

Assignment (이용훈 교수님 연구실)

담당조교:우인희 (dlsgml107@kaist.ac.kr)

-데이터 셋 설명

Data set: Iris dataset

Input features: ['sepal length', 'sepal width', 'petal length', 'petal width']

Target features: ['Virginica', 'Setosa', 'Versicolor'] Classification

-과제설명

Data Science 입문용으로 많이 사용되는 Iris(붓꽃) Classification에 관한 문제입니다.

꽃받침(sepal)의 길이와 너비, 꽃잎(petal)의 길이와 너비를 이용하여 붓꽃의 세가지 품종을 분류하게 됩니다.

2.1에서는 petal width만을 이용하여 'Virginica vs all' 을 구분하는 logistic regression, 2.2에서는 petal width와 petal length를 이용하여 'Virginica vs all', 'Versicolor vs all', 'Setosa vs all' 을 구분하는 logistic regression, 2.3에서는 petal width와 petal length를 이용하여 세가지 품종을 동시에 구분해보는 multinomial regression을 수행하게 됩니다.

-제출물

< 코드 >

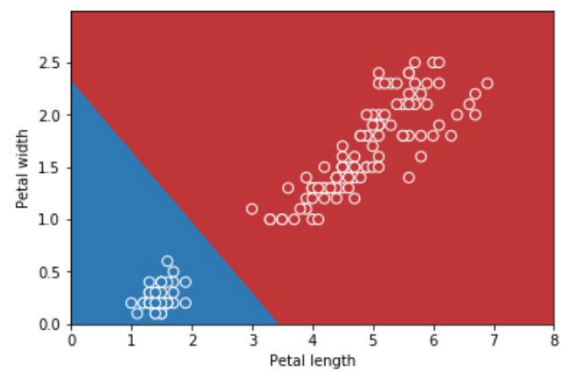
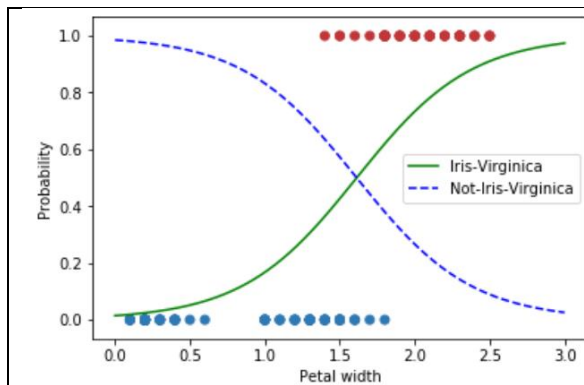
'3. Softmax regression' 부분의 빈칸 3개를 채워주시면 됩니다. 주어진 코드와 7월2일 진행된 이용훈 교수님 강의자료를 참고하시면 됩니다.

< 보고서 >

위 '3. Softmax regression' 코드까지 완성을 하시면, 코드에는 아래와 같이 총 5가지 종류의 그림이 존재하게 됩니다. 이 5가지 그림들에 각각 적용된 알고리즘을 토대로 각자의 분석을 작성해주시면 됩니다.

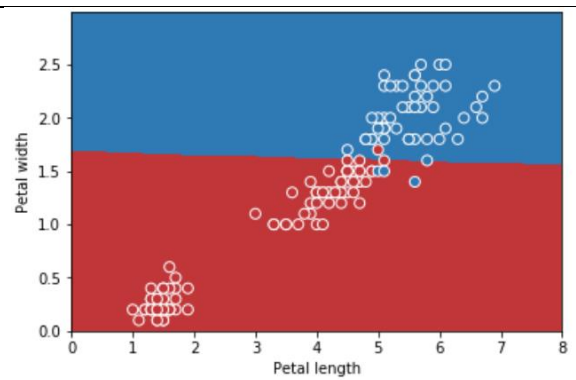
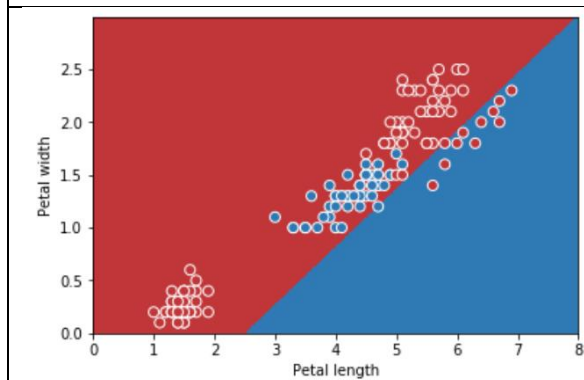
-채점

채점은 코드 완성(2점)과 보고서(8점)으로 채점이 됩니다. 보고서는 양식이나 분량 제한이 없으니, 자유롭게 작성하시기 바랍니다.



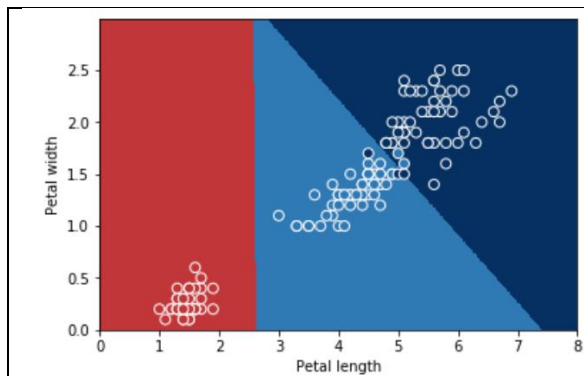
2.1 2-class logistic regression using a single feature

2.2 case 1) Virginica(blue) vs. all(red)



2.2 case 2) Versicolor(blue) vs. all(red)

2.2 case 3) Setosa(blue) vs. all(red)



**3. Softmax regression
(multinomial logistic regression)**