Regression and its Applications

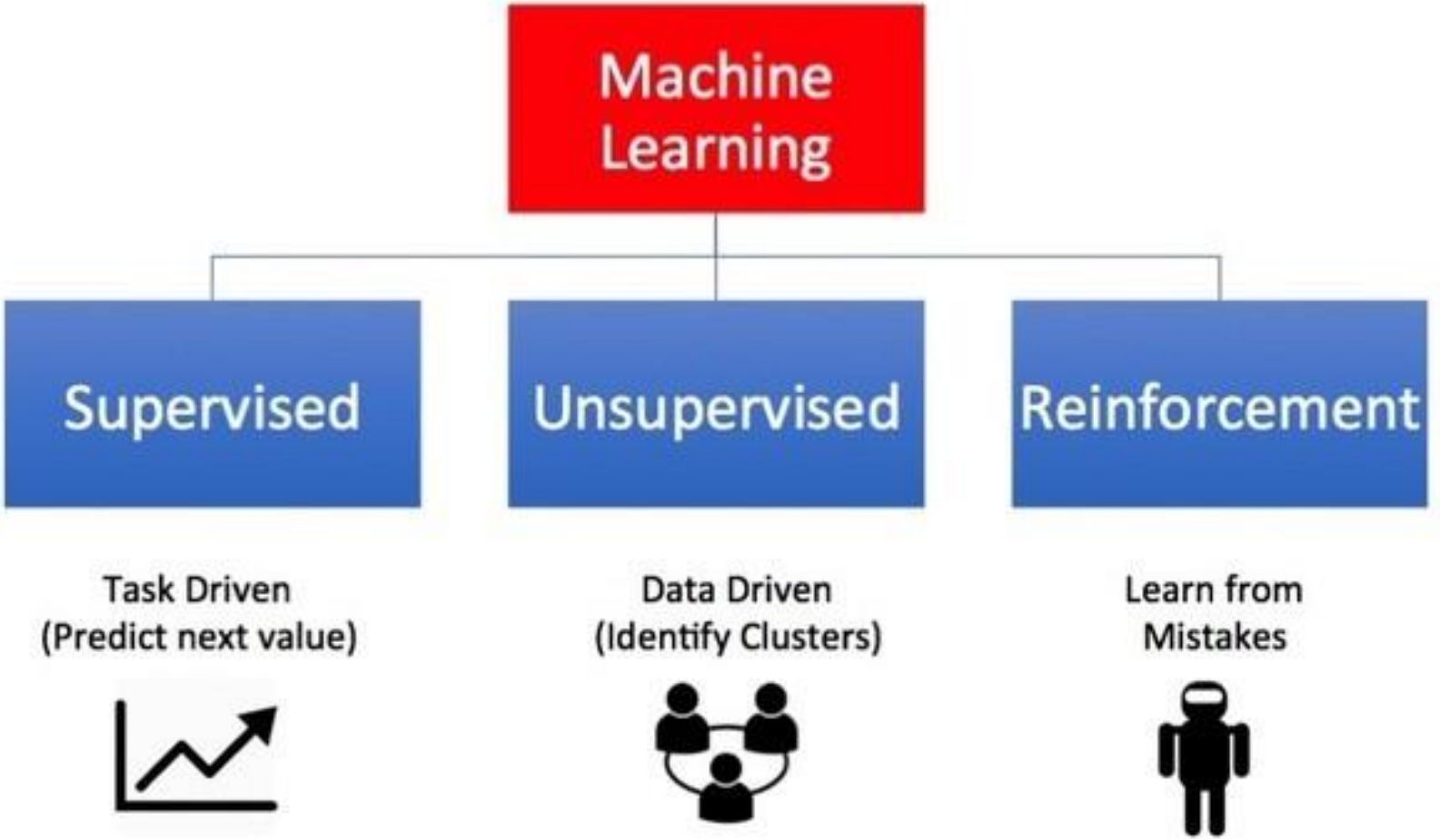
Albert C. Yang, M.D., Ph.D.

Institutes of Brain Science/Digital Medicine Center
National Yang-Ming University

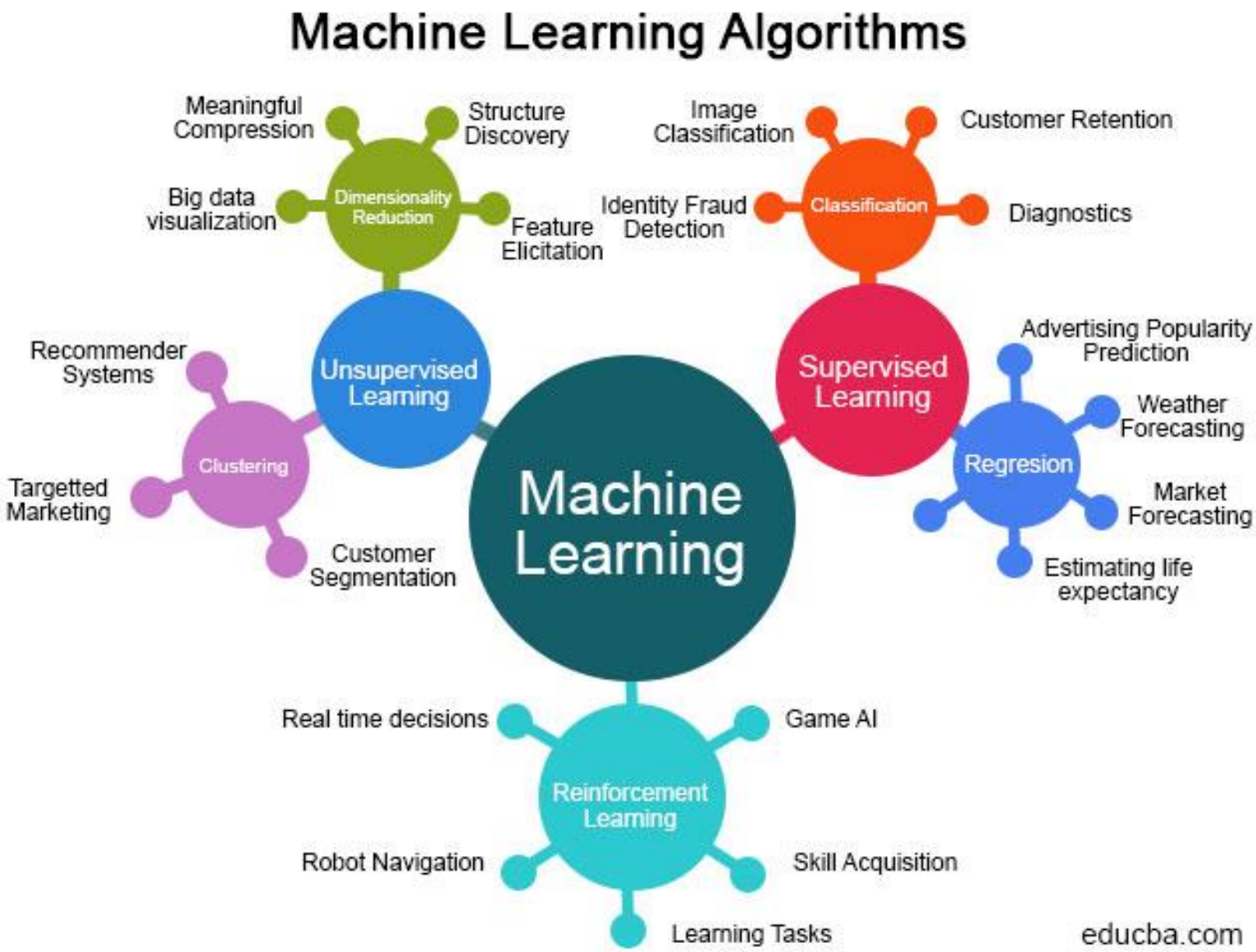
Mar 26, 2019

accyang@gmail.com

Types of Machine Learning

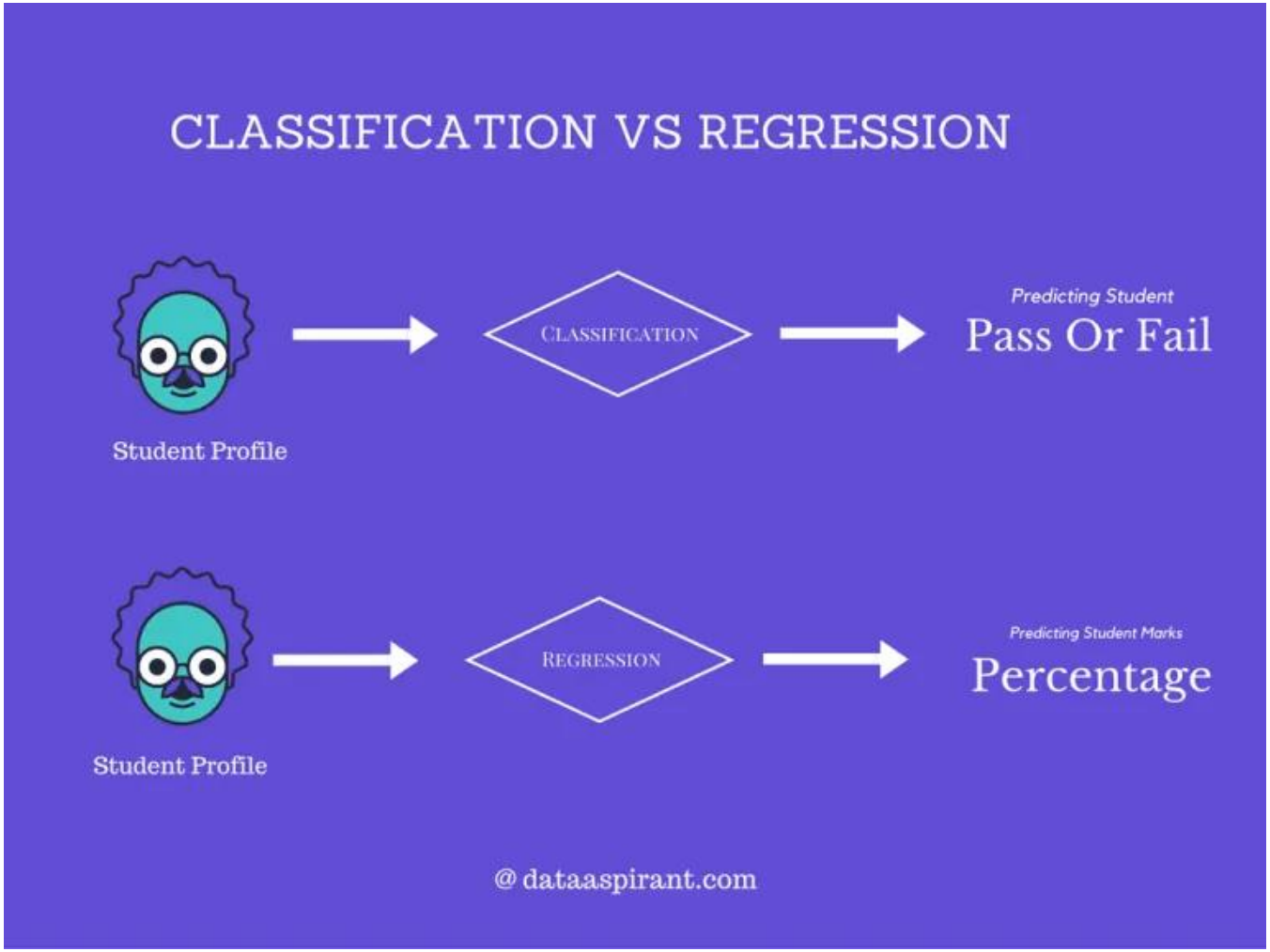


https://towardsdatascience.com/what-are-the-types-of-machine-learning-e2b9e5d1756f



https://www.educba.com/machine-learning-algorithms/

Classification vs. Regression

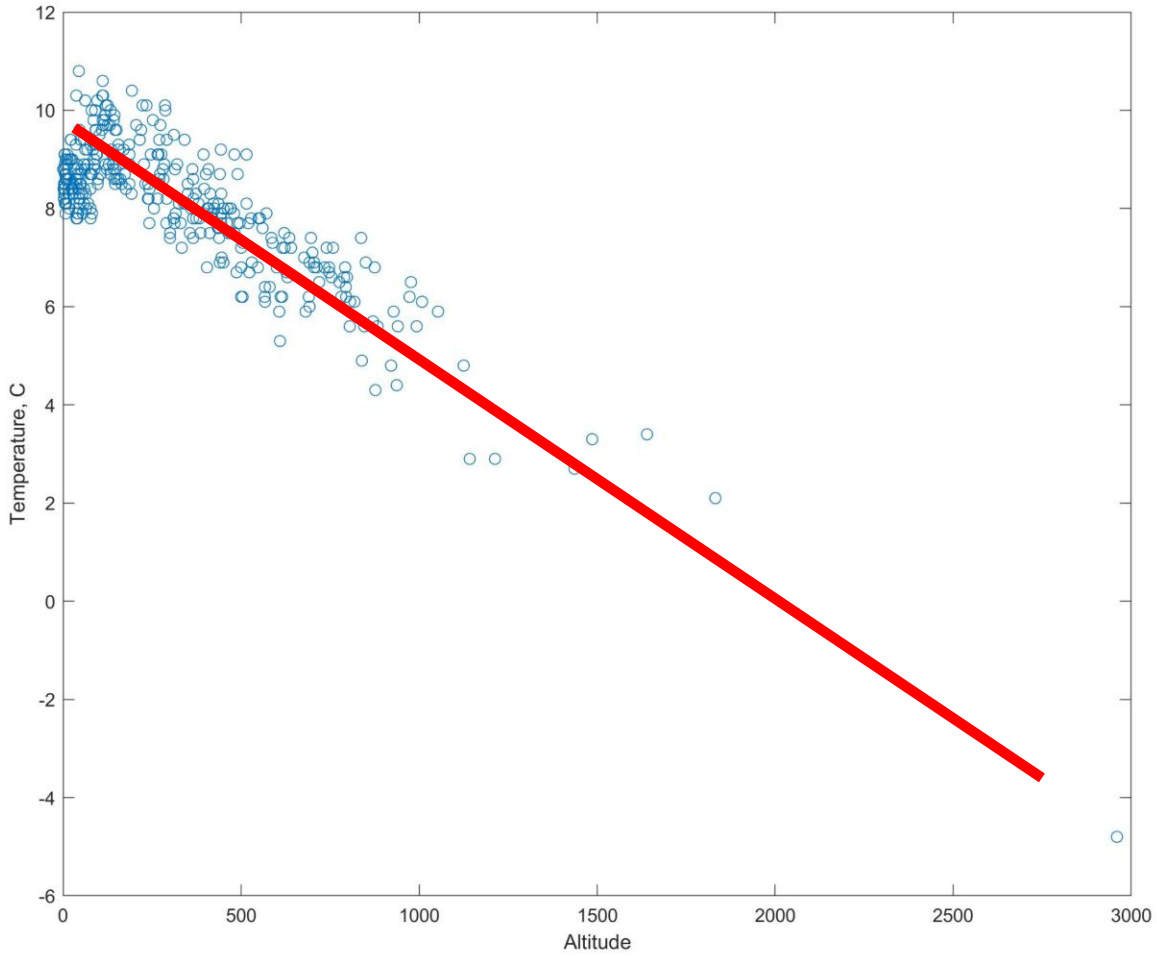


https://dataaspirant.com/2014/09/27/classification-and-prediction/

What is Linear Regression?

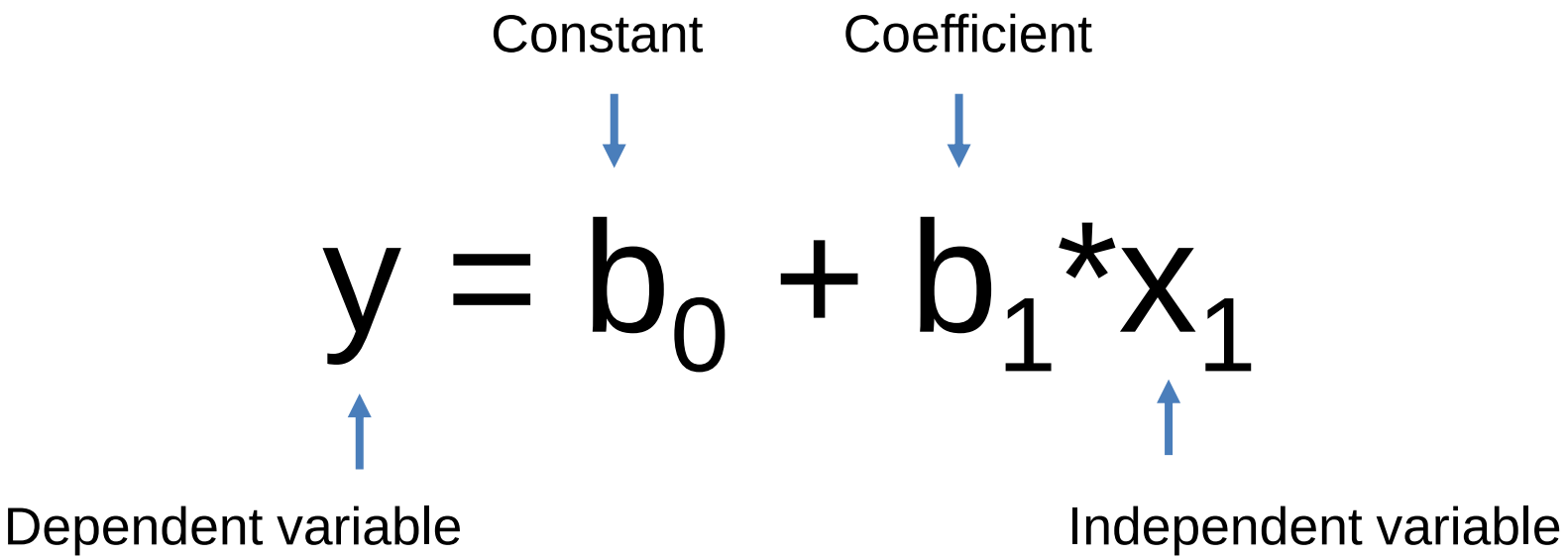
- **Learning**
  - A supervised algorithm that learns from a set of training samples.
  - Each training sample has one or more input values and a single output value.
  - The algorithm learns the line, plane or hyper-plane that best fits the training samples
- **Prediction**
  - Use the learned line, plane or hyper-plane to predict the output value for any
  - Input sample

Altitude vs. Temperature

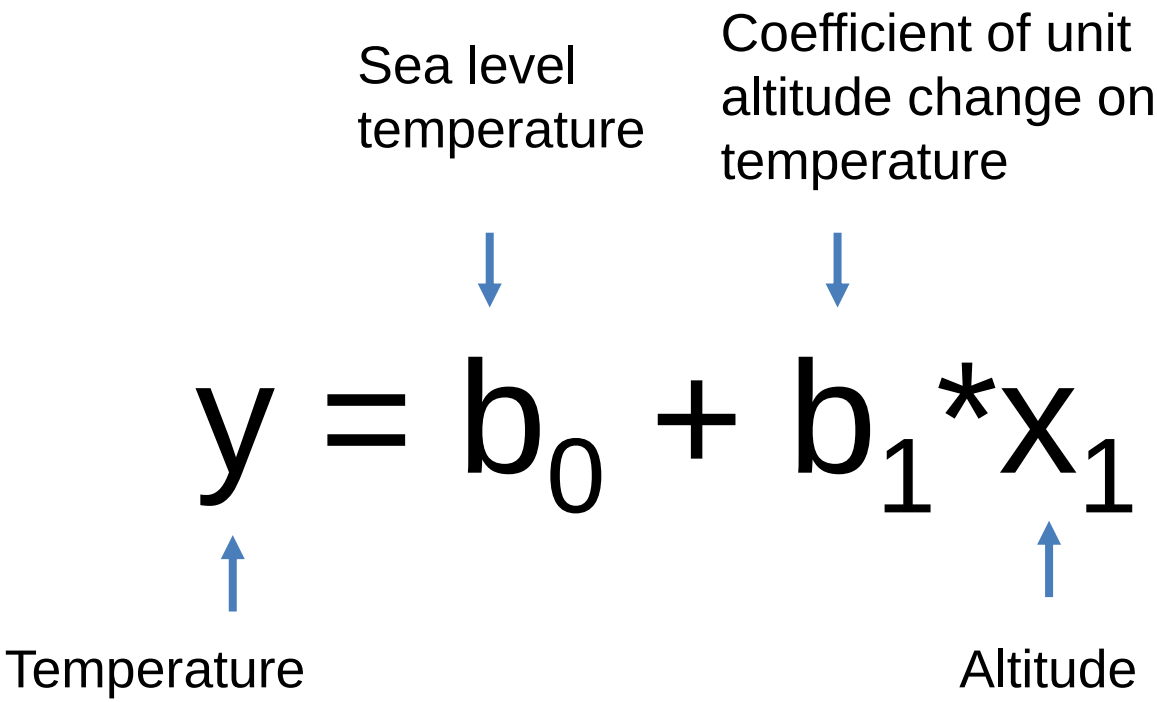


DWD data (Deutscher Wetterdienst)
data was taken at 349 weather station

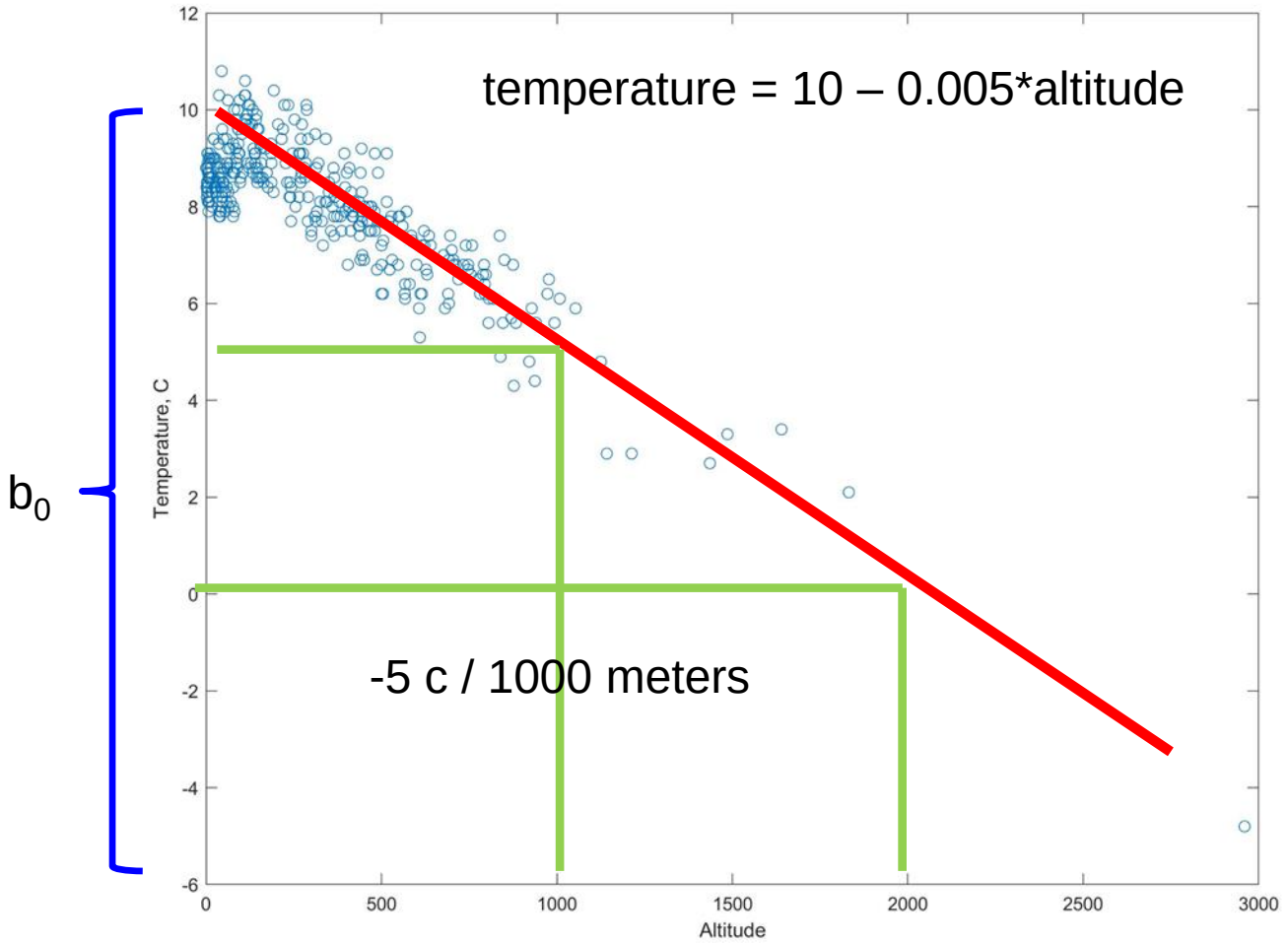
Simple Linear Regression



Simple Linear Regression

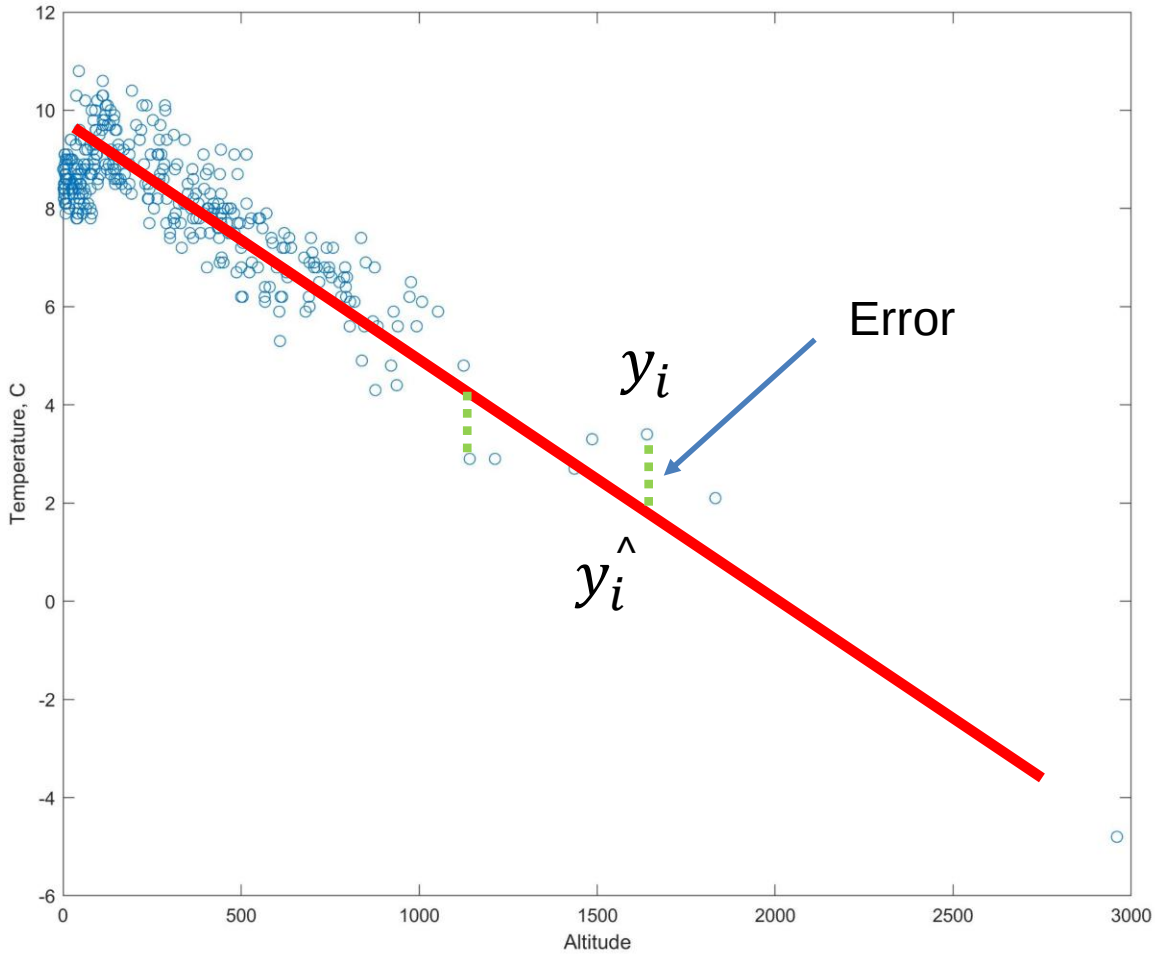


Altitude vs. Temperature

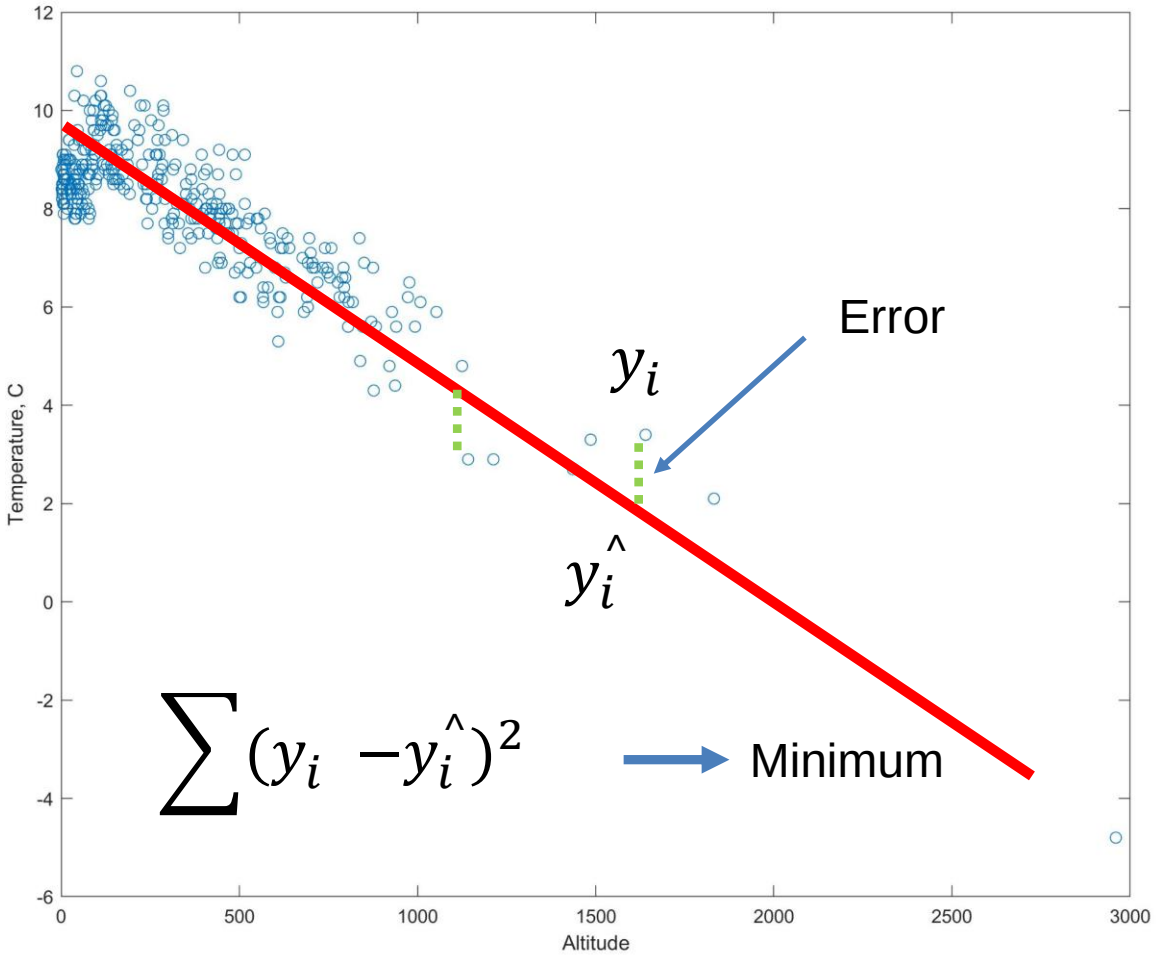


DWD data (Deutscher Wetterdienst)
data was taken at 349 weather station

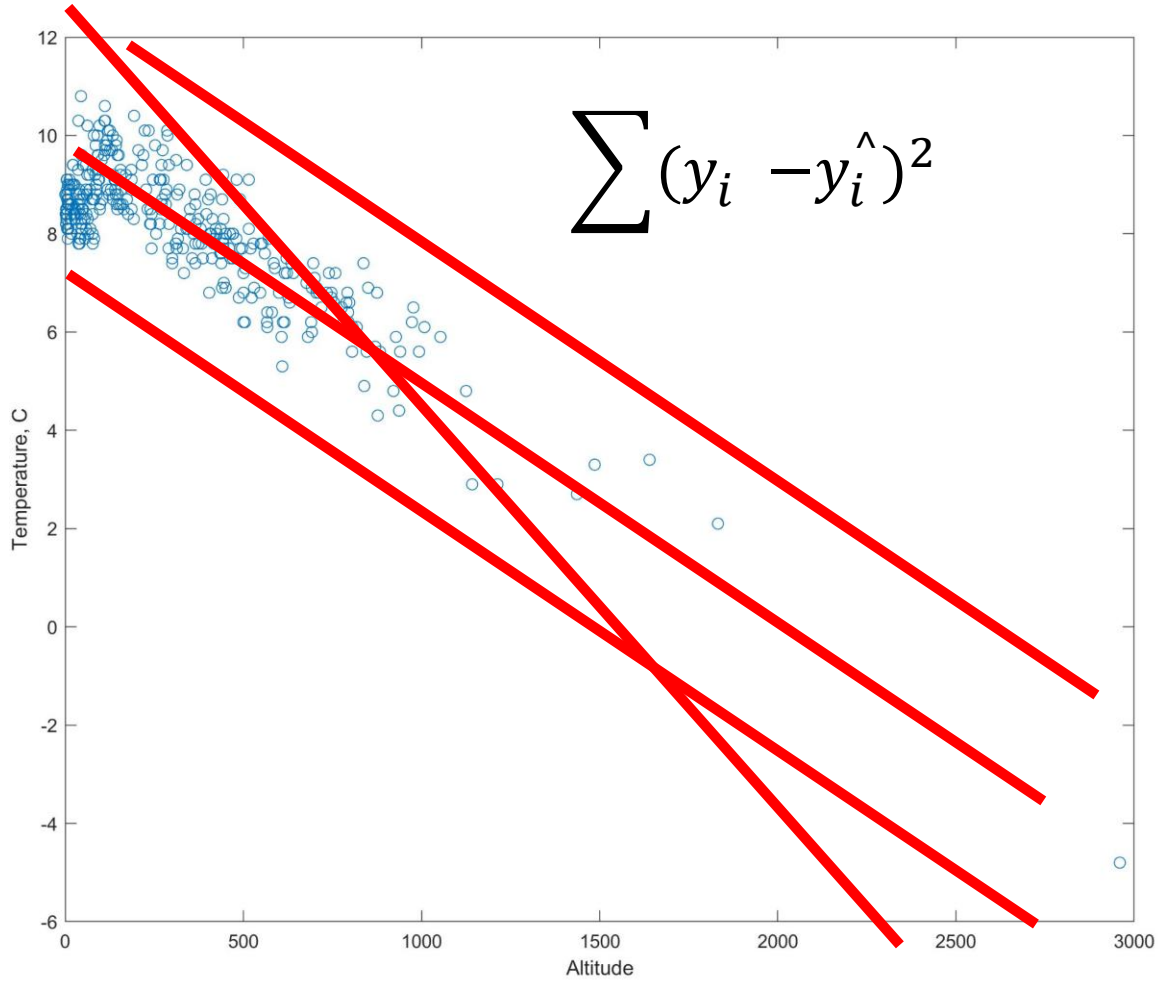
How to find errors?



How to define errors?
Ordinary Least Square



How to find the best fitted line?



Simple vs. Multiple Linear Regression

- Simple Linear Regression

$$y = b_0 + b_1 * x_1$$

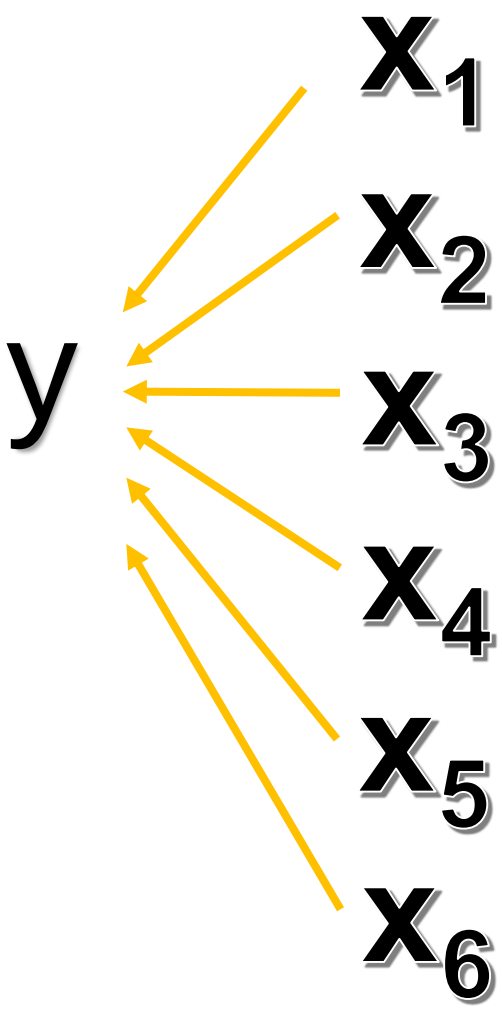
- Multiple Linear Regression

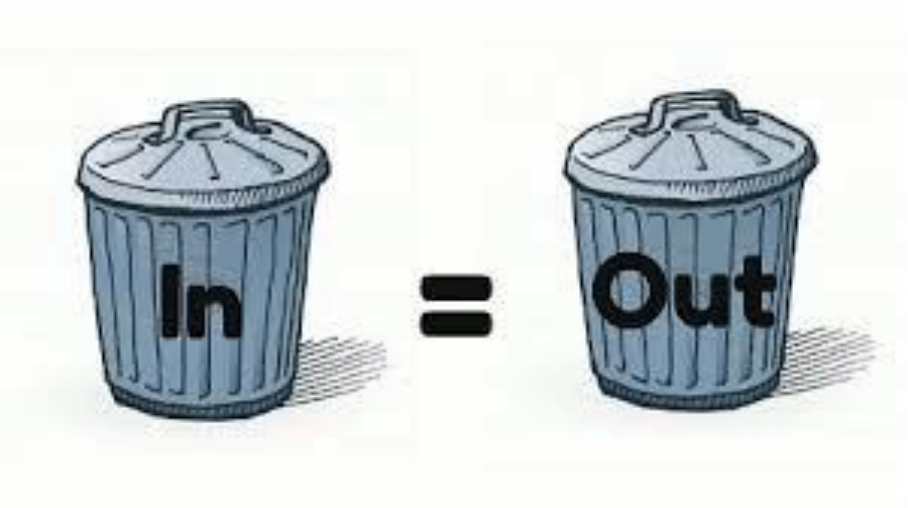
$$y = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + \ldots + b_n * x_n$$

Assumption of Linear Regression

- Linearity
- Homoscedasticity
- Multivariate normality
- Independence of errors
- Lack of multicollinearity

Building a Multiple Regression Model



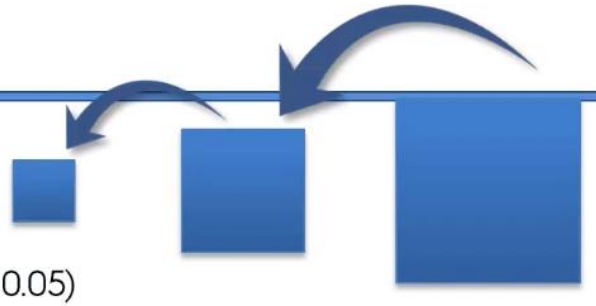


Variable Selection Methods

- All-in
- Backward Elimination
- Forward Selection
- Bidirectional Elimination
- Score comparison

Stepwise Method

Building A Model



Backward Elimination

STEP 1: Select a significance level to stay in the model (e.g. SL = 0.05)

STEP 2: Fit the full model with all possible predictors

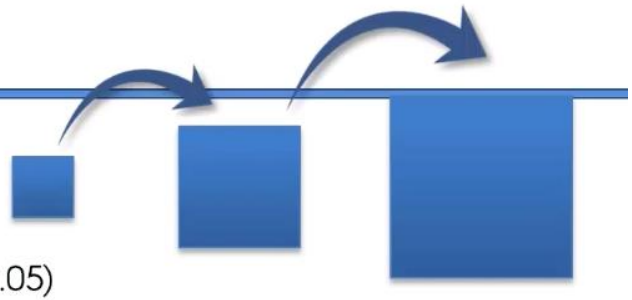
STEP 3: Consider the predictor with the highest P-value. If P > SL, go to STEP 4, otherwise go to FIN

STEP 4: Remove the predictor

STEP 5: Fit model without this variable*

Courtesy of Kirill Eremenko

Building A Model



Forward Selection

STEP 1: Select a significance level to enter the model (e.g. SL = 0.05)

STEP 2: Fit all simple regression models $\mathbf{y \sim x_n}$ Select the one with the lowest P-value

STEP 3: Keep this variable and fit all possible models with one extra predictor added to the one(s) you already have

STEP 4: Consider the predictor with the lowest P-value. If P < SL, go to STEP 3, otherwise go to FIN

Courtesy of Kirill Eremenko

Regressions

- Simple Linear Regression

  $y = b_0 + b_1 * x_1$

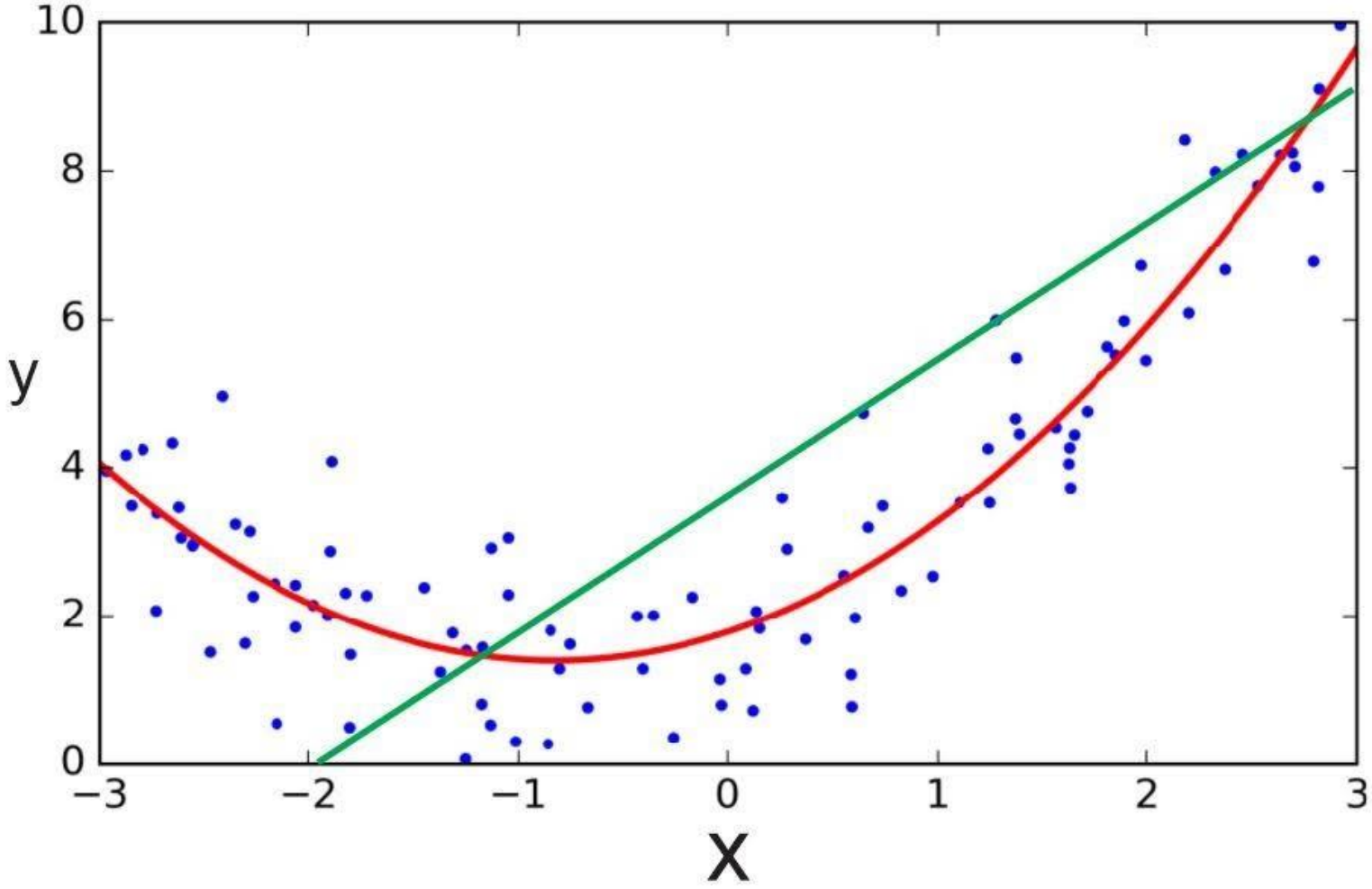
- Multiple Linear Regression

  $y = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + \ldots + b_n * x_n$

- Polynomial Linear Regression

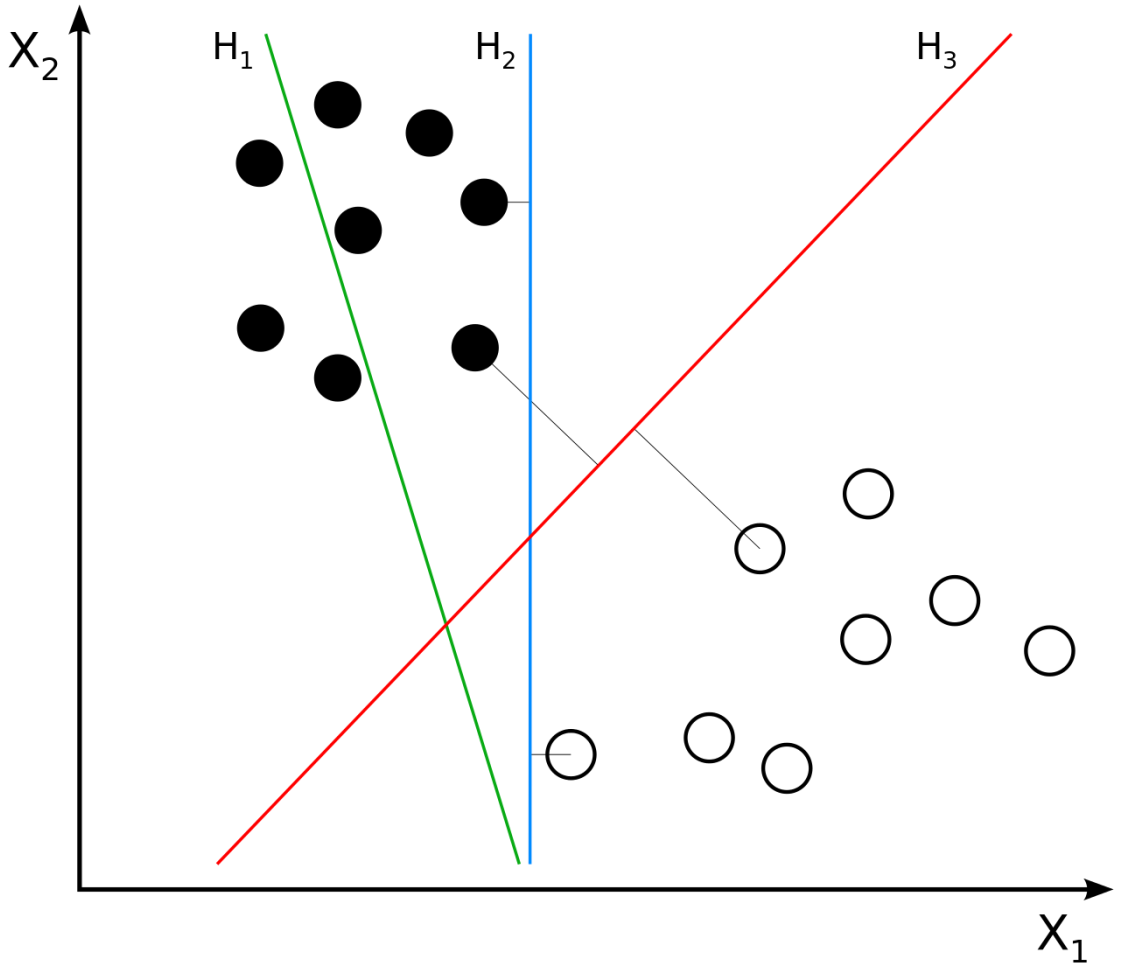
  $y = b_0 + b_1 * x_1 + b_2 * x_1^2 + b_3 * x_1^3 + \ldots + b_n * x_1^n$

Polynomial vs. Linear Regression



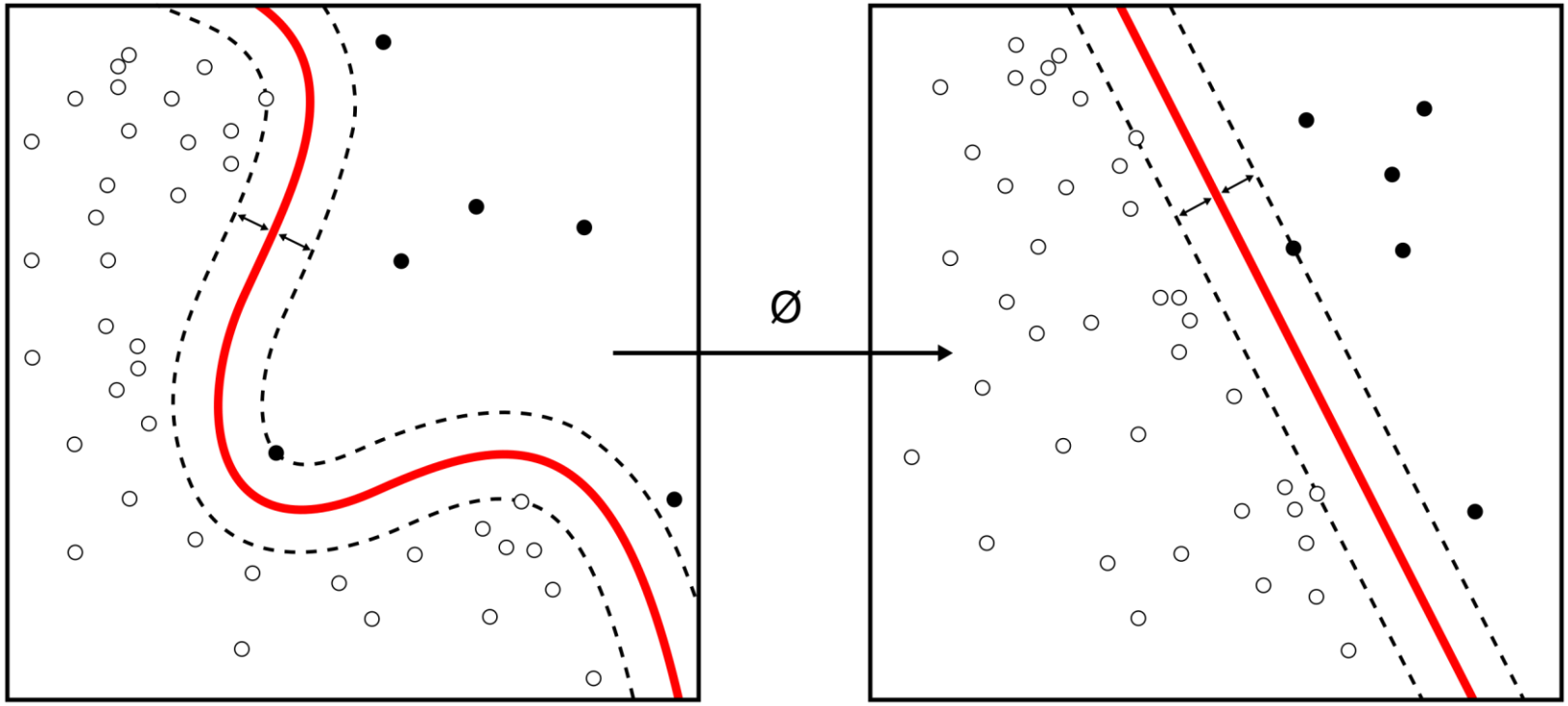
https://harangdev.github.io/machine-learning/5/

Support Vector Regression



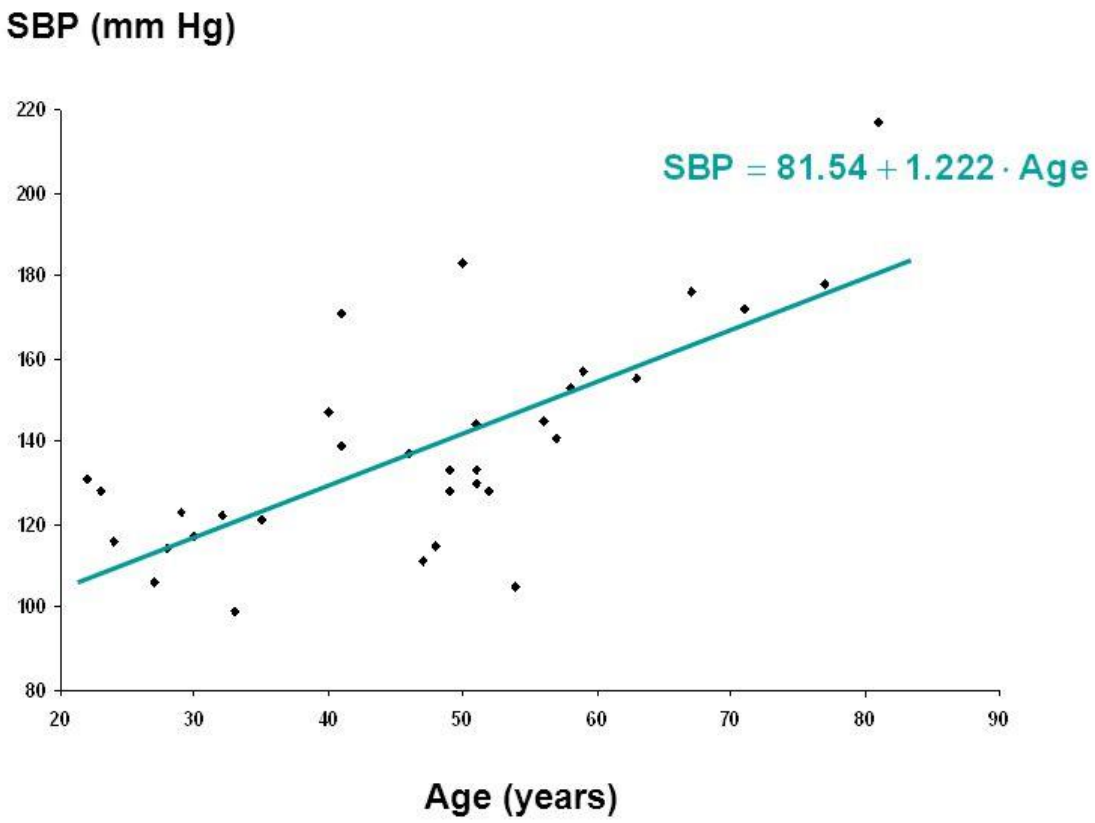
https://en.wikipedia.org/wiki/Support-vector_machine#/media/File:Svm_separating_hyperplanes_(SVG).svg

Support Vector Regression



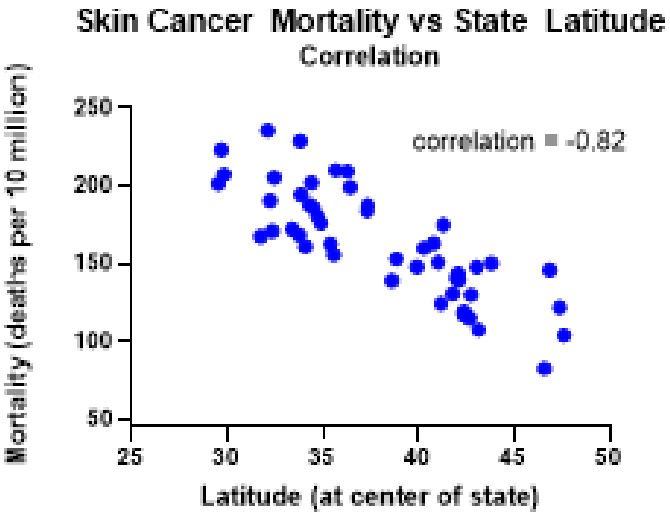
https://en.wikipedia.org/wiki/Support-vector_machine#/media/File:Kernel_Machine.svg

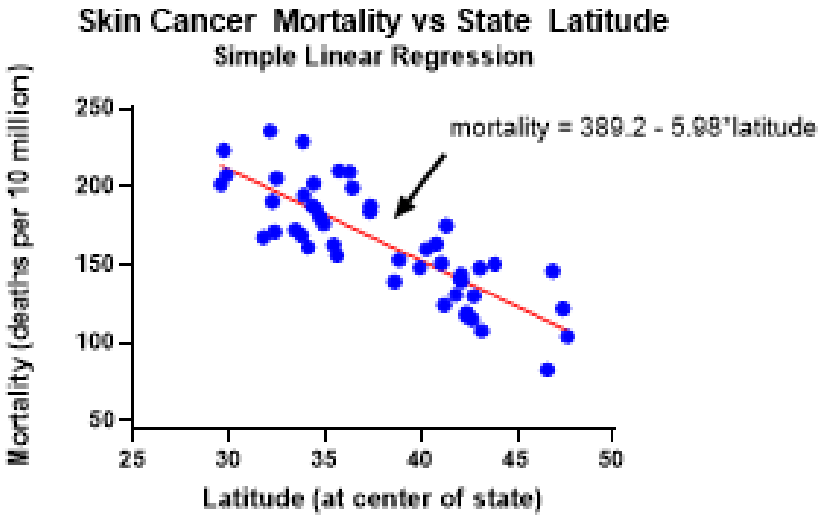
Applications of Regression



adapted from Colton T. Statistics in Medicine. Boston: Little Brown, 1974

Correlation vs. Coefficient





https://www.graphpad.com/support/faq/what-is-the-difference-between-correlation-and-linear-regression/

Summary of Regression Methods

- Widely used in medical research
- Easy implementation
- Good interpretability
- Limited learnability