

Mathematical Data Science HW8

20180617 You SeungWoo

November 26, 2023

Problem 1

Solution. We finally use fine dust distribution data. First, we design a 2D CA simulator and then compare its results with distributed data. We confirmed that this data is provided as public data by many institutions both domestically and overseas. But 2D CA implementation requires a lot of other data.

First of all, we need wind data. So, we extract the wind direction and speed as data from a site that provides real-time wind direction around the world. Of course, this data must be refined and used for our purposes. We will also utilize factory distribution data. From public data from domestic government agencies, the distribution of industrial complexes across the country is applied to the map. Factories may generate fine dust periodically. Since it is difficult to consider all factories, the distribution of factories in the map and region used is calculated and applied proportionally.

Additionally, population numbers, environmental protection areas, mountain ranges, and river distribution can be considered. We plan to consider this data after first completing a prototype simulator and then comparing it with actual fine dust data to find insights.

□

Problem 2

Proof. Define a set of k -simplex be Σ_k . First, consider X . We need $k = 0, 1, 2$.

$$\begin{aligned}\Sigma_0 &= \{[x_0], [x_1], [x_2], [x_3], [x_4]\} \\ \Sigma_1 &= \{[x_0, x_1], [x_1, x_2], [x_0, x_2], [x_2, x_3], [x_2, x_4], [x_3, x_4]\} \\ \Sigma_2 &= \emptyset\end{aligned}$$

Since there is only one graph, $\beta_0 = 1$. For β_1 , define the first chain module C_1 . for $\sigma \in C_1$,

$$\begin{aligned}\sigma &= t_1[x_0, x_1] + t_2[x_0, x_2] + t_3[x_1, x_2] \\ &\quad + t_4[x_2, x_3] + t_5[x_2, x_4] + t_6[x_3, x_4] \quad \text{for } t_i \in \{0, 1\}\end{aligned}$$

By the boundary operator ∂_1 , we get

$$\begin{aligned}\partial_1 \sigma &= t_1([x_0] + [x_1]) + t_2([x_0] + [x_2]) + t_3([x_1] + [x_2]) \\ &\quad + t_4([x_2] + [x_3]) + t_5([x_2] + [x_4]) + t_6([x_3] + [x_4]) \\ &= (t_1 + t_2)[x_0] + (t_1 + t_3)[x_1] + (t_2 + t_3 + t_4 + t_5)[x_2] \\ &\quad + (t_4 + t_6)[x_3] + (t_5 + t_6)[x_4]\end{aligned}$$

Find solutions for t_i . This means we need to solve the below matrix equation:

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

From this, we get

$$\begin{aligned}\sigma &= t_1([x_0, x_1] + [x_0, x_2] + [x_1, x_2]) \\ &\quad + t_4([x_2, x_3] + [x_2, x_4] + [x_3, x_4])\end{aligned}$$

This means there are 2 cycles: $x_0 - x_1 - x_2$ and $x_2 - x_3 - x_4$. Since $\Sigma_2 = \emptyset$, we have no plane. Therefore, $\beta_1 = 2 - 0 = 2$.

Similarly, for Y , $\beta_0 = 1$. But Y has $\Sigma_2 = \{[x_2, x_3, x_4]\}$. This means there is a plane $x_2 - x_3 - x_4$. Therefore, $\beta_1 = 2 - 1 = 1$.

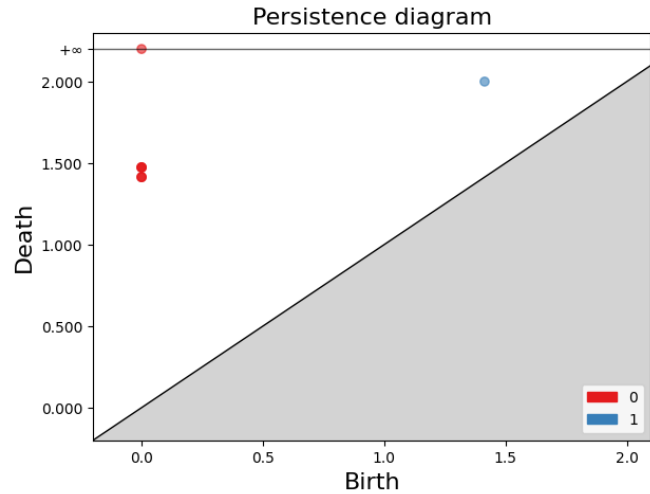
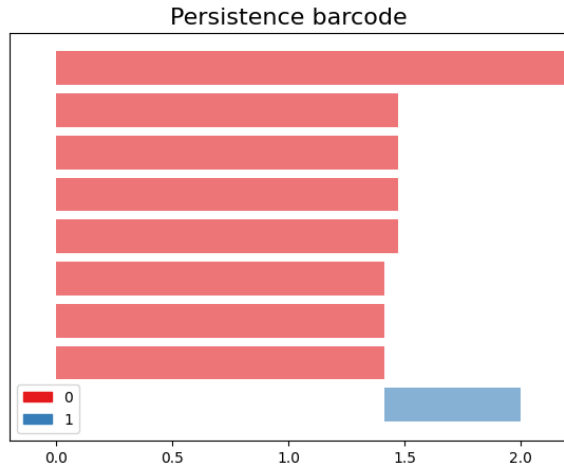
□

Problem 3

Solution. We can expect the barcode as following:

- At first, there are total 8 points. Thus start with 8 0th dimension.
- After some time has passed, since all points have the same distance, they are connected at the same time. Thus 7 points will be dead (note that one 0th dimension is still left) and one 1st dimension complex will be created at that time.

To check this, plot the result by using Python.



□

Problem 4

Solution. 작가의 핵심 전략인 sensemaking에 대해 설명하고 있다. Sensemaking은 인문학에 기초해, 실용적 지혜를 얻는 방식이다. 맥락을 파악하고 그 맥락의 인간 관계로 행동 패턴을 찾아내는 전략을 사용한다. 이 전략은 우리는 인간 사회를 살아가고, 그렇기에 인간 관계에 대한 패턴이 실제 비즈니스에 매우 중요할 것이라는 아이디어에서 출발한다. 이러한 기초 때문에, 작가는 문화에 대한 체험과 깊은 통찰을 중요하게 생각한다. 이러한 아이디어는 데이터와 컴퓨터 알고리즘에 기반한 분석을 중요하게 생각하는 현대의 사고와 정 반대의 전략이다. Sensemaking은 폭 넓은 연결성을 파악한다는 점에서, 빅데이터가 해결하지 못하는 문제를 새로운 관점으로 바라보도록 한다. 또한 빅데이터에서 분석의 결과를 해석하는 것이 중요하지만 이것이 사람에 따라 의미가 달라질 수 있다는 문제가 있다. 이러한 문제는 sensemaking에서 예술과 인문학에 대한 이해로 해결할 수 있다고 주장한다.

□