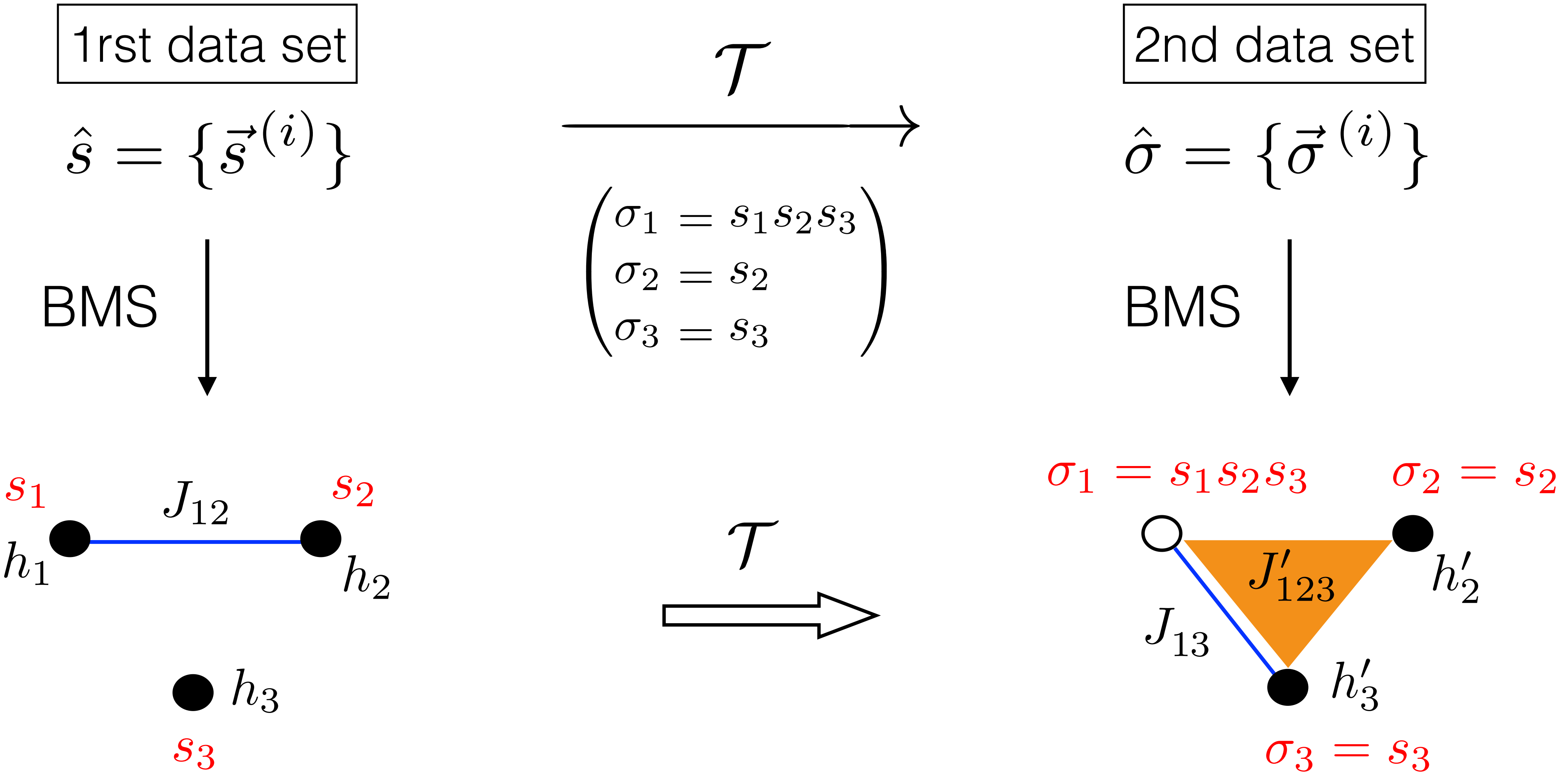


Only fields and pairwise interactions?

- Is it a good idea to restraint the choice to this type of models? Does pairwise/field interactions play a special role? Are these models *simpler?* (less *complex*)

- They don't seem to play a special role:

Ex.
with n=3



1rst data set

$\hat{s} = \{\vec{s}^{\,(i)}\}$

$\xrightarrow{\mathcal{T}}$

$$\begin{pmatrix} \sigma_1 = s_1 s_2 s_3 \\ \sigma_2 = s_2 \\ \sigma_3 = s_3 \end{pmatrix}$$

2nd data set

$\hat{\sigma} = \{\vec{\sigma}^{\,(i)}\}$

BMS

BMS

Bayesian Model Selection

- Using Bayes' theorem:

$$P(\mathcal{M}_i \,|\, \hat{s}) = \frac{P(\hat{s} \,|\, \mathcal{M}_i)\, P_0(\mathcal{M}_i)}{P(\hat{s})}$$

Bayesian Model Selection

- Using Bayes' theorem:

$$P(\mathcal{M}_i\,|\,\hat{s}) = \frac{\overbrace{P(\hat{s}\,|\,\mathcal{M}_i)}^{\text{Likelihood}}\;\overbrace{P_0(\mathcal{M}_i)}^{\text{Prior (uniform)}}}{P(\hat{s})}$$

—> Rank with the *Likelihood* $P(\hat{s}\,|\,\mathcal{M})$

Bayesian Model Selection

- Using Bayes' theorem:

$$P(\mathcal{M}_i \,|\, \hat{s}) = \frac{P(\hat{s} \,|\, \mathcal{M}_i)\, P_0(\mathcal{M}_i)}{P(\hat{s})}$$

—> Rank with the *Likelihood* $P(\hat{s} \,|\, \mathcal{M})$

- For large N, for the spin models, *Log-Likelihood* :

$$\log P(\hat{s} \,|\, \mathcal{M}_i) = \log P(\hat{s} \,|\, \mathcal{M}_i, \theta^*) - \frac{K}{2} \log\left(\frac{N}{2}\right) - c_{\mathcal{M}} + O\left(\frac{1}{N}\right)$$

$\propto N$

Log-Likelihood

Maximum Log-Likelihood

Bayesian Model Selection

- Using Bayes' theorem:

$$P(\mathcal{M}_i \,|\, \hat{s}) = \frac{P(\hat{s} \,|\, \mathcal{M}_i)\, P_0(\mathcal{M}_i)}{P(\hat{s})}$$

—> Rank with the *Likelihood* $P(\hat{s} \,|\, \mathcal{M})$

- For large N, for the spin models, *Log-Likelihood* :

$$\log P(\hat{s} \,|\, \mathcal{M}_i) = \log P(\hat{s} \,|\, \mathcal{M}_i, \theta^*) - \frac{K}{2} \log\left(\frac{N}{2}\right) - c_{\mathcal{M}} + O\left(\frac{1}{N}\right)$$

$\propto N$

$\propto \log N$

Log-Likelihood

Maximum Log-Likelihood

Penalty term: number of parameter K

Penalty term: geometrical complexity

Likelihood $P(\hat{s}\,|\,\mathcal{M})$

spin op. give an example

- Consider a model $\mathcal{M}$ with K parameters, $\vec{\theta} = (\theta_1, \ldots, \theta_K)$

$$P(\hat{s}\,|\,\mathcal{M}) = \int \mathrm{d}^K\vec{\theta}\ \underbrace{P\big(\hat{s}\,|\,\vec{\theta}, \mathcal{M}\big)}\ P_0\big(\vec{\theta}\,|\,\mathcal{M}\big)$$

N independent set of data

$$= \prod_{j=1}^{N} P\big(\vec{s}^{\,j}\,|\,\vec{\theta}, \mathcal{M}\big)$$

spin op. give an example

Likelihood $P(\hat{s}\,|\,\mathcal{M})$

- Consider a model $\mathcal{M}$ with K parameters, $\vec{\theta} = (\theta_1, \dots, \theta_K)$

$$P(\hat{s}\,|\,\mathcal{M}) = \int \mathrm{d}^K\vec{\theta}\;\underbrace{P(\hat{s}\,|\,\vec{\theta},\mathcal{M})}\,P_0(\vec{\theta}\,|\,\mathcal{M})$$

N independent set of data

$$= \prod_{j=1}^{N} P(\vec{s}^{\,j}\,|\,\vec{\theta},\mathcal{M})$$

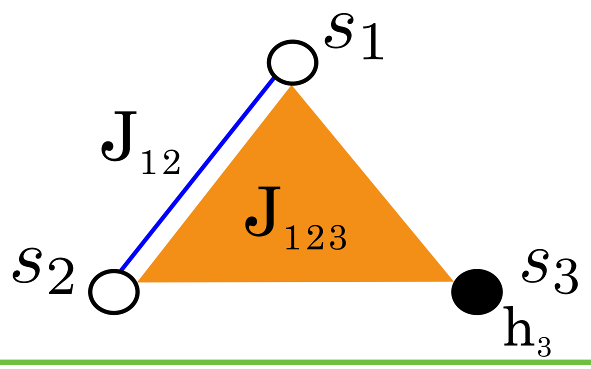
- Probability that the system is in the configuration $\vec{s}^{\,j}$ of the spins:

$$P(\vec{s}^{\,j}\,|\,\vec{\theta},\mathcal{M}) = \frac{\mathrm{e}^{\sum_{k=1}^{K} \boxed{f_k(\vec{s}^{\,j})}\,\theta_k}}{\boxed{Z_{\mathcal{M}}(\vec{\theta})}}$$

Spin operator:
product of the spins
involve in the interaction k

Partition function

Ex.



$$\mathrm{e}^{\,s_3\,h_3 + s_1\,s_2\,J_{12} + s_1\,s_2\,s_3\,J_{123}}$$

spin op. give an example

Likelihood $P(\hat{s}\,|\,\mathcal{M})$

- Consider a model $\mathcal{M}$ with K parameters, $\vec{\theta} = (\theta_1, \dots, \theta_K)$

$$P(\hat{s}\,|\,\mathcal{M}) = \int \mathrm{d}^K\vec{\theta}\ \underbrace{P(\hat{s}\,|\,\vec{\theta}, \mathcal{M})}\ P_0(\vec{\theta}\,|\,\mathcal{M})$$

N independent set of data

$$= \prod_{j=1}^{N} P(\vec{s}^{\,j}\,|\,\vec{\theta}, \mathcal{M})$$

- Probability that the system is in the configuration $\vec{s}^{\,j}$ of the spins:

$$\prod_{j=1}^{N} \qquad P(\vec{s}^{\,j}\,|\,\vec{\theta}, \mathcal{M}) = \frac{\mathrm{e}^{\sum_{k=1}^{K} \boxed{f_k(\vec{s}^{\,j})}\,\theta_k}}{\boxed{Z_{\mathcal{M}}(\vec{\theta})}}$$

Spin operator:
product of the spins
involve in the interaction k

Partition function

- Probability that the system produces the N configurations
$\hat{s} = \{\vec{s}^{\,1}, \dots, \vec{s}^{\,N}\}$ of the spins

Maximum Likelihood $P\big(\hat{s}\,|\,\vec{\theta}^{*},\mathcal{M}\big)$

- Probability that the system produces the N configurations $\hat{s}$ of the spins:

$$P\big(\hat{s}\,|\,\vec{\theta},\mathcal{M}\big) = \mathrm{e}^{\,N\,[\,\overbrace{\underbrace{\vec{F}(\hat{s})}\cdot\vec{\theta} - \log Z_{\mathcal{M}}(\vec{\theta}\,)}\,]}$$

$$= S(\vec{\theta})$$

empirical average

$$F_k(\hat{s}) = \frac{1}{N}\sum_{j=1}^{N} f_k(\vec{s}^{\,j}) = \langle f_k \rangle_{\hat{s}}$$

Maximum Likelihood $P\big(\hat{s}\,|\,\vec{\theta}^{*},\mathcal{M}\big)$

- Probability that the system produces the N configurations $\hat{s}$ of the spins:

$$P\big(\hat{s}\,|\,\vec{\theta},\mathcal{M}\big) = \mathrm{e}^{\,N\,[\,\overbrace{\underbrace{\vec{F}(\hat{s})}_{F_k(\hat{s}) = \frac{1}{N}\sum_{j=1}^{N} f_k(\vec{s}^{\,j}) = \langle f_k\rangle_{\hat{s}}}\cdot\vec{\theta} - \log Z_{\mathcal{M}}(\vec{\theta}\,)}^{= S(\vec{\theta})}\,]}$$

empirical average

- $S(\vec{\theta})$ is a concave function, as

Hessian Matrix: $$S''(\vec{\theta}) = \Big(\partial_{\theta_q}\partial_{\theta_k} S(\vec{\theta})\Big)_{q,k}$$

Maximum Likelihood $P(\hat{s}\,|\,\vec{\theta}^*,\mathcal{M})$

- Probability that the system produces the N configurations $\hat{s}$ of the spins:

$$P(\hat{s}\,|\,\vec{\theta},\mathcal{M}) = \mathrm{e}^{\,N\,[\,\overbrace{\underbrace{\vec{F}(\hat{s})}\cdot\vec{\theta} - \log Z_{\mathcal{M}}(\vec{\theta}\,)}\,]} \qquad = S(\vec{\theta})$$

empirical average

$$F_k(\hat{s}) = \frac{1}{N}\sum_{j=1}^{N} f_k(\vec{s}^{\,j}) = \langle f_k\rangle_{\hat{s}}$$

- $S(\vec{\theta})$ is a concave function, as

Hessian Matrix: $S''_{q,k}(\vec{\theta}) = -J_{q,k}(\vec{\theta})$

$\uparrow$ semi-negative definite $\qquad$ $\uparrow$ semi-positive definite

Fisher Information Matrix

$$\boxed{J(\vec{\theta}) = \big(\partial_{\theta_q}\partial_{\theta_k}\log Z_{\mathcal{M}}\big)_{q,k}}$$

- $P(\hat{s}\,|\,\vec{\theta},\mathcal{M})$ passes by a maximum at $\vec{\theta}^*$ such that:

$$\begin{cases} \vec{\nabla}S(\vec{\theta}^*) = \vec{0} \\ \det S''(\vec{\theta}^*) \neq 0 \end{cases} \quad\Longrightarrow\quad \boxed{\langle f_k\rangle_{\vec{s}}(\vec{\theta}^*) = \langle f_k\rangle_{\hat{s}}}$$

ensemble average $\qquad$ empirical average

Likelihood $P(\hat{s}\,|\,\mathcal{M})$ for large N

- We can re-write the Likelihood:

$$P(\hat{s}\,|\,\mathcal{M}) = \int \mathrm{e}^{\,N\,[\,\vec{F}(\hat{s})\cdot\vec{\theta}\,-\,\log Z_{\mathcal{M}}(\vec{\theta}\,)\,]}\; P_0\big(\vec{\theta}\,|\,\mathcal{M}\big)\,\mathrm{d}^K\vec{\theta}$$

- Expansion for large N, using the Saddle-point method in $\vec{\theta}^*$:

$$P(\hat{s}\,|\,\mathcal{M}) = \left[\frac{2\pi}{N}\right]^{K/2} \frac{P(\hat{s}\,|\,\vec{\theta}^*,\mathcal{M})}{\sqrt{\det J(\vec{\theta}^*)}} \left[P_0(\vec{\theta}^*\,|\,\mathcal{M}) + O\left(\frac{1}{N}\right)\right]$$

- Log-likelihood for large N

$$\log P(\hat{s}\,|\,\mathcal{M}) = \log P(\hat{s}\,|\,\mathcal{M},\theta^*) - \frac{K}{2}\log\left(\frac{N}{2}\right) - c_{\mathcal{M}} + O\left(\frac{1}{N}\right)$$

$$c_{\mathcal{M}} = \log\left[\frac{\sqrt{\det J(\vec{\theta}^*)}}{P_0(\vec{\theta}^*\,|\,\mathcal{M})}\right]$$

?

Complexity

- Given a model $\mathcal{M}$, the family of probability distribution: $\{P(\vec{f}\,|\,\vec{\theta}, \mathcal{M})\}$ forms a Riemannian manifold in the space of all distributions.

- In this space, each point, parametrised by $\vec{\theta}$, corresponds to a probability distribution $P(\vec{s}\,|\,\vec{\theta}, \mathcal{M})$

- The natural metric on this manifold is given by the FIM:

$$\begin{aligned} J(\vec{\theta}) &= \big(\partial_{\theta_q}\partial_{\theta_k} \log Z_{\mathcal{M}}\big)_{q,k} \\ &= \langle f_q f_k \rangle - \langle f_q \rangle \langle f_k \rangle \end{aligned}$$

- Varying the parameters of the model from a small $\mathrm{d}^K\vec{\theta}$ gives rise to similar distributions that correspond to nearby points in this space. The small volume that is formed by the variation $\mathrm{d}^K\vec{\theta}$ is then

$$\mathrm{d}V = \sqrt{\det J(\vec{\theta})}\,\mathrm{d}^K\vec{\theta}$$

Complexity

- Choice of the prior on the values of the parameters:

Jeffreys' prior $$P_0(\vec{\theta}\,|\,\mathcal{M}) = \frac{\sqrt{\det J(\vec{\theta})}}{\int \sqrt{\det J(\vec{\theta})}\,\mathrm{d}^K\vec{\theta}}$$

Best choice in absence of any information: its form stay invariant under re-parametrisation (doesn't give more importance to any parametrisation).

Geometrical complexity

$$c_{\mathcal{M}} = \log\left[\int_{\mathbb{R}^K} \sqrt{\det J(\vec{\theta})}\,\mathrm{d}^K\vec{\theta}\right]$$

- It is the volume of the entire manifold: Complexity represents how broad the model is in term of describing various probability distributions.
  A model is complex if it can fit a wide range of data. In a way, it also means that it is a poorly informative model.

$$J(\vec{\theta}) = \big(\partial_{\theta_q}\partial_{\theta_k}\log Z_{\mathcal{M}}\big)_{q,k}$$

Partition function

$$Z_{\mathcal{M}}(\vec{\theta}) = \sum_{\vec{s}=\{\pm 1\}^n} \mathrm{e}^{\sum_{k=1}^{K} f_k(\vec{s})\theta_k}$$

$$J(\vec{\theta}) = \left(\partial_{\theta_q}\partial_{\theta_k} \log Z_{\mathcal{M}}\right)_{q,k}$$

Partition function

$$Z_{\mathcal{M}}(\vec{\theta}) = \sum_{\vec{s}=\{\pm 1\}^n} \prod_{k=1}^{K} \mathrm{e}^{f_k(\vec{s})\theta_k}$$

Trick: $f_k(\vec{s}) \in \{-1, +1\}$

$$\mathrm{e}^{\theta_k f_k(\vec{s})} = \cosh(\theta_k)\left[1 + f_k(\vec{s})\tanh(\theta_k)\right]$$

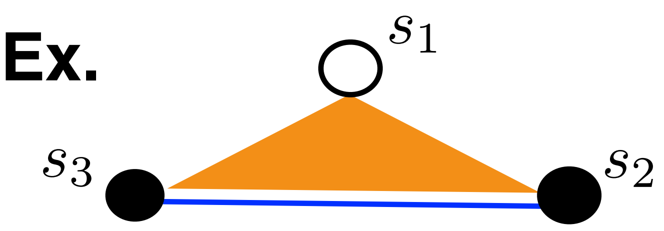
- We can re-write the partition function:

$$Z_{\mathcal{M}}(\vec{\theta}) = Z_0(\vec{\theta})\left[1 + \sum_{\ell\in\mathcal{L}}\prod_{\mu\in\ell}\tanh(\theta_\mu)\right], \qquad \text{where} \;\; Z_0(\vec{\theta}) = 2^n \prod_{\mu\in\mathcal{M}} \cosh(\theta_\mu)$$

$\mathcal{L}$ = Set of loops of the model

Loop = a set of operators of $\mathcal{M}$ such that the product of all operators of the set is = +1

Ex.



1 loop of 3 operators:

$$\mathcal{L} = \{\{s_2 s_2, s_2, s_3\}\}$$

$$J(\vec{\theta}) = \left(\partial_{\theta_q}\partial_{\theta_k} \log Z_{\mathcal{M}}\right)_{q,k}$$

Partition function

$$Z_{\mathcal{M}}(\vec{\theta}) = \sum_{\vec{s}=\{\pm 1\}^n} \prod_{k=1}^{K} \mathrm{e}^{f_k(\vec{s})\theta_k}$$

Trick: $f_k(\vec{s}) \in \{-1, +1\}$

$$\mathrm{e}^{\theta_k f_k(\vec{s})} = \cosh(\theta_k)\left[1 + f_k(\vec{s})\tanh(\theta_k)\right]$$

- We can re-write the partition function:

$$Z_{\mathcal{M}}(\vec{\theta}) = Z_0(\vec{\theta})\left[1 + \sum_{\ell\in\mathcal{L}}\prod_{\mu\in\ell}\tanh(\theta_\mu)\right], \qquad \text{where } \; Z_0(\vec{\theta}) = 2^n \prod_{\mu\in\mathcal{M}} \cosh(\theta_\mu)$$

$\mathcal{L}$ = Set of loops of the model

- The structure of $Z_{\mathcal{M}}(\vec{\theta})$ depends only on the structure of loops $\mathcal{L}$:
  — the number of loops
  — the length of each loop
  — how the operators are in relations through the loops

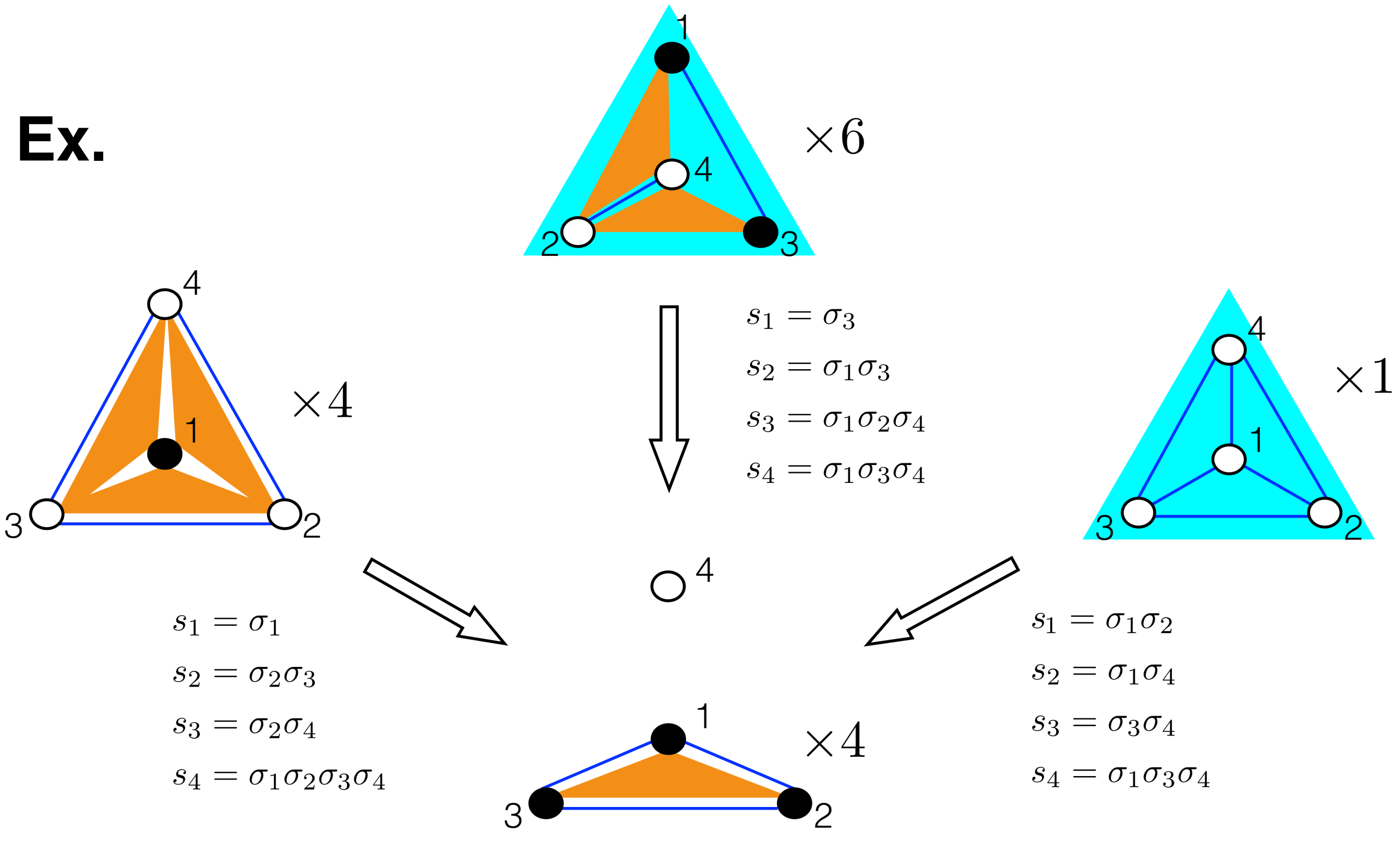
—> Models with the same loop structure has the same Complexity

Gauge Transformations

- It exists a certain number of transformations that conserve the loop structure of the models:
  they are the **bijections that map the a set of n spins $\vec{s}$** to another set of n **spins $\vec{\sigma}$** (while conserving the structure the operators that can be created with this spins)

- *Gauge Transformations* (Automorphisms of the group of operators): they conserve the geometry of the model

Ex.



$s_1 = \sigma_3$

$s_2 = \sigma_1\sigma_3$

$s_3 = \sigma_1\sigma_2\sigma_4$

$s_4 = \sigma_1\sigma_3\sigma_4$

$s_1 = \sigma_1$

$s_2 = \sigma_2\sigma_3$

$s_3 = \sigma_2\sigma_4$

$s_4 = \sigma_1\sigma_2\sigma_3\sigma_4$

$s_1 = \sigma_1\sigma_2$

$s_2 = \sigma_1\sigma_4$

$s_3 = \sigma_3\sigma_4$

$s_4 = \sigma_1\sigma_3\sigma_4$

7 interactions

15 loops

7 loops of 3 interactions
7 loops of 4 interactions
1 loop of 7 interactions

Gauge Transformations

- It exists a certain number of transformations that conserve the loop structure of the models: they are the bijections that map the set of spin s to another set of spin sig, while conserving the structure of all the operators

- = *Gauge Transformations* (Automorphisms of the group of operators)

- Total number of *Gauge Transformations* for a system with n spins:

$$\mathcal{N}_{GT}(n) = \prod_{i=0}^{n-1} (2^n - 2^i)$$

Ex. $\mathcal{N}_{GT}(3) = 168$

$\mathcal{N}_{GT}(4) = 20160$

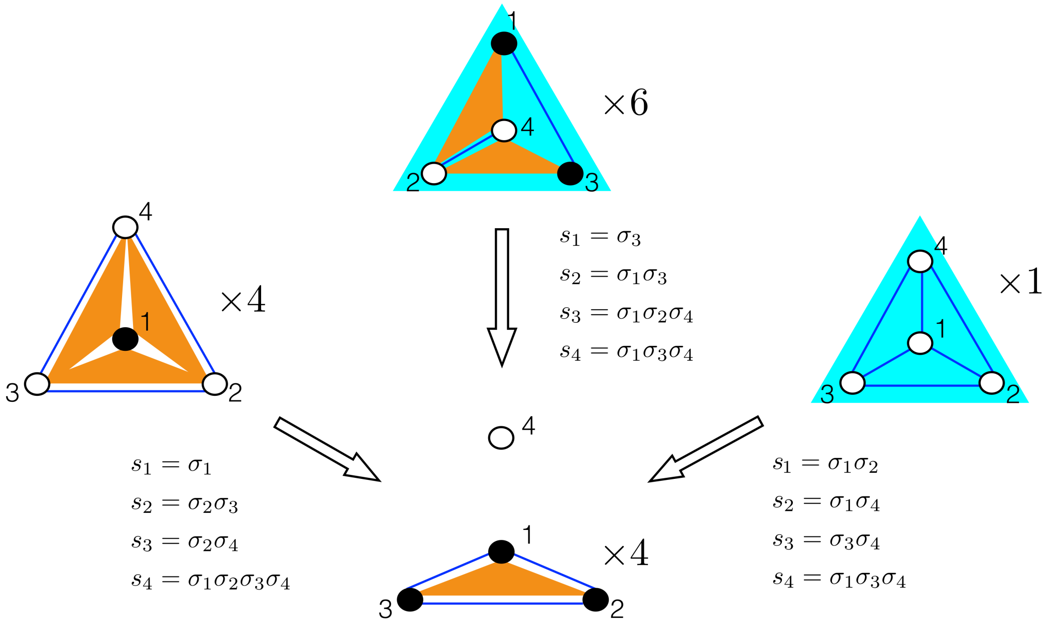
Complexity Classes

- Gauge Transformations (GT) allow us to partition the space of models into classes of models that are images from one to another by GT.

- Within a class models have
  — the same geometrical properties (loop structure)
  — the same complexity
  ——> "*Complexity Classes*"

- The cardinality of a class is necessary $\leq \mathcal{N}_{GT}(n)$ and is more precisely given by the invariants under GT

Ex.



$s_1 = \sigma_3$
$s_2 = \sigma_1\sigma_3$
$s_3 = \sigma_1\sigma_2\sigma_4$
$s_4 = \sigma_1\sigma_3\sigma_4$

$s_1 = \sigma_1$
$s_2 = \sigma_2\sigma_3$
$s_3 = \sigma_2\sigma_4$
$s_4 = \sigma_1\sigma_2\sigma_3\sigma_4$

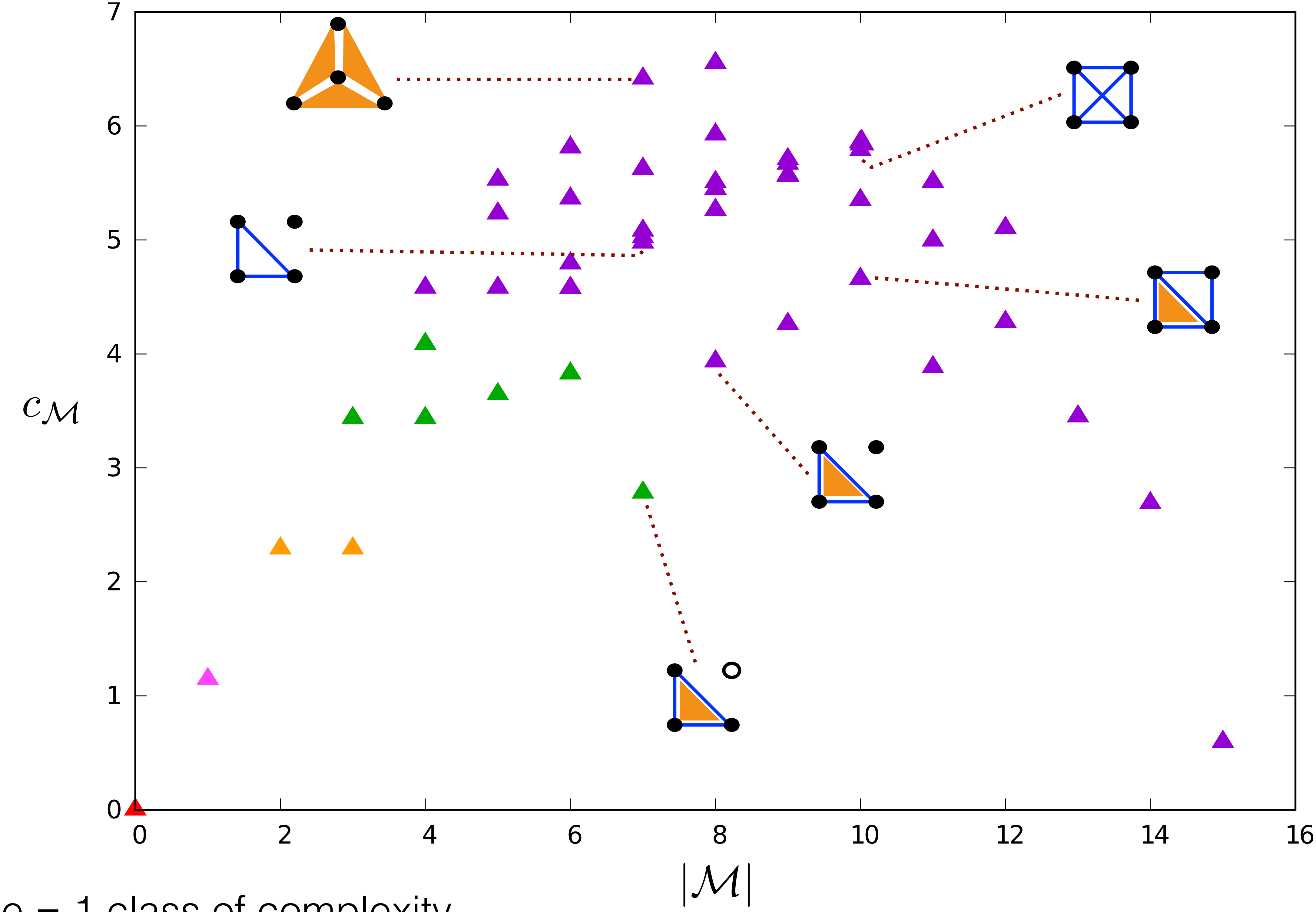
$s_1 = \sigma_1\sigma_2$
$s_2 = \sigma_1\sigma_4$
$s_3 = \sigma_3\sigma_4$
$s_4 = \sigma_1\sigma_3\sigma_4$

Class with 15 models

classe

Complexity for n=4

32768 models,   only **46** complexity classes

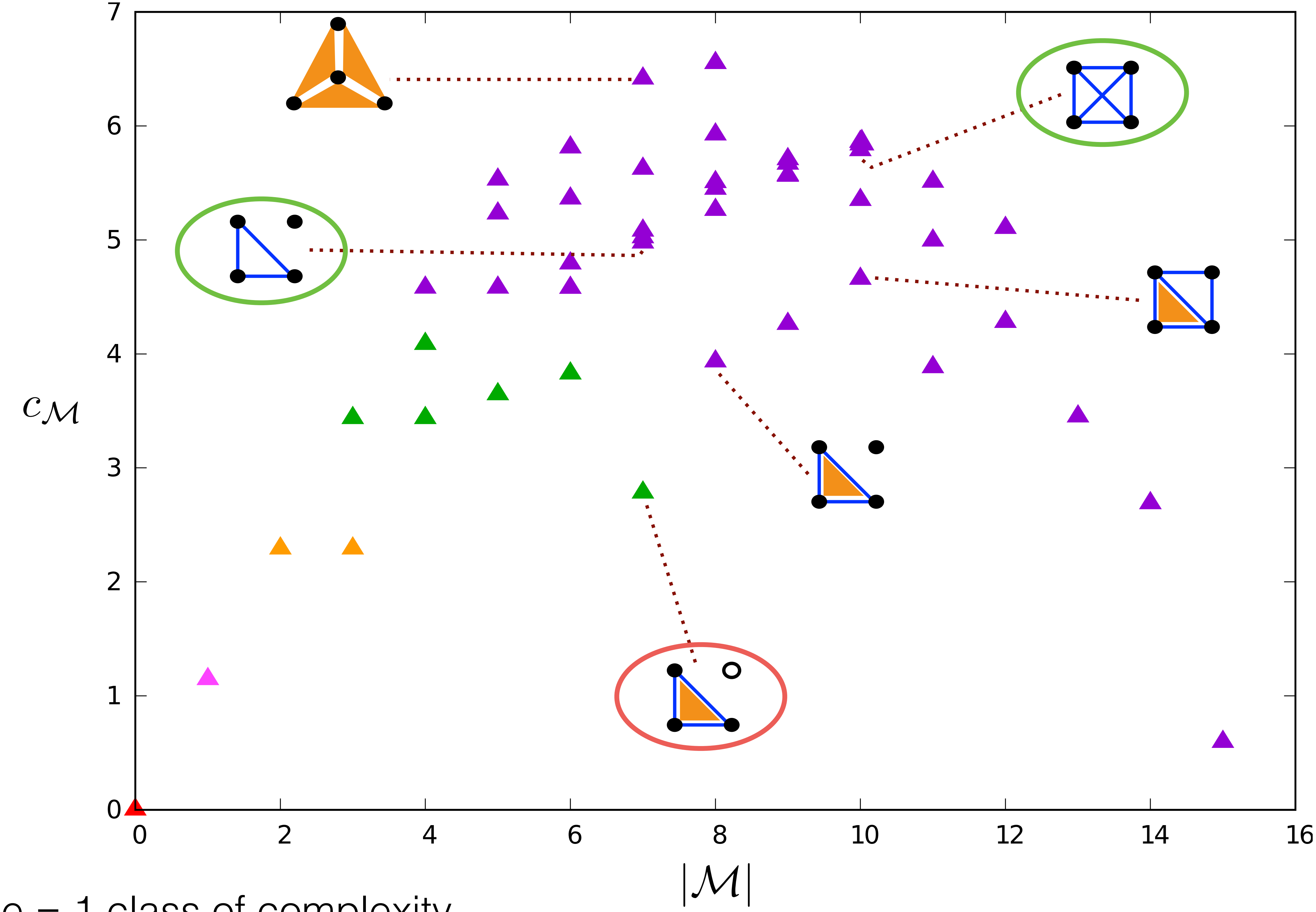


1 triangle = 1 class of complexity

number of parameter

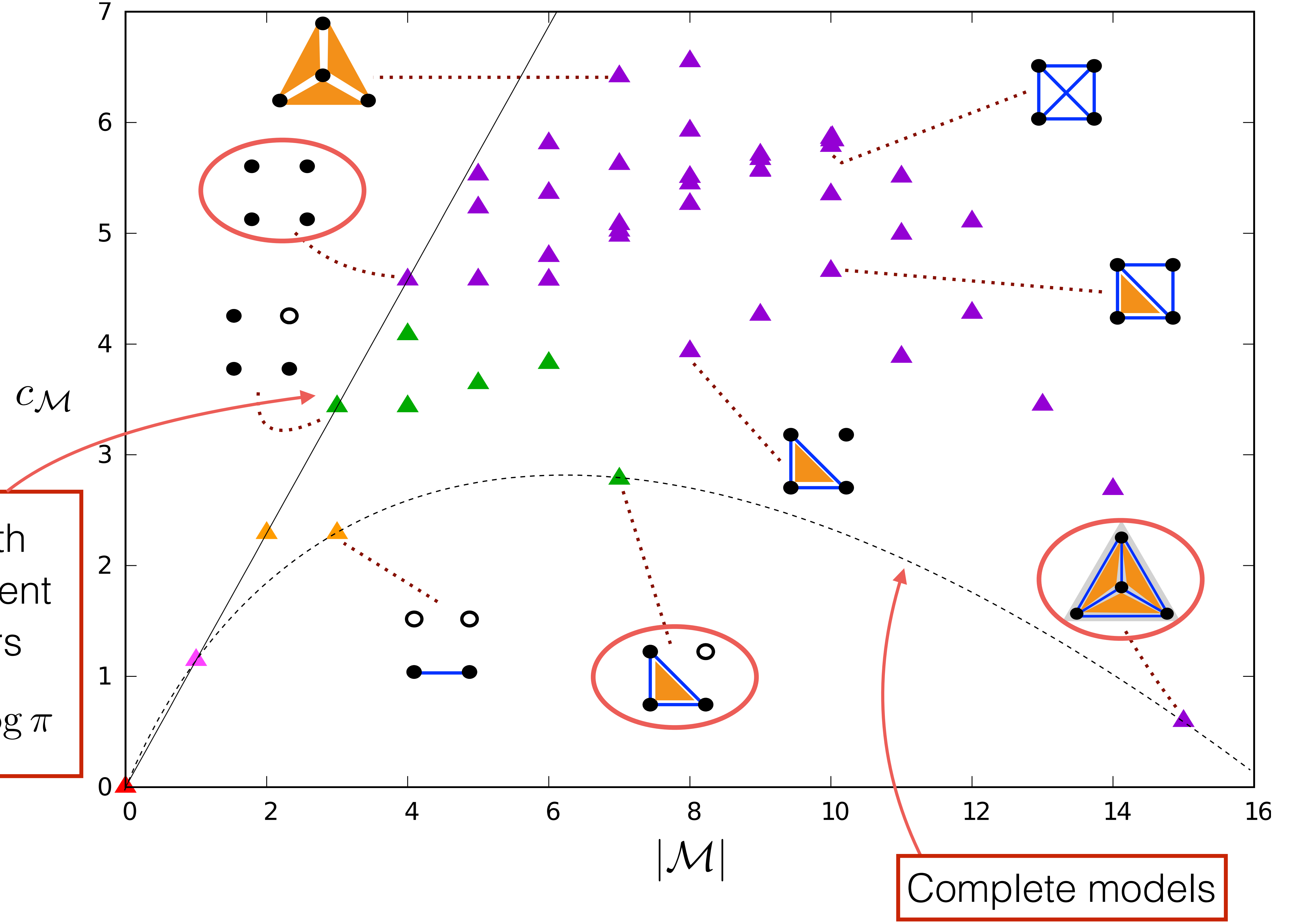
Complexity for n=4

32768 models, only **46** complexity classes



1 triangle = 1 class of complexity

Complexity for n=4



Models with K independent parameters

$c_{\mathcal{M}} = |\mathcal{M}| \log \pi$

Complete models

$\longrightarrow c_{\mathcal{M}}(n=5) = -9.58$

Complexity

- Complexity is the characteristic of a model that can account for a broad range of data:
  - — the most complex models: where all parameters are independent
  - — the simplest models: the *complete models*

- Models are partitioned into ***complexity classes***, in which models are connected by ***Gauge Transformations*** and share the same geometrical structure (loop structure, same invariants)

- If there were a way of choosing the best model in term of its internal geometries this will highly reduce the number of models that should be tested Geometries of the models are directly related to the patterns that are hidden inside the data.

- Complexity is not monotonic with the number K of parameters:

$$\log P(\hat{s}\,|\,\mathcal{M}) = \log P(\hat{s}\,|\,\mathcal{M},\theta^*) - \frac{K}{2}\log\left(\frac{N}{2}\right) - c_{\mathcal{M}} + O\left(\frac{1}{N}\right)$$

Depending on the values of N, which method should we adopt?