

- Klassifikation mit logistischer Regression -

Beantworten Sie folgende Fragen zum Titanic-Datensatz über Passagiere, die überlebt bzw. nicht überlebt haben (überleben ist wahr bzw. falsch).

1. Beschreiben Sie die Größen, die aus dem Training (*engl. fit*) des Modells in der logistischen Regression als Funktion des Geschlechts resultieren.

```
#Train the model
log_model.fit(X = pd.DataFrame(encoded_sex),
              y = titanic_train["Survived"])
```

2. Beurteilen Sie die Aussagekraft der logistischen Regression für die Berechnung der Überlebenswahrscheinlichkeit als Funktion des Geschlechts anhand der Ergebnisse der Kreuztabelle:

```
# Make predictions
preds = log_model.predict_proba(X= pd.DataFrame(encoded_sex))
preds = pd.DataFrame(preds)
preds.columns = ["Death_prob", "Survival_prob"]

# Generate table of predictions vs Sex
pd.crosstab(titanic_train["Sex"], preds.loc[:, "Survival_prob"])
```

Survival_prob	0.19314664603456402	0.7342488525817924
Sex		
female	0	314
male	577	0

Kreuztabelle (*engl. crosstab*) für die Überlebenswahrscheinlichkeit als Funktion des Geschlechts

3. Interpretieren Sie die Ergebnisse der Wahrheitsmatrix, die aus dem Training eines Modells mit logistischer Regression von insgesamt vier Variablen entstanden ist:

Survived	0	1
row_0		
0	458	89
1	91	253

Wahrheitsmatrix