MACHINE LEARNING WITH PYTHON

LINEAR REGRESSION

Themistoklis Diamantopoulos

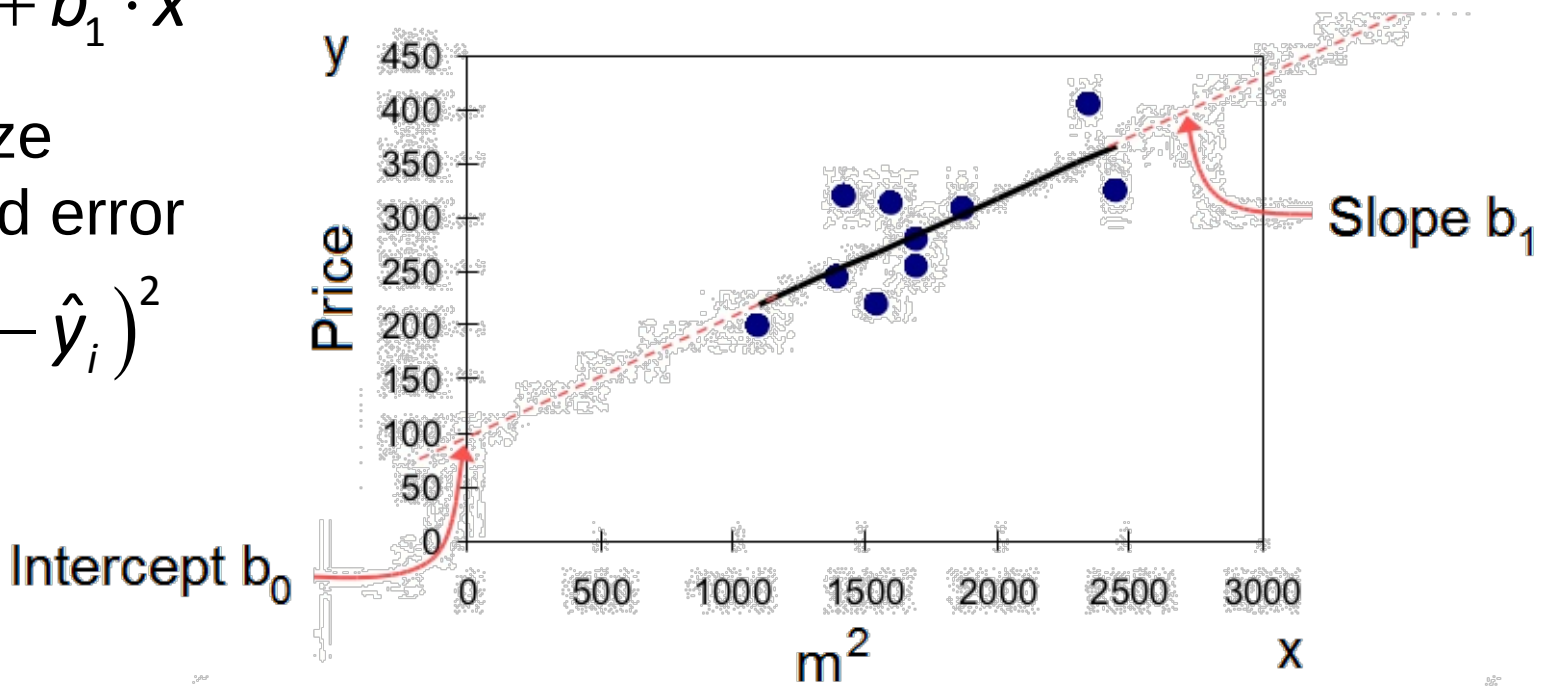
The Problem of Regression

- Find model that best fits the data

$$y = b_0 + b_1 \cdot x$$

- Minimize squared error

$$\sum_i \left(y_i - \hat{y}_i\right)^2$$



Source: https://www.slideshare.net/oxygen024/regression-analysis-8672920
https://web.williams.edu/Mathematics/sjmiller/public_html/BrownClasses/54/handouts/MethodLeastSquares.pdf

Example Least Squares Fitting

- Find model that best fits the data
  $y = b_0 + b_1 \cdot x$
- Minimize squared error
  $\sum_i \left(y_i - \hat{y}_i\right)^2$

- Derive error w.r.t. *b* and solve

$$b_1 = \frac{\sum_i \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum_i \left(x_i - \overline{x}\right)^2}$$

$$b_0 = \overline{y} - b_1 \overline{x}$$

Source: https://www.varsitytutors.com/hotmath/hotmath_help/topics/line-of-best-fit

Regression Evaluation

- Mean Absolute Error

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|$$

- Mean Squared Error and Root Mean Squared Error

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \qquad RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}$$

- Coefficient of Determination

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

Source: https://www.ritchieng.com/machine-learning-evaluate-linear-regression-model/