

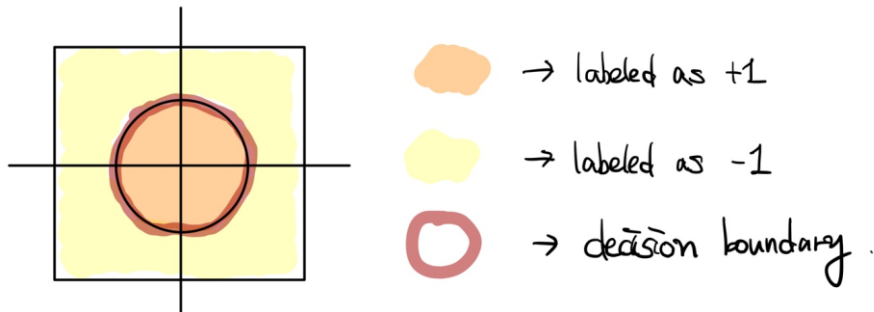
MAS374 Programming HW #3 Report

20180127 Woojin Kim

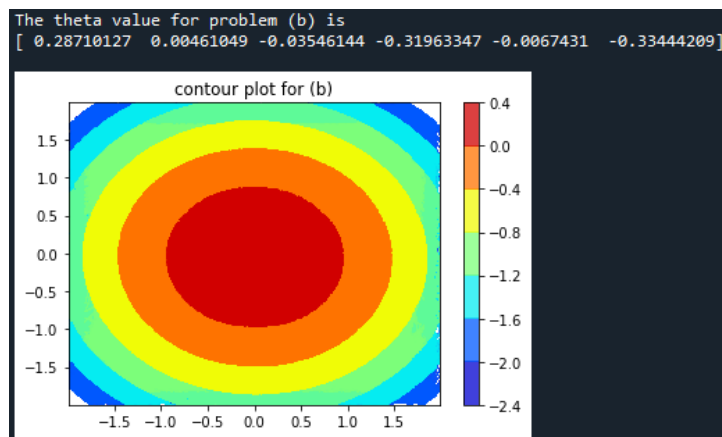
1. Answers to the question in (b)

Question: Make a guess on how the decision boundary would look like, and verify it. Give reasons for your guess and observations.

문제 (b)에서는 모든 샘플이 $[-2, 2] \times [-2, 2]$ 의 사각형 내부에만 존재하고, 중심이 원점이고 반지름이 1인 원 내부에 위치한 샘플만 +1로, 나머지는 -1로 labeling된다. 직관적으로 생각해보면, +1로 labeling되는 조건에 필요한 원의 둘레가 decision boundary라고 볼 수 있다. 이 경계를 중심으로 각 샘플의 label 바뀌기 때문이다. Labelling되는 샘플의 영역과 decision boundary는 다음과 같이 볼 수 있다.



위 그림처럼 decision boundary가 형성되는지 확인하기 위해, 코드를 통해서 계산한 predicted value들이 어떤 분포를 갖고 있는지 출력해보았다. 문제 (a)에서 작성한 `my_lstsq()` 함수를 통해 계산된 theta 값으로 함수 값을 다시 계산해보는 방식을 사용하였다. 이는 코드 실행 시에도 출력된다.

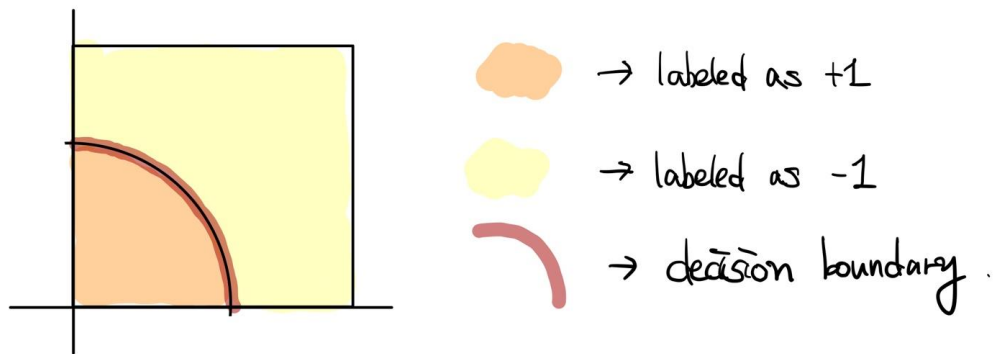


각 샘플에 대응되는 predicted label value가 위 그래프에 표현되어 있다. 주황색-빨간색 경계에서 predicted label value가 0이 되는 것을 확인할 수 있는데, 이 형태 자체는 앞서 이론적으로 추측한 decision boundary의 형태와 유사하나, 범위에 오차가 있다. 하지만, 이 predicted label value들은 polynomial function에 의해서 계산되기 때문에 -1~+1 사이의 다양한 값을 가지기 때문에 실제 decision boundary는 원 둘레로 명확하게 나타나지 않을 수 있다고 생각한다.

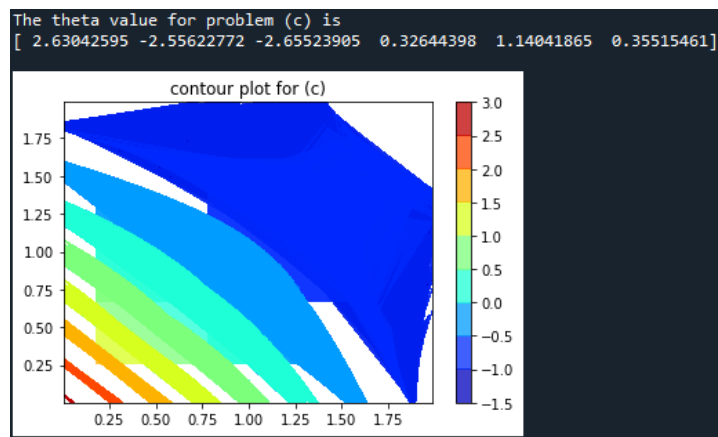
2. Answers to the question in (c)

Question: Make a guess on the decision boundary. What is the optimal polynomial that minimizes (3) in this case? Does the result coincide with your intuition? Explain your answers.

문제 (c)에서는 모든 샘플이 $[0, 2] \times [0, 2]$ 의 사각형 내부에만 존재하고, 중심이 원점이고 반지름이 1인 원과 겹치는 부분에 위치한 샘플만 +1로, 나머지는 -1로 labeling된다. 직관적으로 생각해보면, +1로 labeling되는 조건에 필요한 원의 둘레의 일부가 decision boundary라고 볼 수 있다. 이 경계를 기준으로 각 샘플의 label 바뀌기 때문이다. Labelling되는 샘플의 영역과 decision boundary는 다음과 같이 볼 수 있다.



문제 (b)에 대해서 확인한 것처럼, 이 추측이 옳은 지 확인하고자 아래와 같이 predicted label value를 출력해보았다.



연두색-민트색 경계에서 predicted label value들이 0이 되는 것을 확인할 수 있는데, 여전히 오차가 존재하지만 앞서 세운 추측과 유사한 결과를 보여준다고 생각한다. 위 결과와 대응되는 optimal polynomial은 다음과 같다.

$$f(x) = 2.630 - 2.556x_1 - 2.655x_2 + 0.326x_1^2 + 1.140x_1x_2 + 0.355x_2^2$$