

LETTER CLASSIFICATION

Group M



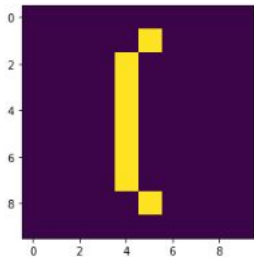
INTRODUCTION & BACKGROUND

- **classify a letter given a string of 100 zeros and ones using persistent homology**
 - **Computing vector of feature for doing comparison**
 - **Using different ways to scan letters**
- **scan all 26 letters of the Latin alphabet in different ways by using persistent homology**
 - **Persistent homology**
 - **Lower Star Image Filtrations**

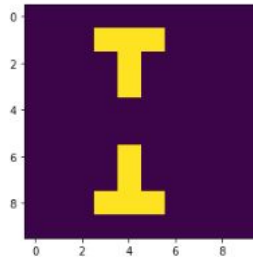
EXPERIMENT

- **create a test feature block which will contain all the implemented features**
 - **First, comparing feature vector to feature matrix**
 - **Second, make some changes on the input sequence of zeros and ones**
 - **Take out points randomly from the sequence**
 - **Take out points manually to see what happens next.**

```
[8 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0  
0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
feature vect of input list is:[96. 95. 96. 96. 96. 0.]
standard feature vect of I is:[97. 97. 96. 96. 96. 0.]
the best fitted letter is: 8 th letter
the closest distance is: 2.23606797749979
```



```
[8 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
feature vect of input list is: [191. 192. 190. 190. 190.      0.]
standard feature vect of I is: [97. 97. 96. 96. 96. 0.]
the best fitted letter is: 4 th letter
the closest distance is: 201.45604322958934
```



ADVANTAGE AND DISADVANTAGE

Pro

- successfully differentiated between letters
- able to recognize letters with original input, with no points taken off
- complete control over the algorithm
- can bring our classifications into other systems
- compared to other ML sets we do not need to train our algorithm every time

Con

- if we randomly took points off a letter, the system tend to produce the wrong result.
- Especially when letters are broken into different components, the system produce a vector with abnormally large norm, and tend to recognize the letter as E, since it has the largest norm.