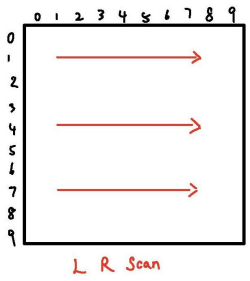
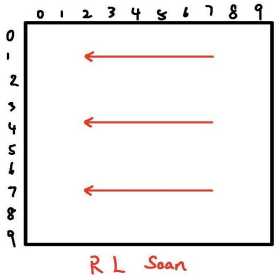
Letter classification

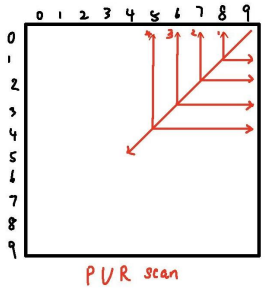
Group M

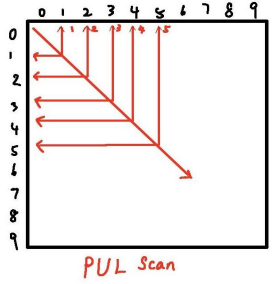
Introduction & Background

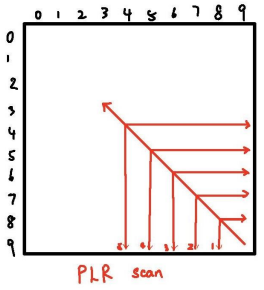
- **Classify a letter given a string of 100 zeros and ones using persistent homology**
  - **Using different ways to scan letters**
    - **Compute feature vector for doing comparison**









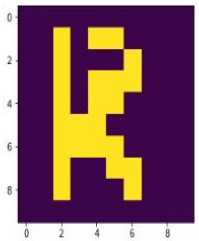


- **Scan all 26 letters of the Latin alphabet in different ways by using persistent homology**
  - **Persistent homology**
  - **Lower Star Image Filtrations**

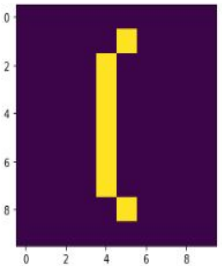
Experiment

- **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.**

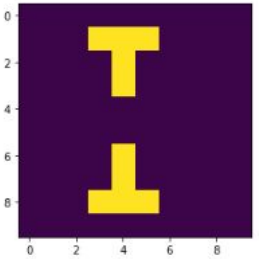
```
[11  0  0  0  0  0  0  0  0  0  0  0  0  1  0  1  1  0  0  0  0  0  0  1
  0  0  0  1  0  0  0  0  0  0  1  0  1  1  1  1  0  0  0  0  0  1  0  1  1  0
  0  0  0  0  0  1  1  1  1  0  0  0  0  0  0  0  1  1  1  1  1  0  0  0  0  0
  0  1  0  0  1  1  0  0  0  0  0  1  0  0  0  1  0  0  0  0  0  0  0  0  0  0
  0  0  0  0  0]
feature vect of input list is: [101.          99.          98.          97.
   1.95084465]
the best fitted letter is: 21 th letter
 the closest distance is: 2.7131596267111178
```



```
[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 1 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 1 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 1 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 1 1 1 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 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 1 0 0 0 0 0 0 0 0 1 1 1 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
```



Advantages & Disadvantages

Pro

- Able to recognize letters without any points taken off
- Complete control over the algorithm
- Compared to other ML sets we do not need to train our algorithm everytime
- Can bring our classifications into other systems

Con

- If we randomly took points off a letter, the system tends to produce the wrong result
- Especially when letters are broken into different components, the system produces a vector with abnormally large norm.