

문제1.

37페이지의 original scale의 decision boundary를 어떻게 계산할 수 있는지 설명하고,  
이를 사용해서 38페이지와 같은 그래프를 그리시오.

문제2.

입력변수  $x_1$ (exam1),  $x_2$ (exam2)에  $x_1^2$ ,  $x_2^2$ ,  $x_1x_2$  을 입력변수로 추가한 후 logistic regression 모델을 학습하여라.

학습을 수행하여 34페이지와 같은 learning curve를 그리라.

cost 값을 입력변수 추가하기 전과 비교하여 보아라.

문제3.

문제2에서 학습한 모델의 decision boundary의 식을 표현하고

38페이지와 같은 형식의 그래프로 그려보아라.

문제4.

[Kaggle Titanic Dataset](#)을 학습하여 생존 여부를 판단하는 logistic model을 만들라. (train.csv 파일 사용) 강의 및 실습에서 연습한 것처럼 logistic regression은 패키지를 사용하지 않고 직접 구현하고 학습하는 알고리즘을 수행해야한다.

입력 변수는 전체 변수를 사용할 필요는 없고 사용할 수 있는 변수들만 선택해서 사용하면 된다.  
예측 대상 변수는 생존 여부를 나타내는 **Survived**이다.

Variable	Definition	Key
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Sex	
Age	Age in years	
sibsp	# of siblings / spouses aboard the Titanic	
parch	# of parents / children aboard the Titanic	
ticket	Ticket number	
fare	Passenger fare	
cabin	Cabin number	
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton

#### 문제5.

문제 4에서 학습한 모델을 이용해서 **test.csv** 파일의 승객들에 대한 생존 여부를 예측하여라. 예측 결과를 **gender\_submission.csv**와 같이 작성하여 kaggle에 제출하여 성능을 평가해보아라. **train.csv**의 예측성능과 비교해보고 overfitting이거나 underfitting인지 설명해보자