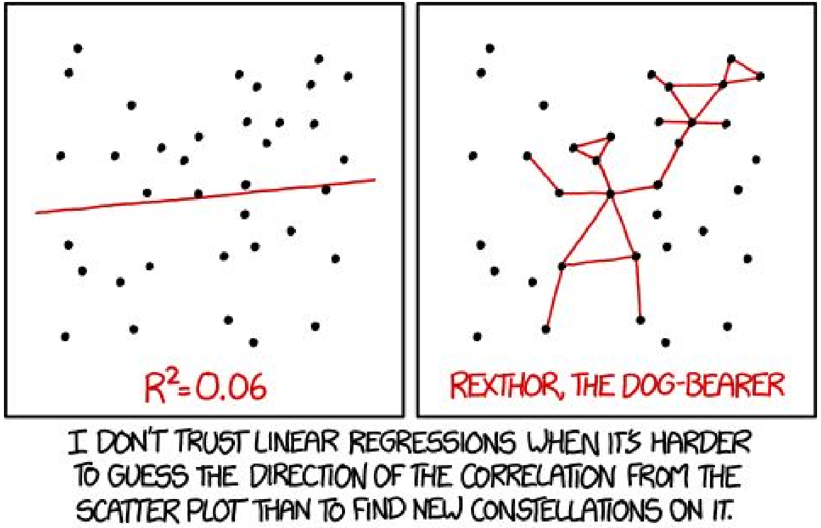
Linear Regression (1)



Seoul AI Meetup

Martin Kersner, 2017/10/14

What Is Linear Regression?

Linear approach for modeling of relationship between

- a scalar dependent variable (denoted as y)
- and one or more independent variables (denoted as X).

- Simple Linear Regression
- Multivariate Linear Regression

Classification - targets are nominal values.
Regression - targets are numeric and continuous.

Linear Regression Methods

Introduction

- **Ordinary Least Squares Regression**

Basic methods

- Polynomial Regression
- Locally Weighted Linear Regression

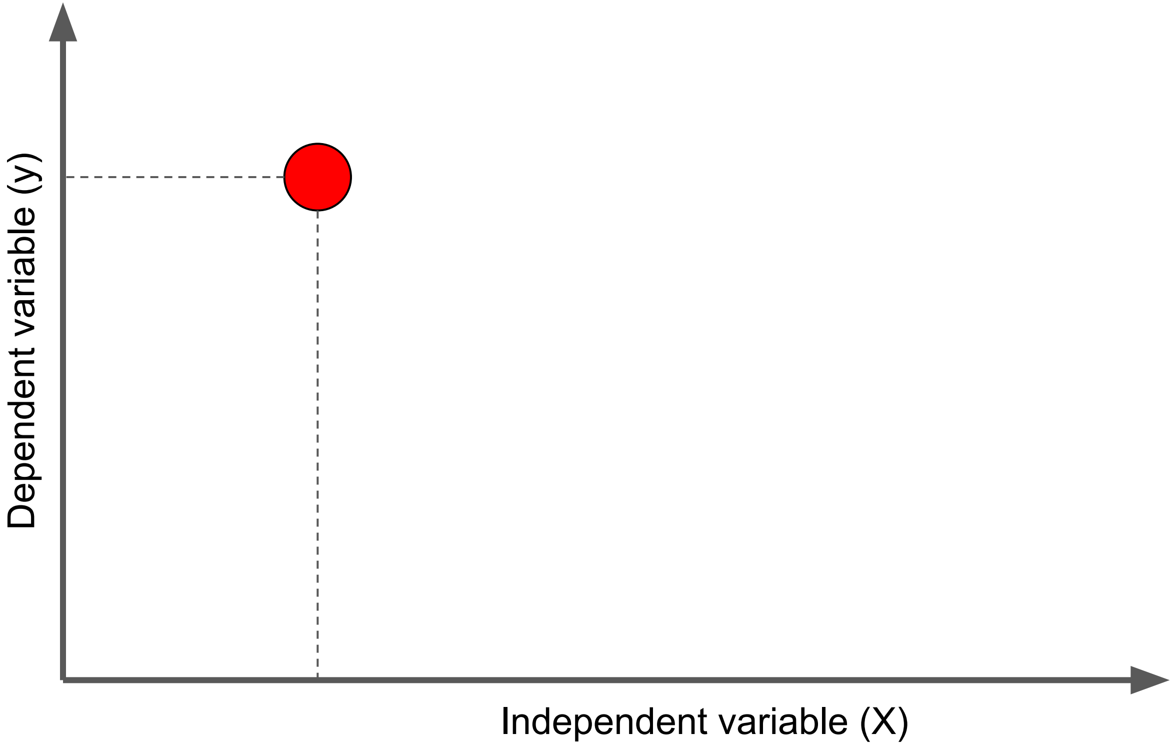
Shrinkage methods

- Ridge Regression
- Lasso
- Forward Stagewise Regression

Others

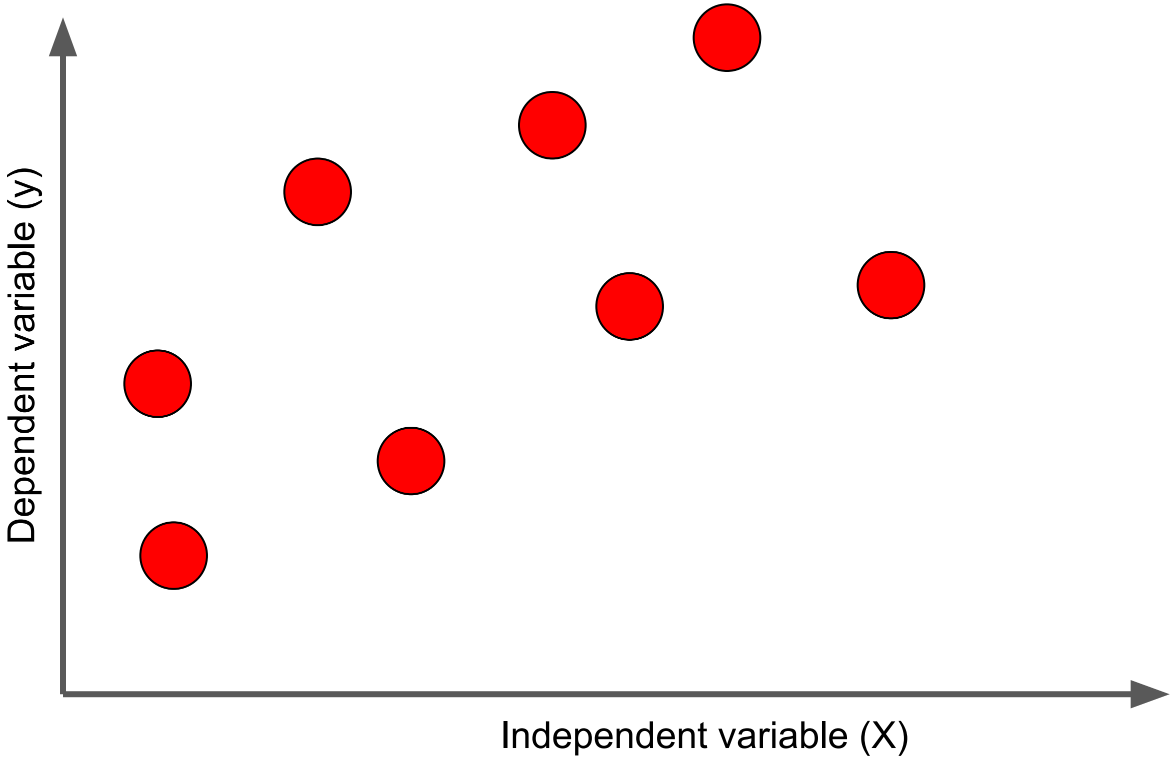
- TensorFlow Lattice

Ordinary Least Squares

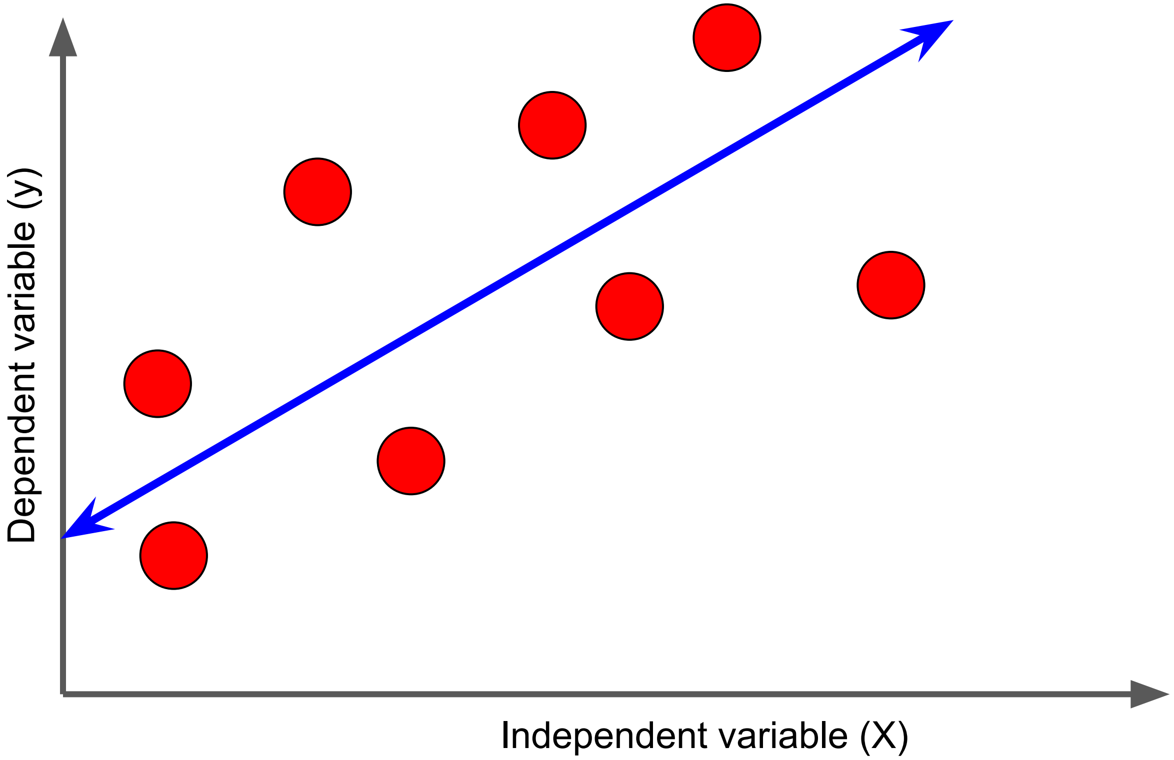


Ordinary Least Squares

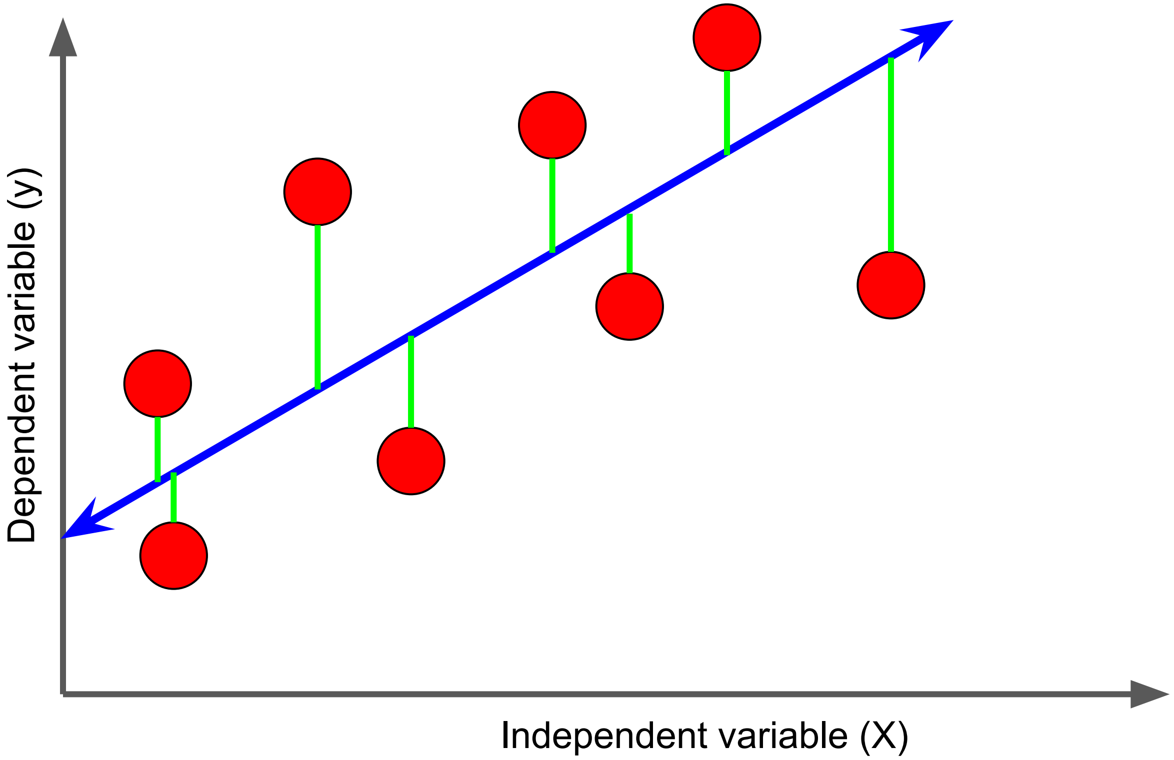
Intuition



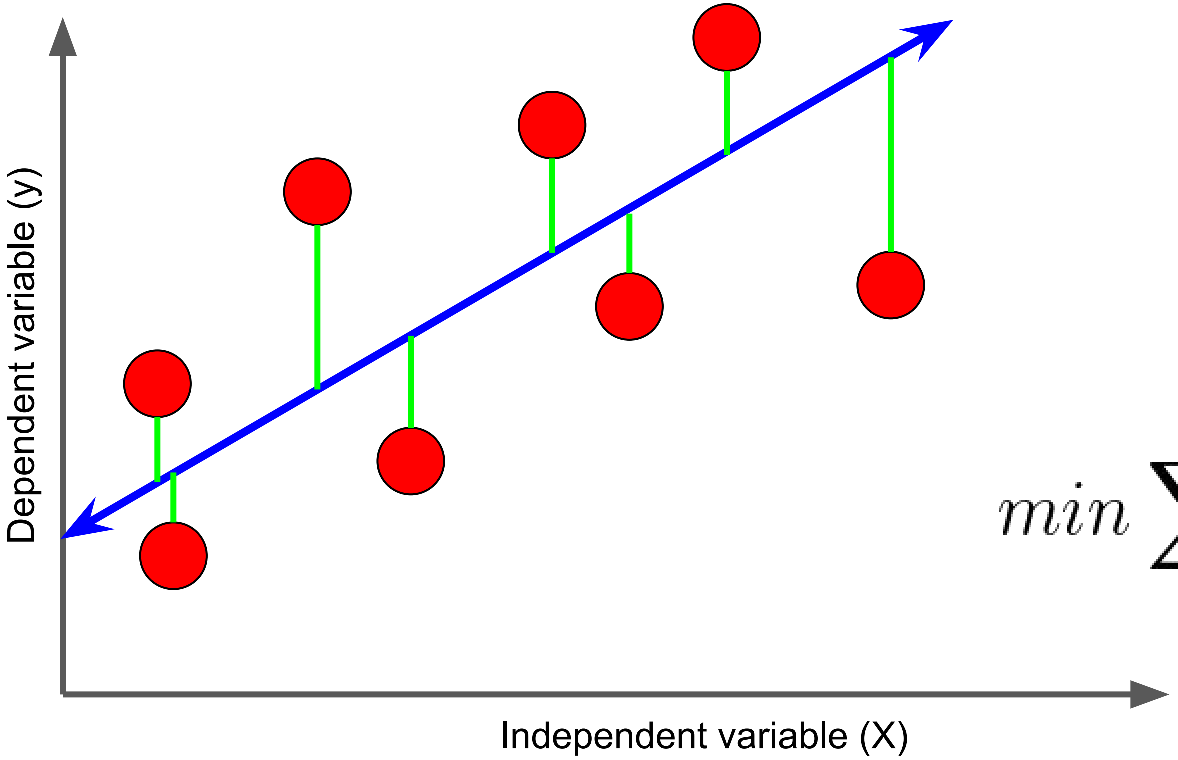
Ordinary Least Squares

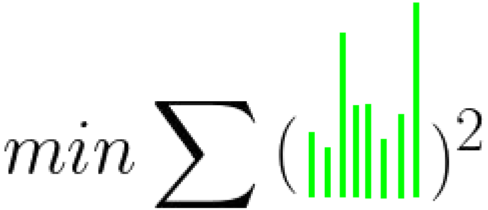


Ordinary Least Squares



Ordinary Least Squares





Ordinary Least Squares

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$y_i = \boxed{\beta_0 1} + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$$

bias

$$where \;\; i = 1, \ldots, n$$

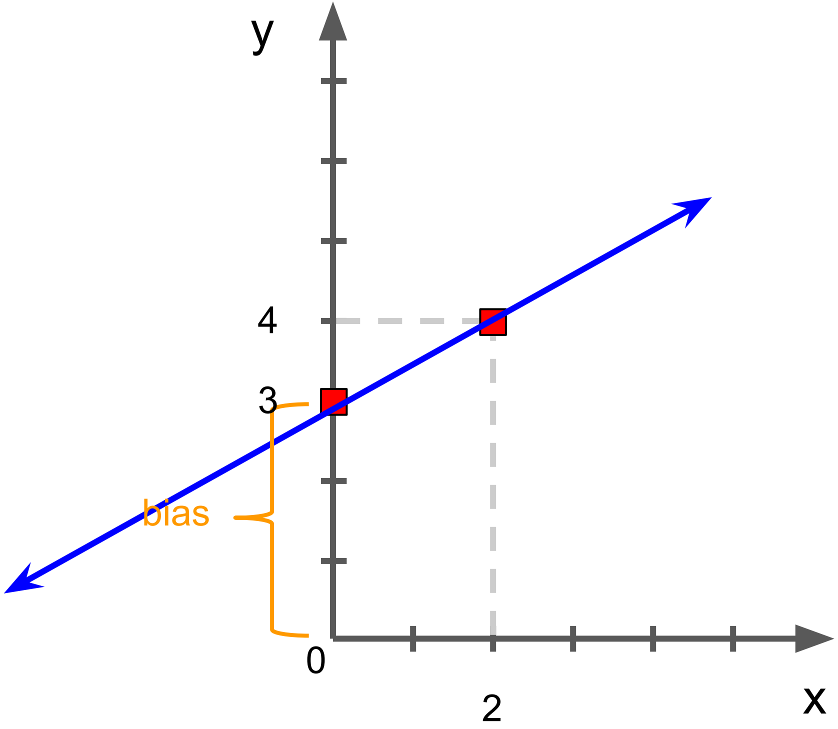
p data dimensionality

n number of data points

ε error variable

Bias

Without bias every line goes through $[0, 0]$.



$$y = \boxed{3} + x \cdot \frac{1}{2}$$

$$3 = \boxed{3} + 0 \cdot \frac{1}{2}$$

$$4 = \boxed{3} + 2 \cdot \frac{1}{2}$$

Ordinary Least Squares

Mathematical Interpretation

The goal is to minimize the residual sum of squares.

$$min \sum_{i=1}^{n} (y_i - \boldsymbol{x_i}^\top \boldsymbol{\beta})^2$$

$$\boxed{(\boldsymbol{y} - \boldsymbol{X\beta})^\top (\boldsymbol{y} - \boldsymbol{X\beta})}$$

$$\frac{\partial}{\partial \boldsymbol{\beta}} (\boldsymbol{y} - \boldsymbol{X\beta})^\top (\boldsymbol{y} - \boldsymbol{X\beta}) \ = \ \boldsymbol{X}^\top (\boldsymbol{y} - \boldsymbol{X\beta})$$

https://web.stanford.edu/~mrosenfe/soc_meth_proj3/matrix_OLS_NYU_notes.pdf

Ordinary Least Squares

Mathematical Interpretation

$$\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = 0$$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$$

Ordinary Least Squares


```
def fit(X, y):
    XtX = X.T * X

    # matrix must be nonsingular
    assert(np.linalg.det(XtX) != 0.0)

    return XtX.I * X.T * y
```

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

Ordinary Least Squares


```python
def predict(X, beta_hat):
    return X * beta_hat
```

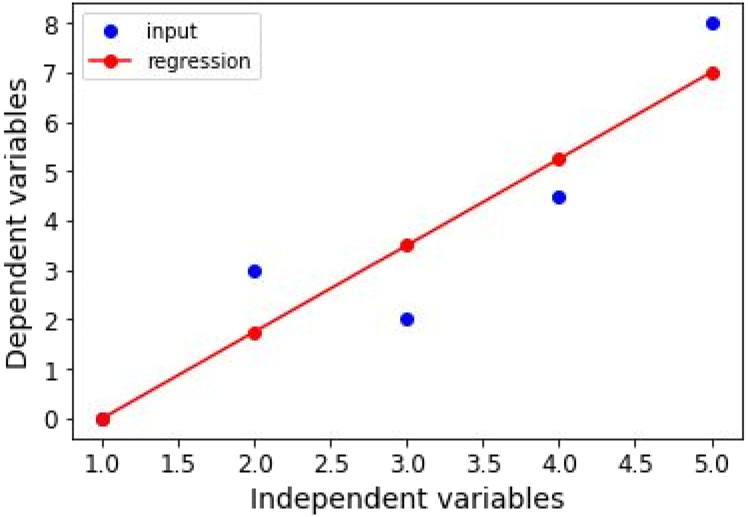
$$\hat{y} = \boldsymbol{X}\hat{\beta}$$

Ordinary Least Squares


```
X = np.matrix([[1,1], [1,2], [1,3], [1,4], [1,5]])
y = np.matrix([[0], [3], [2], [4.5], [8]])

beta_hat = fit(X, y)
# y_hat = predict(X, beta_hat)
```

Ordinary Least Squares



Evaluation

R^2

Coefficient of determination, $\mathbf{R^2}$

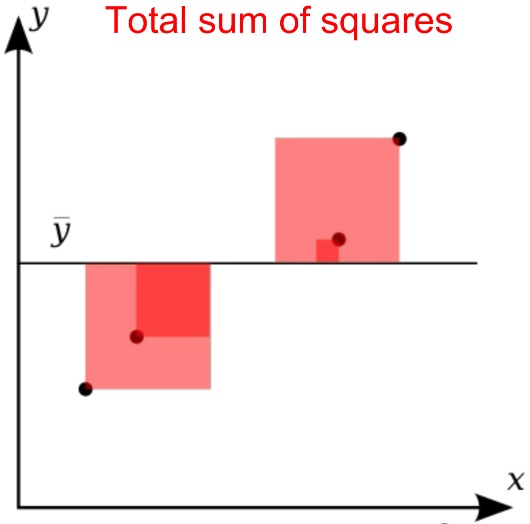
- best possible score is 1.0
- 0.0 when always predicting mean of dependent variables

Measure of how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model.

Evaluation

Coefficient of determination, $\mathbf{R^2}$

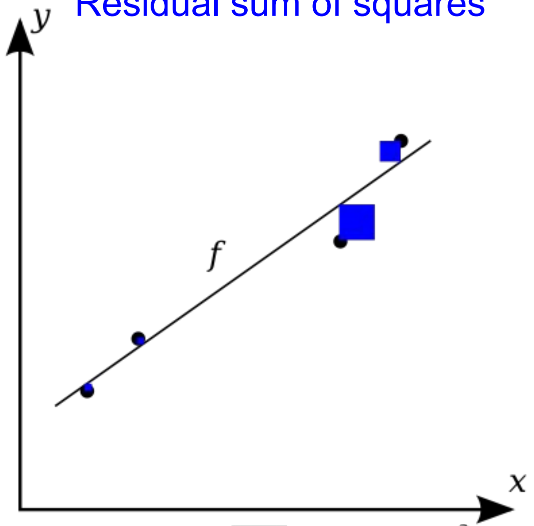
Total sum of squares



$$SS_{tot} = \sum_i (y_i - \bar{y})^2$$

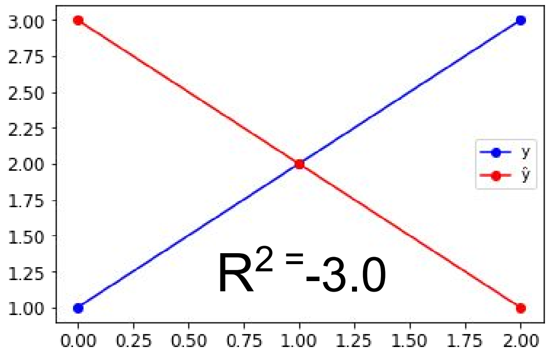
$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

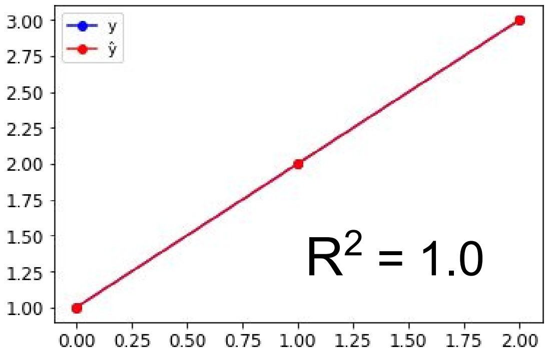
Residual sum of squares



$$SS_{res} = \sum_i (y_i - \hat{y}_i)^2$$

Evaluation





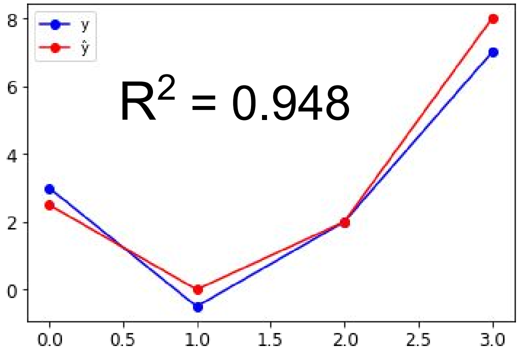
R^2 scores:

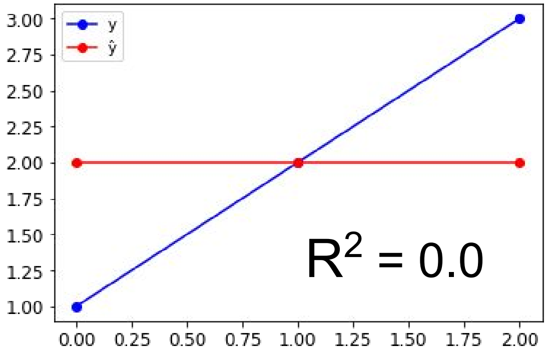
0.948

1.0

-3.0

0.0

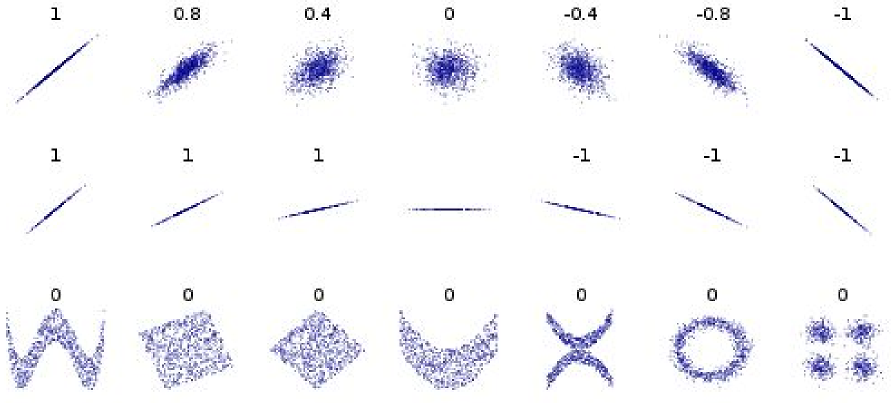




Evaluation

Pearson Correlation Coefficient

- Sensitive only to a linear relationship between two variables
- Values between -1 and 1



$$\mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Correlation reflects the noisiness and direction of a linear relationship (top row), but not the slope of that relationship (middle).

Ordinary Least Squares

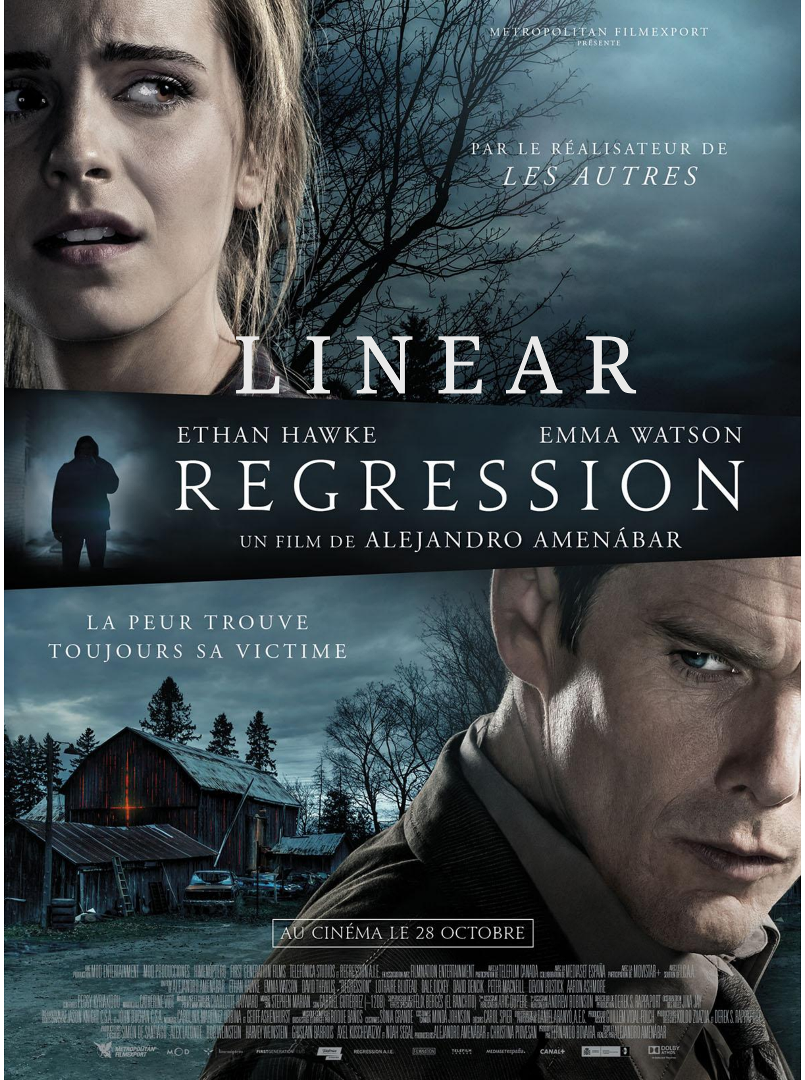
scikit-learn

sklearn.linear_model.LinearRegression
sklearn.metrics.r2_score

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

lr = LinearRegression()
lr.fit(X, y)
y_hat = lr.predict(X)

r2_score(y, y_hat)
```


METROPOLITAN FILMEXPORT
PRÉSENTE
PAR LE RÉALISATEUR DE
LES AUTRES
LINEAR
ETHAN HAWKE
EMMA WATSON
REGRESSION
UN FILM DE ALEJANDRO AMENÁBAR
LA PEUR TROUVE
TOUJOURS SA VICTIME
AU CINÉMA LE 28 OCTOBRE

References

Regression

https://en.wikipedia.org/wiki/Regression_analysis

Linear Regression

https://en.wikipedia.org/wiki/Linear_regression

https://en.wikipedia.org/wiki/Linear_predictor_function

R^2

https://en.wikipedia.org/wiki/Coefficient_of_determination

https://en.wikipedia.org/wiki/Correlation_and_dependence

Nonsingular matrix

- n-by-n square matrix A is called nonsingular if there exists an n-by-n square matrix B such that

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$$

- A square matrix is singular if and only if its determinant is 0.
- If A is a square matrix, then A is invertible if and only if A has rank n (that is, A has full rank).
- Rank of matrix A equals to maximal number of linearly independent columns of matrix A.
- Column rank of a matrix equals its row rank.