

Assignment 10

Alberto Mejia

RIN: 661514960

CSCI 4100 - Machine Learning from Data

November 17, 2019

1. (100) LFD Exercise 6.1

(a)

High Cosine Similarity and Low Euclidean distance Similarity

Vector 1 = [1, 2, 3]

Vector 2 = [1000, 2000, 3000]

The Euclidean distance Similarity = 3738

The Cosine Similarity = 1.00

Low Cosine Similarity and High Euclidean distance Similarity

Vector 3 = [0.1, 0.2, 0.3]

Vector 4 = [-0.0001, -0.0002, -0.0003]

The Euclidean distance Similarity = 0.3745

The Cosine Similarity = -1

(b) The cosine similarity does change but only by a little because it only depends on the angle between vectors not the magnitude. Therefore, it is a measure of orientation and would not be affected that much by a change of origin. The Euclidean distance similarity on the other hand does not change.

2. (100) LFD Exercise 6.2

Let

$$f(x) = \begin{cases} +1, & \text{if } \pi(x) \geq \frac{1}{2} \\ -1, & \text{otherwise.} \end{cases}$$

Show

$$e(f(x)) = P[f(x) \neq y] = \min\{\pi(x), 1 - \pi(x)\} \text{ and } e(f(x)) \leq e(h(x)) \text{ for any other } h$$

$$e(f(x)) = P[f(x) \neq y]$$

$$e(f(x)) = P[\pi(x) \neq y]$$

$$= P[f(x) = +1, y = -1] + P[f(x) = -1, y = +1]$$

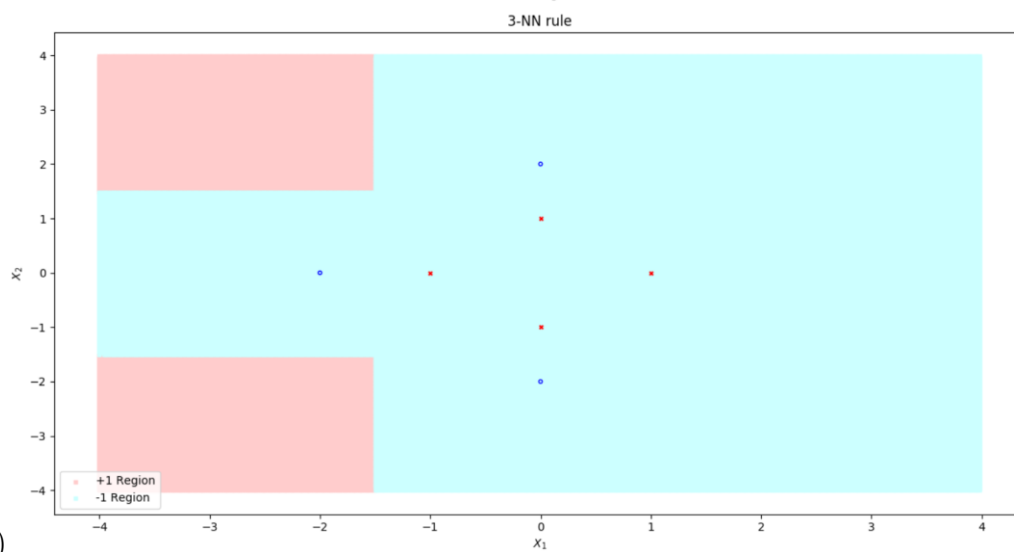
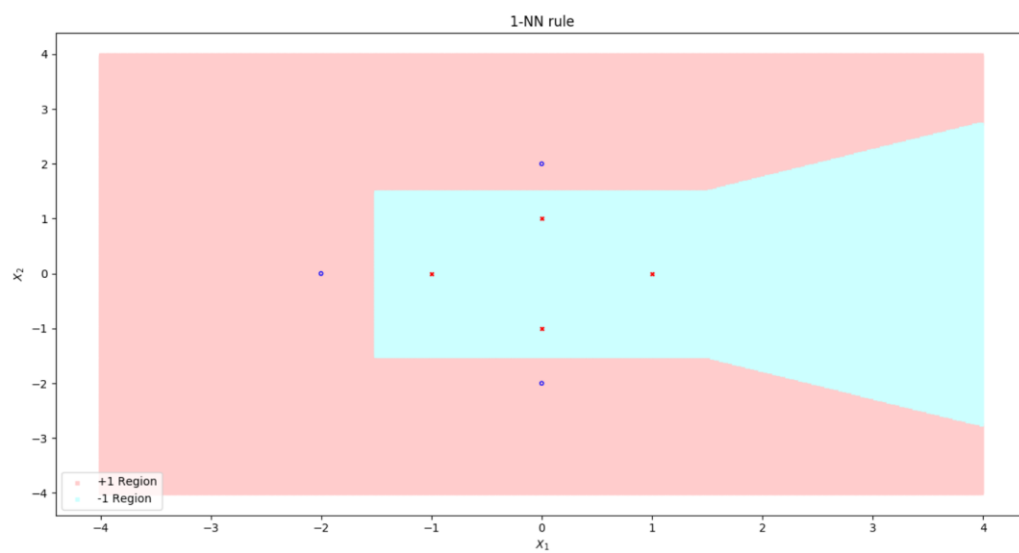
$$= \frac{1}{2}\pi(x) \times (1 - \pi(x)) + (1 - \pi(x)) \times \frac{1}{2}\pi(x)$$

$$= \pi(x) \times (1 - \pi(x))$$

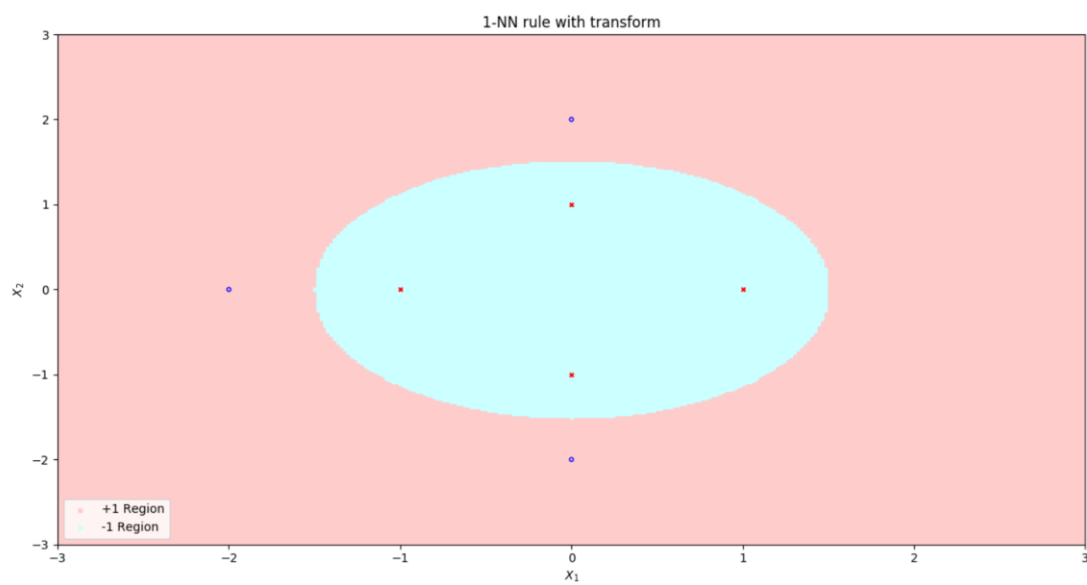
$$e(f(x)) = \min[\pi(x), (1 - \pi(x))]$$

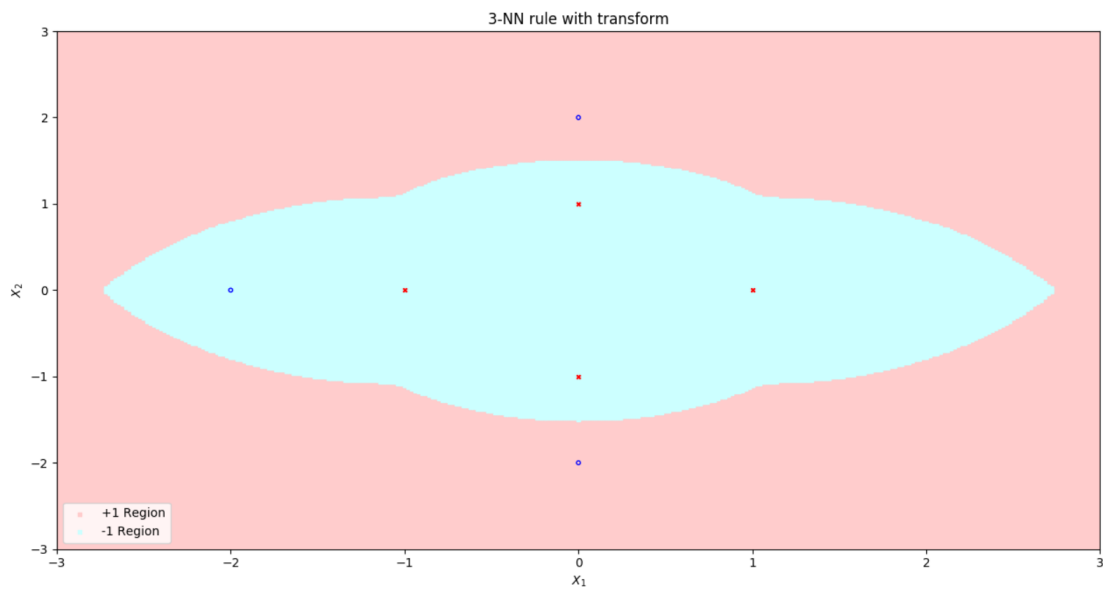
3.(200) **LFD Problem 6.1**

(a)

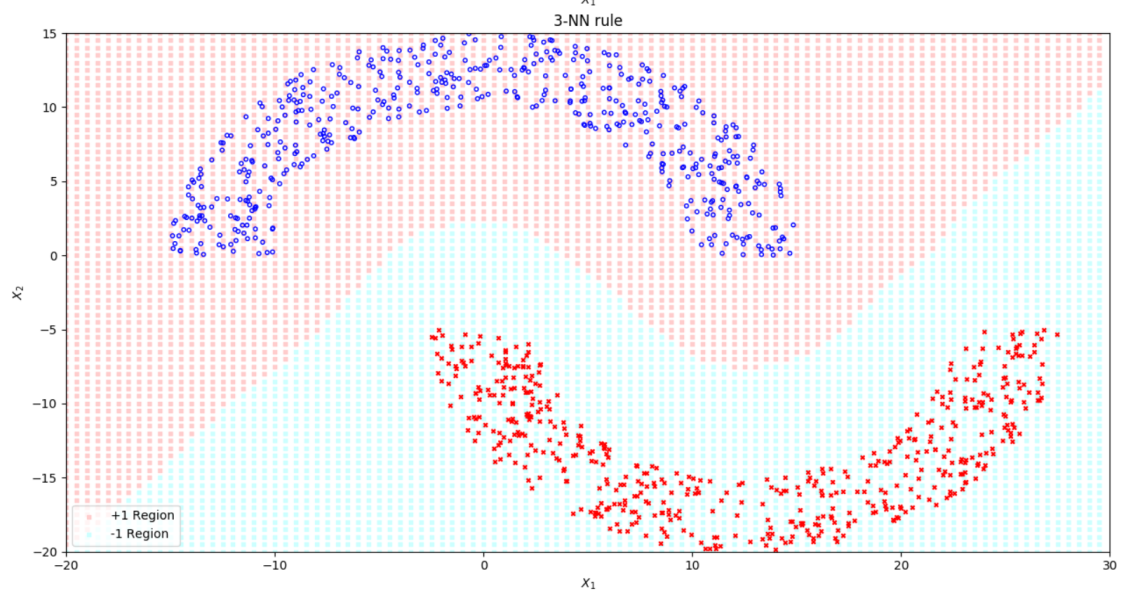
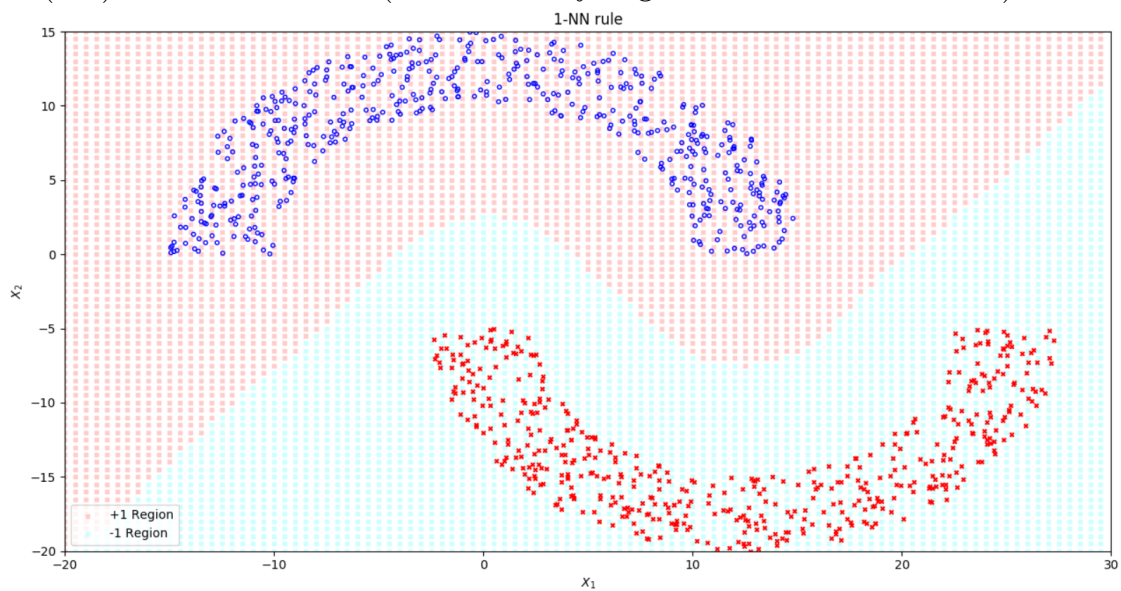


(b)





4. (200) **LFD Problem 6.4** (use the data you generated for Problem 3.1)



5. (400) **LFD Problem 6.16**

(a)

Finding NN with partition took 31.6s

Finding NN with brute force took 201.93s

(b)

Finding NN with partition took 30.77s

Finding NN with brute force took 207.2s

(c) Doing the brute force way forces us to search our entire problem space, which is 10,000 points. By using partitions, we can cluster together the points we want to search (using their relative distances to other points), reducing our search space immensely.

(d) No, it does not affect the decision. This is because the branch and bound will perform better than the brute force method most of the time.