



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

Carlos J. Costa

REGRESSIONS



Regression

- Is a set of statistical processes for estimating the relationships among variables.
- Dependent variable, outcome variable, target
- Independent variables, predictor, covariates, or features



Regression

- simple regression/multivariate regression

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

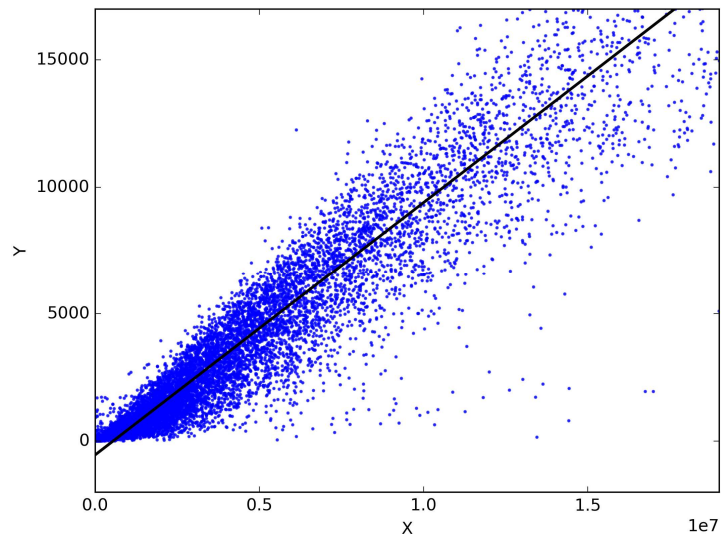
$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i.$$



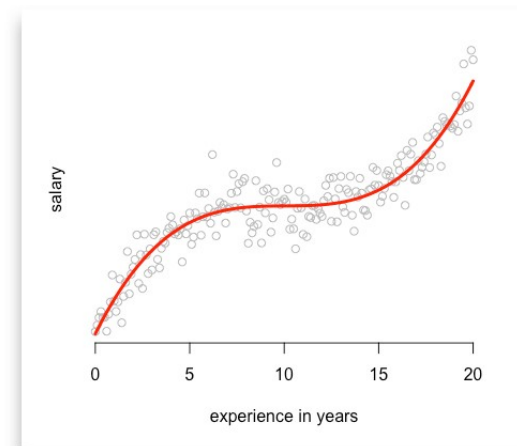
Regression

- .Linear/non linear

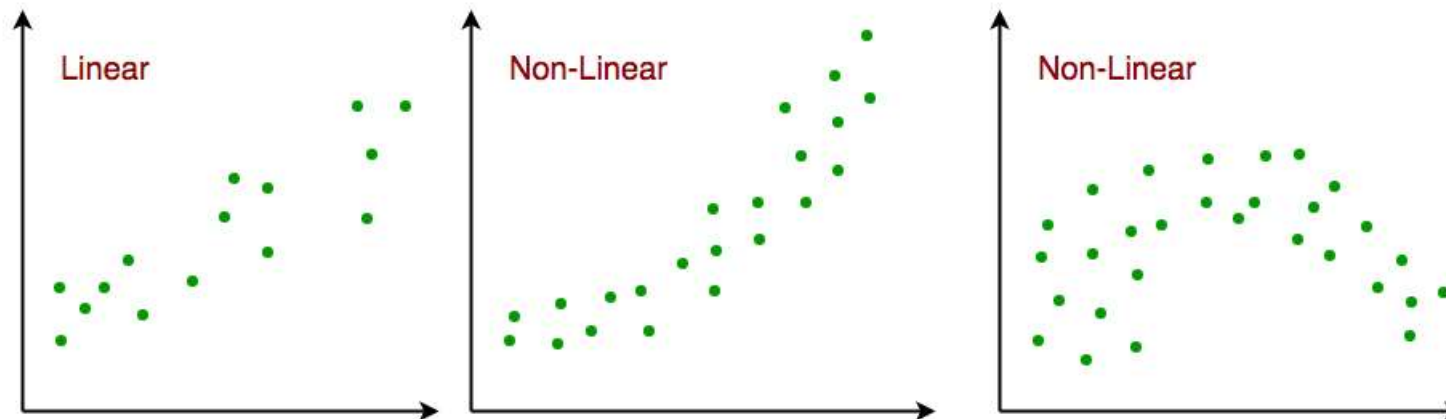
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n.$$



$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \quad i = 1, \dots, n.$$



Regression



Regression

```
import statsmodels.api as sm
X = sm.add_constant(X)
result = sm.OLS(y,X).fit()
result.summary()
```

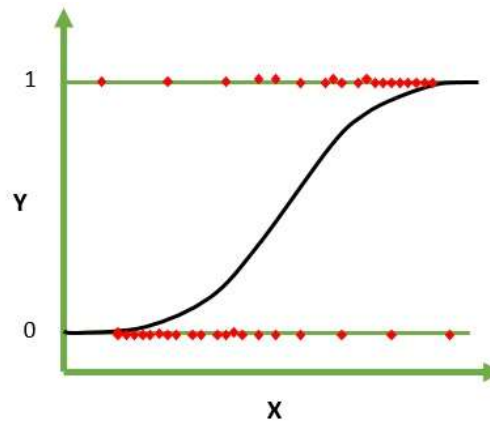
Regression

```
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.3)
#fit intercept is by default: LinearRegression(fit_intercept = False)
linear_regressor = LinearRegression()
results=linear_regressor.fit(X_train, y_train)
y_test_pred=results.predict(X_test)
y_train_pred=results.predict(X_train)
print("R2 score =", round(sm.r2_score(y_test, y_test_pred), 2))
```



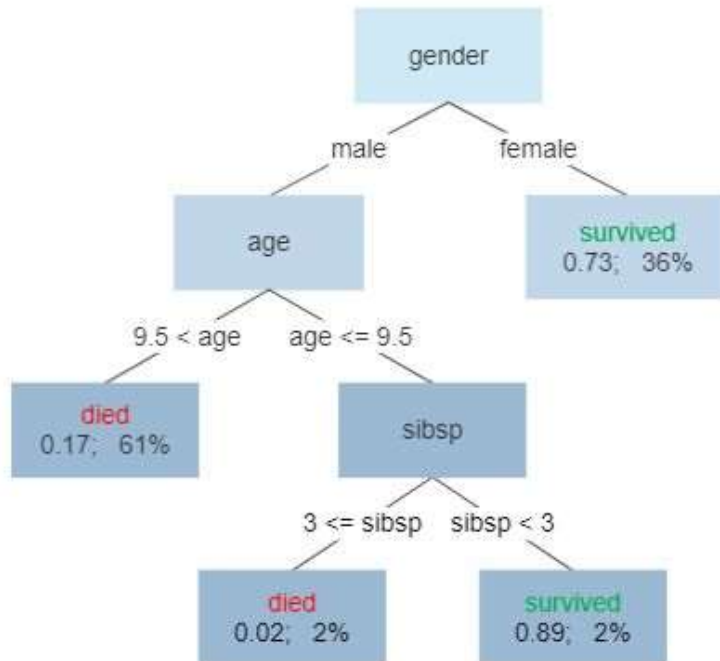
Logistics Regression

- A regression that having binary dependent variable
- in its basic form, uses a logistic function to model a binary dependent variable



Decision Tree

Survival of passengers on the Titanic



- Decision tree builds classification or regression models in the form of a tree structure.
- It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed.
- The final result is a tree with decision nodes and leaf nodes.

Random Forest

- are an ensemble learning method for classification, regression and other tasks
- operates by constructing a multitude of decision trees at training time
- outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

