CSCI 4100 Assignment 10

Boliang Yang 661541863

1. Exercise 6.1 in LFD

   (a) Two vectors with very high cosine similarity but with very low Eucliddean distance similarity can be chosen as two vectors with same direction but different length. For example, $\boldsymbol{x} = \begin{bmatrix} 10 \\ 0 \end{bmatrix}$, $\boldsymbol{x}' = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, then $\text{CosSim}(\boldsymbol{x}, \boldsymbol{x}') = \frac{\boldsymbol{x} \cdot \boldsymbol{x}\prime}{||x||||\boldsymbol{x}'||} = 1$, $d(\boldsymbol{x}, \boldsymbol{x}') = 9$.

   Two vectors with very low cosine similarity but with very high Eucliddean distance similarity can be chosen as two vectors with opposite direction but same length. For example, $\boldsymbol{x} = \begin{bmatrix} 0.01 \\ 0 \end{bmatrix}$, $\boldsymbol{x}' = \begin{bmatrix} -0.01 \\ 0 \end{bmatrix}$, then $\text{CosSim}(\boldsymbol{x}, \boldsymbol{x}') = \frac{\boldsymbol{x} \cdot \boldsymbol{x}\prime}{||x||||\boldsymbol{x}'||} = -1$, $d(\boldsymbol{x}, \boldsymbol{x}') = 0.02$.

   (b) If the origin of the coordinate system changes, neither of the measures of similarity change. Euclidean distance measures the raw distance between two vectors, and cosine similarity measures the angle between two vectors. Thus, both of the measures are not dependent on the origin of the coordinate system.

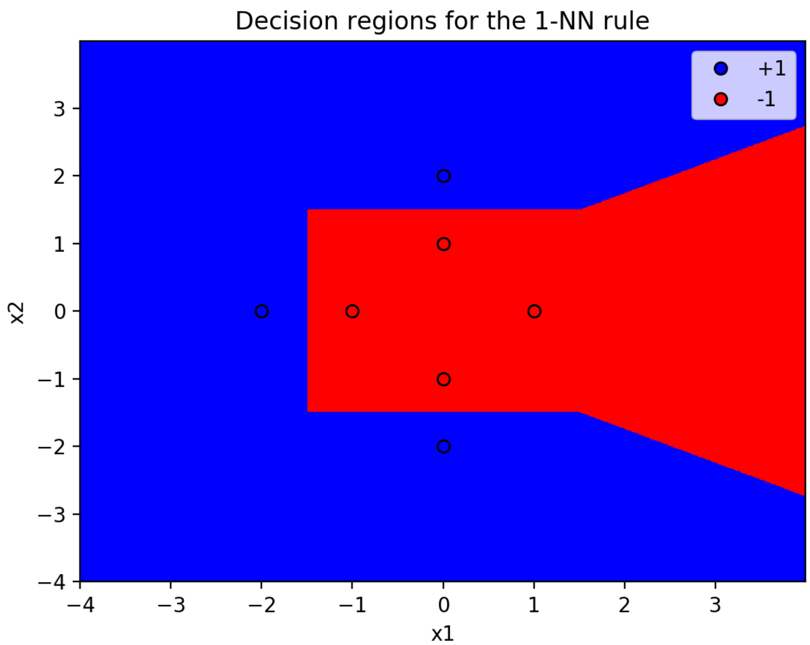
2. Exercise 6.2 in LFD

   By definition, $\pi(x) = \mathbb{P}[y = +1|x]$, $1 - \pi(x) = \mathbb{P}[y = -1|x]$. Then we can obtain $e\big(f(x)\big) = \mathbb{P}[f(x) \neq y] = \begin{cases} 1 - \pi(x) \text{ if } \pi(x) \geq \frac{1}{2} \\ \pi(x) \quad\quad \text{if } \pi(x) < \frac{1}{2} \end{cases}$. It follows that $1 - \pi(x) \leq 0.5$ or $\pi(x) < 0.5$. Since $f(x)$ picks what seems to be the most optimal, then $e\big(f(x)\big) = \mathbb{P}[f(x) \neq y] = \min\{1 - \pi(x),\ \pi(x)\}$.

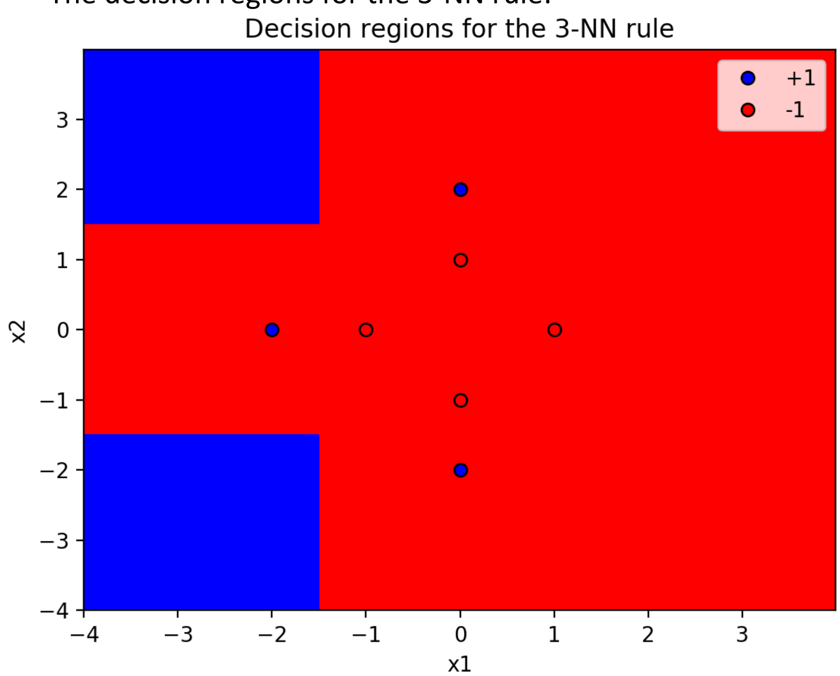
   Also since $f(x)$ picks what seems to be the most optimal, any other hypothesis will produce more errors than $f(x)$ does. For example, when $\pi(x) \geq 0.5$, $f(x) = +1$, and $e\big(f(x)\big) = 1 - \pi(x) < 0.5$. If other hypothesis h disagrees with $f(x)$ then $h(x) = -1$ and $e\big(h(x)\big) = \pi(x) \geq 0.5$. Thus, $e\big(f(x)\big) \leq e(h(x))$ for any other hypothesis h.

3. Problem 6.1 in LFD

(a) The decision regions for the 1-NN rule:

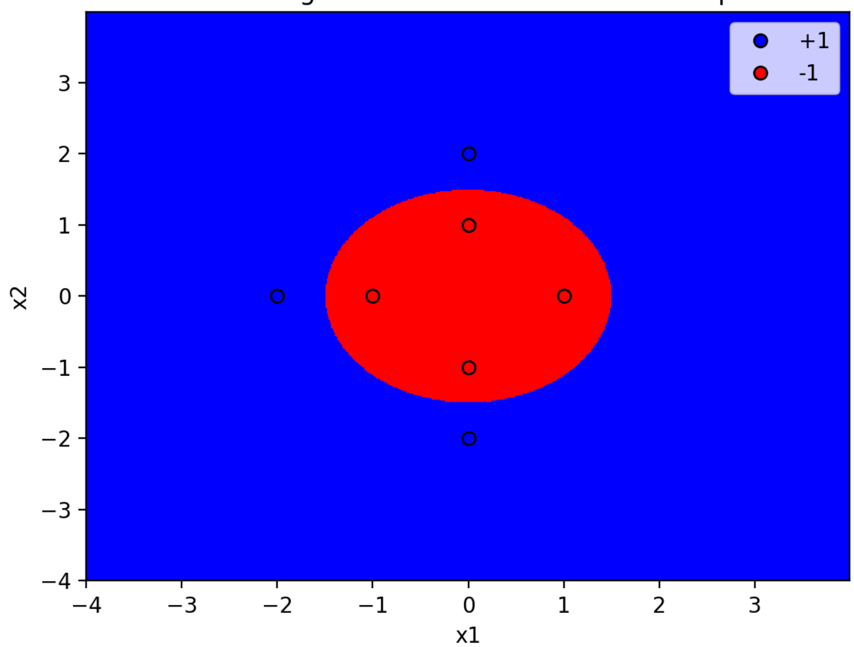


The decision regions for the 3-NN rule:



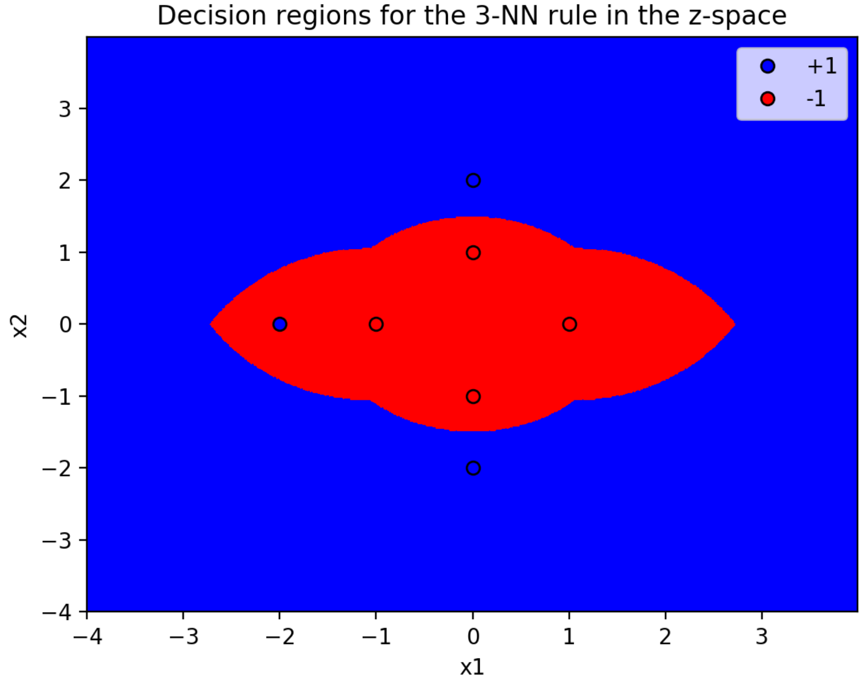
(b) The classification regions for 1-NN rule in the z-space (with non-linear transform):

Decision regions for the 1-NN rule in the z-space



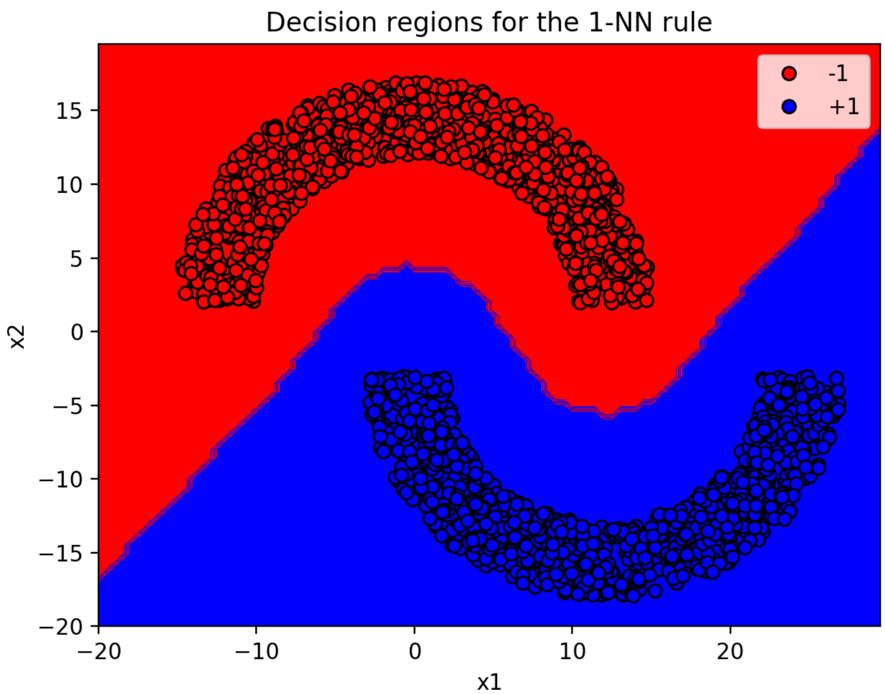
The classification regions for 3-NN rule in the z-space(with non-linear transform):

Decision regions for the 3-NN rule in the z-space

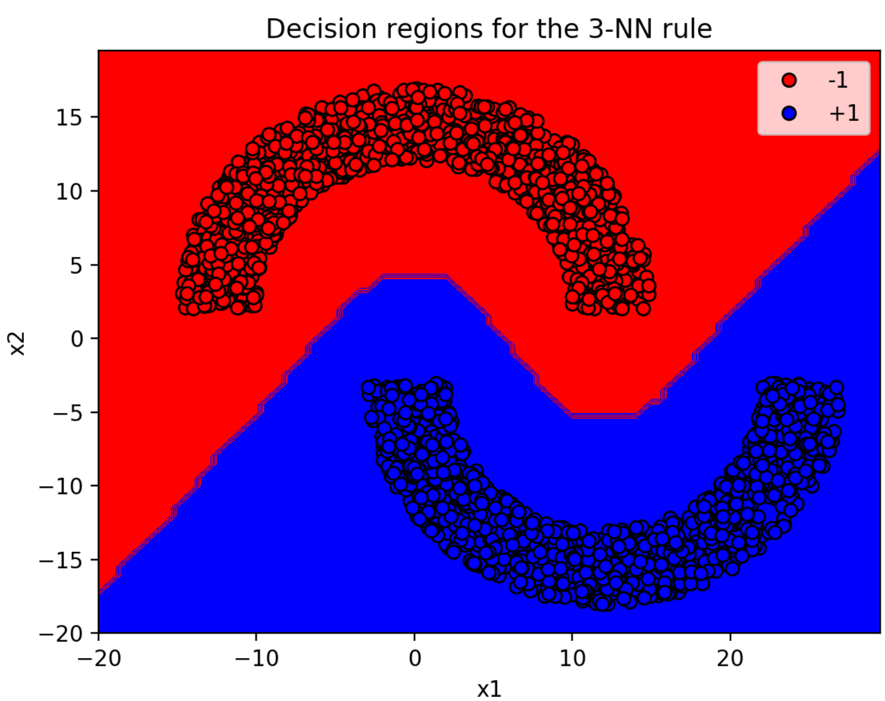


4. Problem 6.4 in LFD

The decision regions for the 1-NN rule:



The decision regions for the 3-NN rule:



5. Problem 6.16 in LFD
    (a) For uniform distribution, the running time for branch and bound is 13.4529459476 seconds; the running time for brute force is 28.6138601303 seconds. Since branch and bound allows us to ignore a large number of points that are far away in the dataset, the running time for branch and bound reduces significantly.
    (b) For the gaussian distribution, the running time for branch and bound is 4.82971596718 seconds; the running time for brute force is 28.1432440281 seconds. The running time for branch and bound reduces more than in part a.
    (c) Since the points are closer to each center under 10 gaussian distribution than in uniform distribution, and the radius of the clusters are much smaller than the distance between clusters, then the partition allows us to disregard most of the clusters. It follows that, for gaussian distribution, the running time in branch and bound is much lower.
    (d) No, the decision to use the branch and bound technique does not depend on the number of test points, but rather the distribution of the points. If the points group into a couple clusters, then branch and bound algorithm is very efficient.