

```
In [ ]: import pandas as pd
import numpy as np
```

```
In [ ]: train_url = "http://s3.amazonaws.com/assets.datacamp.com/course/Kaggle/train.csv"
train = pd.read_csv(train_url) #training set
test_url = "http://s3.amazonaws.com/assets.datacamp.com/course/Kaggle/test.csv"
test = pd.read_csv(test_url) #test set
print(train.head())
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	

	Name	Sex	Age	SibSp	\
0	Braund, Mr. Owen Harris	male	22.0	1	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	
2	Heikkinen, Miss. Laina	female	26.0	0	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	
4	Allen, Mr. William Henry	male	35.0	0	

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S

```
In [ ]: print("T8) median is "+str(train["Age"].median()))
train["Age"] = train["Age"].fillna(train["Age"].median())
```

T8) median is 28.0

```
In [ ]: train.loc[train["Embarked"] == "S", "Embarked"] = 0
train.loc[train["Embarked"] == "C", "Embarked"] = 1
train.loc[train["Embarked"] == "Q", "Embarked"] = 2
train["Embarked"] = train["Embarked"].fillna(train["Embarked"].mode()[0])

print("T9)mode is "+str(train["Embarked"].mode()[0]))
```

T9)mode is 0

```
In [ ]: train.loc[train["Sex"] == "female", "Sex"] = 1
train.loc[train["Sex"] == "male", "Sex"] = 0
```

```
In [ ]: def h(x):
    return 1/(1+np.exp(-x))

def dj_dtheta(theta,X,Y,j):
    result=0
    for i in range(np.size(Y)):
        error=Y[i]-h(np.dot(theta,X[i]))
        result+=(error*X[i,j])
    return result

def grad_des(X,Y,theta):
```

```

lnr=0.0001
epoch=50000
for i in range(epoch):
    if(i%5000==0):
        print(i,theta)
        new_theta=[]
        for j in range(X.shape[1]):
            a=theta[j]+lnr*dj_dtheta(theta,X,Y,j)
            new_theta.append(a)
        theta=np.array(new_theta)

    return theta

```

```

In [ ]: data = np.array(train[["Pclass","Sex","Age","Embarked"]].values, dtype = float)
data=np.insert(data, 0, 1, axis=1)
normalizer=max(data[:,[3]])
data[:,[3]]=data[:,[3]]/normalizer
print(data)

```

```

[[1.    3.    0.    0.275 0.    ]
 [1.    1.    1.    0.475 1.    ]
 [1.    3.    1.    0.325 0.    ]
 ...
 [1.    3.    1.    0.35  0.    ]
 [1.    1.    0.    0.325 1.    ]
 [1.    3.    0.    0.4   2.    ]]

```

```

In [ ]: theta=np.array([0,0,0,0,0])
Y=np.array(train["Survived"].values, dtype = float)
theta=grad_des(data,Y,theta)

print("theta",theta)

```

```

0 [0 0 0 0 0]
5000 [ 1.48197875 -1.08577968  2.57184137 -1.77383924  0.31735197]
10000 [ 1.87108062 -1.15823869  2.57182532 -2.38374666  0.3195396 ]
15000 [ 2.00231865 -1.18313964  2.57389922 -2.58985458  0.32041422]
20000 [ 2.04750068 -1.19176484  2.57484796 -2.66085443  0.32073113]
25000 [ 2.06316981 -1.19476226  2.57520473 -2.68548207  0.32084291]
30000 [ 2.0686178  -1.19580518  2.5753321  -2.69404541  0.32088201]
35000 [ 2.07051369 -1.1961682  2.57537683 -2.69702552  0.32089564]
40000 [ 2.07117367 -1.19629459  2.57539245 -2.69806293  0.32090039]
45000 [ 2.07140344 -1.19633859  2.57539789 -2.6984241  0.32090204]
theta [ 2.07148343 -1.19635391  2.57539979 -2.69854984  0.32090262]

```

```

In [ ]: def isSurvived(theta,X,m):
    result=[]
    for i in range(m):
        ans=h(np.dot(theta,X[i]))
        if(ans>=0.5):
            ans=1
        else:
            ans=0
        result.append(ans)
    return result

dic=dict()

```

```
dic["PassengerId"]=np.array(test["PassengerId"].values, dtype = int)
test.loc[test["Sex"] == "female", "Sex"] = 1
test.loc[test["Sex"] == "male", "Sex"] = 0
test.loc[test["Embarked"] == "S", "Embarked"] = 0
test.loc[test["Embarked"] == "C", "Embarked"] = 1
test.loc[test["Embarked"] == "Q", "Embarked"] = 2
data=np.array(test[["Pclass", "Sex", "Age", "Embarked"]].values, dtype = float)
data=np.insert(data, 0, 1, axis=1)
data[:,3]=data[:,3]/normalizer
dic["Survived"]=isSurvived(theta,data,418)
df=pd.DataFrame(dic)
df.to_csv("OT10_11.csv",index=False)
```