

A Toric Variety from Machine Learning

Zhangsheng Lai

Orlando Marigliano

April 6, 2018

Given a directed graph G without multiple edges, we introduce a biology-inspired statistical model, the *McCulloch-Pitts process (MPP)*, for recurrent neural networks. We associate a toric variety to such a model and compute its Hilbert polynomial in a special case.

The statistical model

Given a directed graph $G = (V, E)$ without multiple edges, with vertex weights $\beta_i > 0$ and edge weights $\alpha_{ij} > 0$, a *McCulloch-Pitts process (MPP)* is an activity-based process with binary states $x \in \{0, 1\}^{|V|}$ and transitions xy where state y is one-bit away from state x . If y and x differs in the i -th bit, we define the transition rate

$$F_{xy} = [\beta_i^{\sigma_i} \alpha_i^{x\sigma_i}]^{1/\tau} \text{ what is } \sigma_i?$$

where $\alpha_i^x = \alpha_{1i}^{x_1} \alpha_{2i}^{x_2} \dots \alpha_{di}^{x_d}$ where $|V| = d$.

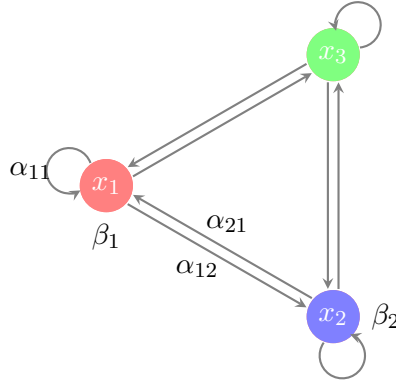


Figure 1: McCulloch-Pitts process with three neurons.

Working with the example in Figure 1 and choosing $\tau = 1$, the transition rate matrix is given by

$$F = \begin{array}{c|cccccccc} & 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\ \hline 000 & * & \beta_3 & \beta_2 & 0 & \beta_1 & 0 & 0 & 0 \\ 001 & \beta_3^{-1}\alpha_{33}^{-1} & * & 0 & \beta_2\alpha_{32} & 0 & \beta_1\alpha_{31} & 0 & 0 \\ 010 & \beta_2^{-1}\alpha_{22}^{-1} & 0 & * & \beta_3\alpha_{23} & 0 & 0 & \beta_1\alpha_{21} & 0 \\ 011 & 0 & \beta_2^{-1}\alpha_{22}^{-1}\alpha_{32}^{-1} & \beta_3^{-1}\alpha_{23}^{-1}\alpha_{33}^{-1} & * & 0 & 0 & 0 & \beta_1\alpha_{21}\alpha_{31} \\ 100 & \beta_1^{-1}\alpha_{11}^{-1} & 0 & 0 & 0 & * & \beta_3\alpha_{13} & \beta_2\alpha_{12} & 0 \\ 101 & 0 & \beta_1^{-1}\alpha_{11}^{-1}\alpha_{31}^{-1} & 0 & 0 & \beta_3^{-1}\alpha_{13}^{-1}\alpha_{33}^{-1} & * & 0 & \beta_2\alpha_{12}\alpha_{32} \\ 110 & 0 & 0 & \beta_1^{-1}\alpha_{11}^{-1}\alpha_{21}^{-1} & 0 & \beta_2^{-1}\alpha_{12}^{-1}\alpha_{22}^{-1} & 0 & * & \beta_3\alpha_{13}\alpha_{23} \\ 111 & 0 & 0 & 0 & (\beta_1\alpha_1^{111})^{-1} & 0 & (\beta_2\alpha_2^{111})^{-1} & (\beta_3\alpha_3^{111})^{-1} & * \end{array}$$

where $*$ denotes the negative of the sum of its corresponding row, and $\alpha_i^{111} = \alpha_{1i}\alpha_{2i}\alpha_{3i}$.

Simulation

The simulation of the McCulloch-Pitts process starts by drawing an initial state $x^{(0)}$ from a distribution $p^{(0)}$. Then for any state x , it holds for some time

$$\Delta t \sim \text{Exp}(\lambda_x)$$

where $\lambda_x = \sum_{y \neq x} F_{xy}$ and it transits to state y which is one hop away with probability

$$F_{xy}/\lambda_x$$

Thus the temporal data obtained from the simulation are binary tuples of length $|V|$ and an associated holding time for each pair of consecutive state.

Could you write something about the connection to machine learning/recurrent neural networks? Thanks!

Toric Variety

I would remove this section if this is ok. I wouldn't redefine what it means to be a toric variety etc, as we want to keep this short.

With the statistical model defined proper, we would like to show that given directed graph $G = (V, E)$ with $|V| = d$, with non-negative vertex and edge weights β_i, α_{ij} respectively, we have a map $F : \mathbb{C}^{d(d+1)} \rightarrow \mathbb{C}^{d \cdot 2^d}$ defined by $(\beta_i, \alpha_{ij}) \mapsto (F_{xy})$, from the space of non-negative parameters to the space of stochastic space matrices and the closure(note: do we really need to take its closure or it is guaranteed to be close?) of the image is a toric variety.

The Zariski closure of $X \subset \mathbb{C}$, is

$$\overline{X} := \{p \in \mathbb{C}^n \mid f(p) = 0 \text{ whenever } f(X) = 0\}$$

Definition 1. A toric variety is an irreducible variety V such that

- $(\mathbb{C}^*)^n$ is a Zariski open subset of V
- the action of $(\mathbb{C}^*)^n$ on itself extends to an action of $(\mathbb{C}^*)^n$ on V .

Next we would like know the degree and dimension of our toric variety. By first principles, we can obtain them by....

I would like to give some heuristics here, like the degree is the number of interections when we stab the toric variety with a line, plane or in general with a hyperplane. What good does knowing the dimension and degree helps us with and how the method of finding the dimension we used on Wednesday is equivalent to finding the dimension and degree by the method we discuss above.

The Toric Variety

We consider the space of weights $W := \mathbb{C}^{d+|E|} = \{(\beta_i, \alpha_{jk}) \mid i \in V, (j, k) \in E\}$ and the space of transition rates $T := \mathbb{C}^{2^d} = \{(F_{xy}) \mid x, y \text{ binary states differing at one bit}\}$. We have a map $f: W \rightarrow T$ defined by $f(\alpha, \beta) = (F_{xy}(\alpha, \beta))_{xy}$. We define the toric variety X as the Zariski closure of the image of f .

In the above example, we get the map $f: \mathbb{C}^{12} \rightarrow \mathbb{C}^{24}$ and the induced toric variety X . Using Polymake, we can compute the Hilbert series of (the closure in \mathbb{P}^{24} of) X . We obtain The Hilbert series

$$\frac{P(x)}{(1-x)^{12}}$$

with

$$P(x) = x^6 + 12x^5 + 51x^4 + 88x^3 + 51x^2 + 12x + 1.$$

The dimension of X is thus 12, i. e. the degree of the denominator, and the degree of X is $P(1) = 216$.

A group action

Let π be a graph isomorphism of G , seen as a map $V \rightarrow V$. The map π acts on the space of weights by sending an element $(\beta_i, \alpha_{j,k})$ to $(\beta_{\pi i}, \alpha_{\pi j, \pi k})$, and similarly it acts on the space of transition rates. Since f is a Laurent map, for all weights (α, β) we have $\pi f(\alpha, \beta) = f\pi(\alpha, \beta)$. Hence we obtain a group action of $\text{Aut}(G)$ on the variety X .