

[Data Science] 데이터 사이언스 개념 - 6.분류문제 — 나무늘보의 개발 블로그

노트북: 첫 번째 노트북

만든 날짜: 2021-01-19 오후 10:02

URL: <https://continuous-development.tistory.com/216?category=833358>

Data Science

[Data Science] 데이터 사이언스 개념 - 6.분류문제

2021. 1. 14. 05:17 수정 삭제 공개

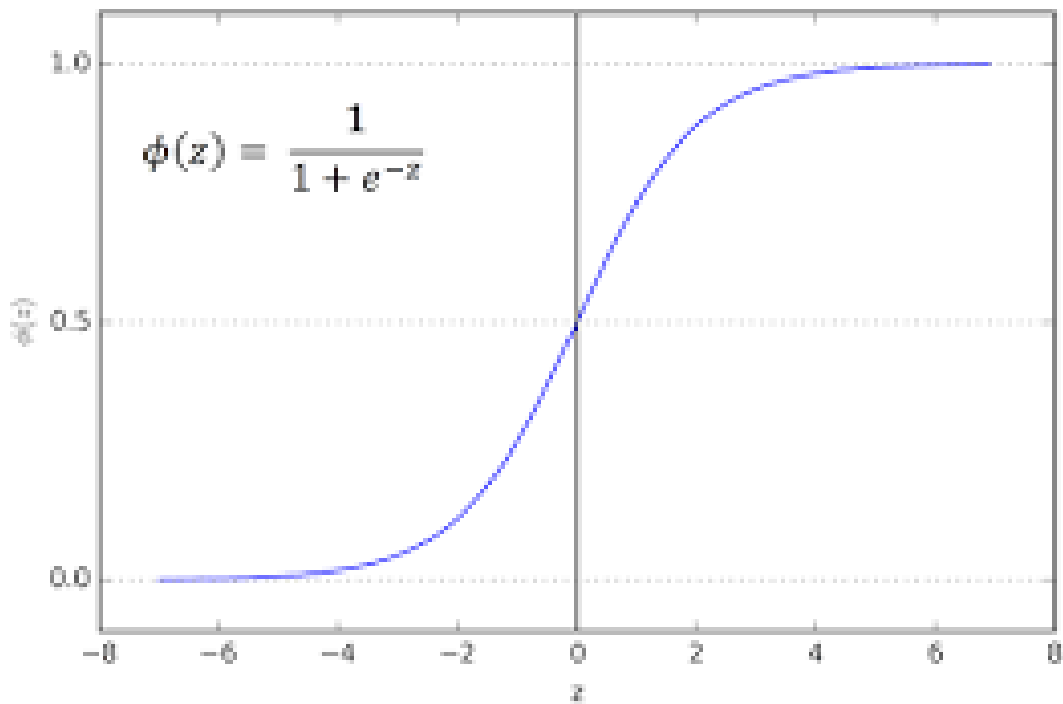


분류문제

1.로지스틱 회귀

로지스틱 회귀 - 회귀를 사용하여 데이터가 어떤 범주에 속할 확률을 0에서 1 사이의 값으로 예측하고 그 확률에 따라 가능성이 더 높은 범주에 속하는 것으로 분류해주는 지도 학습 알고리즘

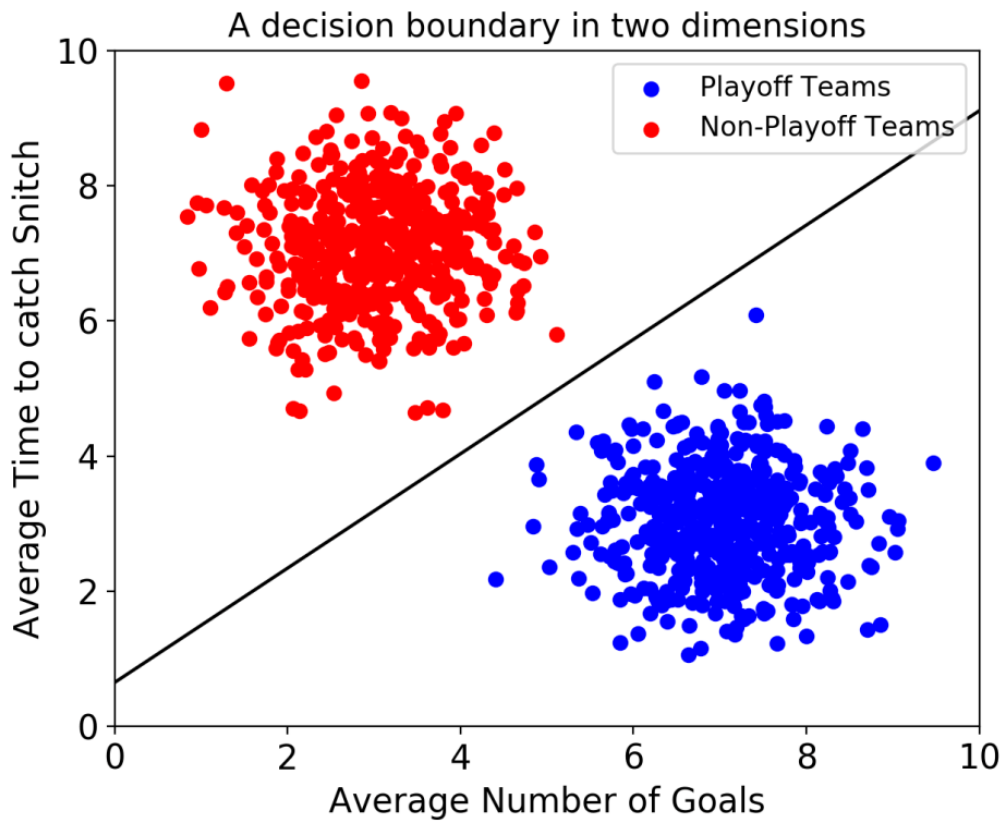
출력 변수가 두가지인 데이터를 자연스럽게 모델링하고자, 0과 1이 나타나는 확률값이 출력되도록 선형회귀 모델을 확장한 것



값이 0~1사이로 들어온다.

추정은 제곱오차를 최소화하지 않고, 우도를 최대화해서 구한다.

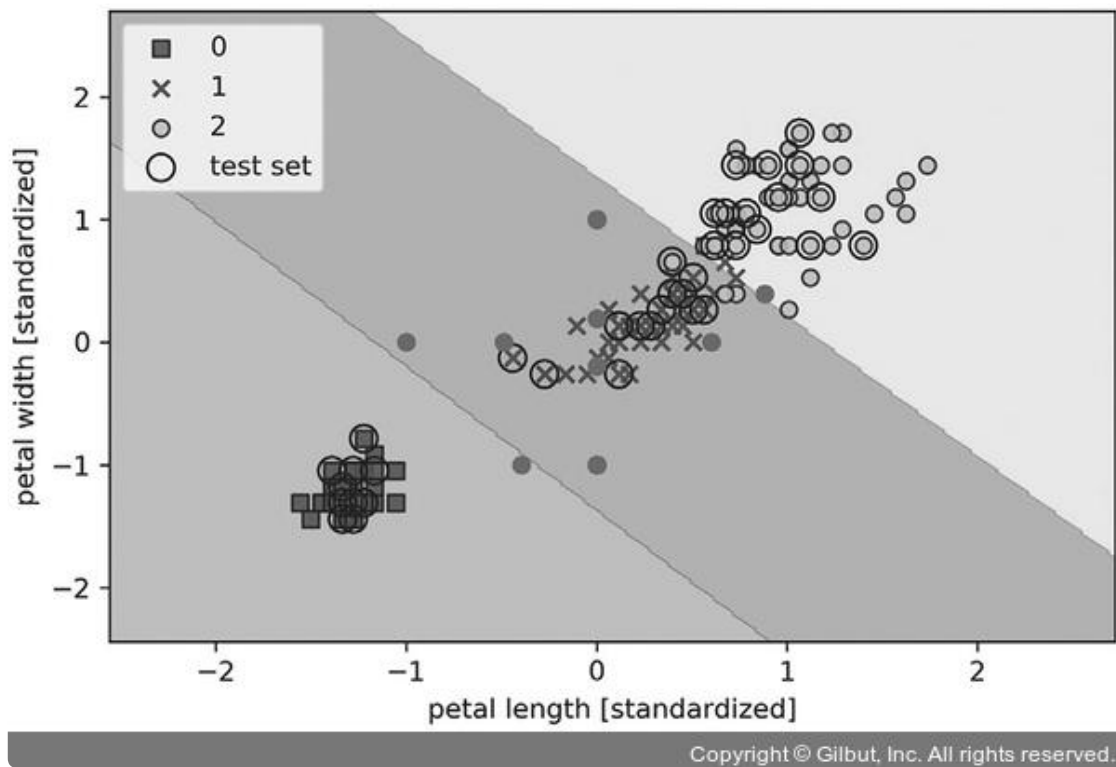
2.선형 서포트 벡터 머신



선을 그은다음에 위면 빨간색 아닐경우 파란색 이런식으로 분류하는게 선형 서포트 벡터 머신의 요점이다.

이 선을 긋는 함수를 최대 마진 초평면이라고 한다.

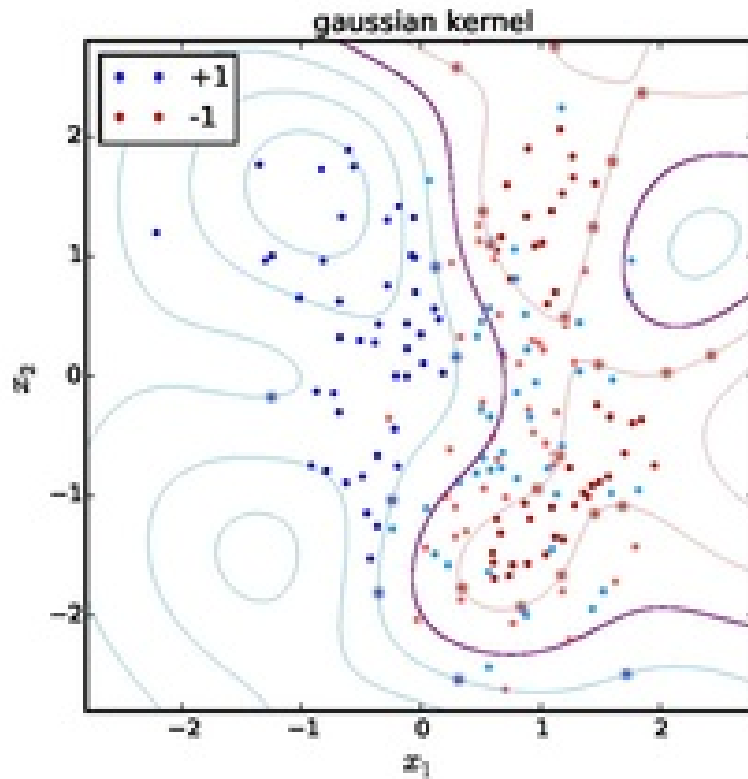
이 선을 긋는걸로 분리가 잘안될때는 슬랙변수라는 것을 정의함으로써 추정한다.



분리 초평면을 정의할 때 사용하는 몇 개의 점을 서포트 벡터라고 부른다. 이 서포트 벡터를 잘 활용하면 서포트 벡터 머신을 비선형으로 확장 할 수 있다.

3.서포트 벡터 머신

특징량의 관계가 비선형 일때는 선형 서포트 벡터 머신으로 풀 수 없다. 이럴 때 커널함수로 불리는 비선형적으로 2개의 벡터 거리를 측정하는 함수로 치환한다.



그 결과 위와같이 비 선형적인 서포트 벡터 머신을 구현 할 수있다.

본 내용은 그림으로 배우는 DataScience 데이터 과학을 참고한 내용입니다

'Data Science' 카테고리의 다른 글

[Data Science] 데이터 사이언스 개념 - 8.토픽 모델 / 네트워크 분석

[Data Science] 데이터 사이언스 개념 - 7.비지도 학습

[Data Science] 데이터 사이언스 개념 - 6.분류문제

[Data Science] 데이터 사이언스 개념 - 5.앙상블 학습

[Data Science] 데이터 사이언스 개념 - 4.회귀 모델

[Data Science] 데이터 사이언스 개념 - 3.과적합과 모델 선택

로지스틱 회귀

분류문제

서포트 벡터 머신

선형 서포트 벡터 머신



나아무늘보

혼자 끄적끄적하는 블로그 입니다.