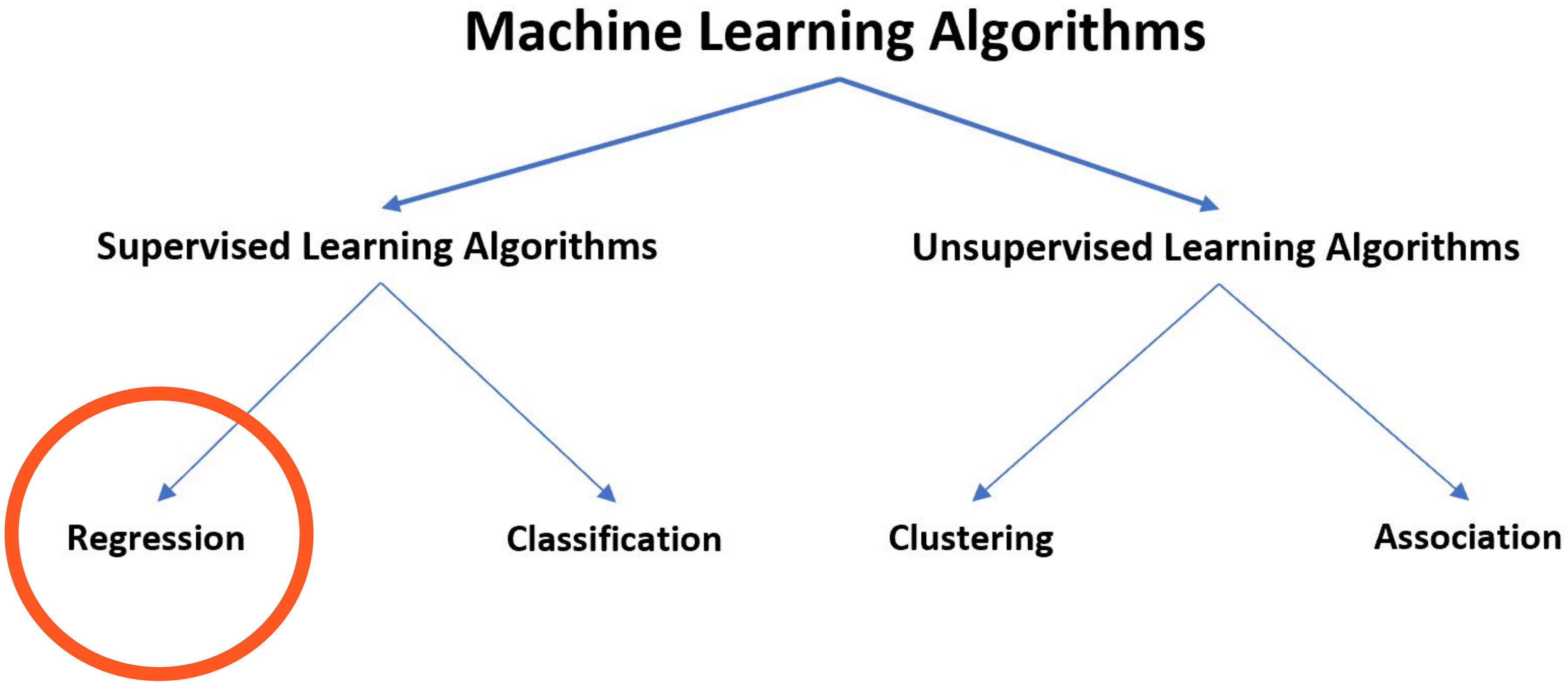
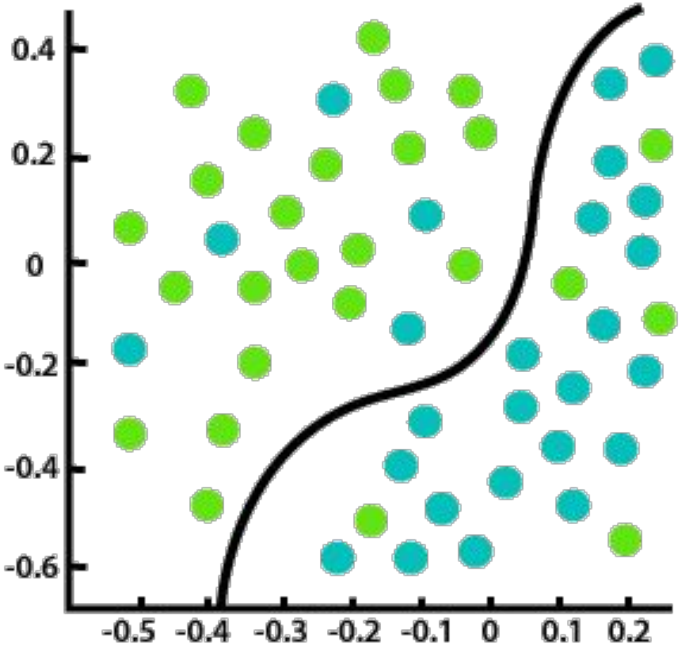
# Regression

Golem Bootcamp - XI 2024 - Wojciech Zarzecki

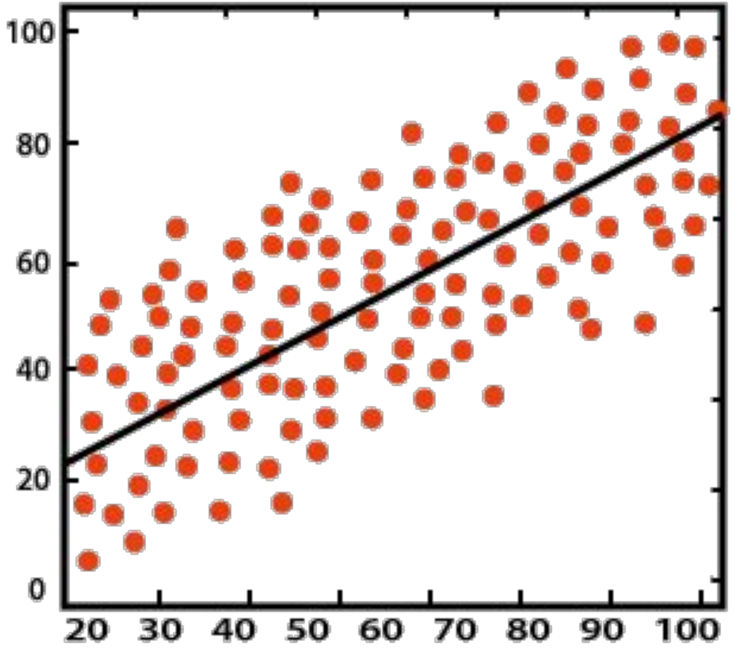
# Supervised learning



# Regression vs classification




Classification




Regression

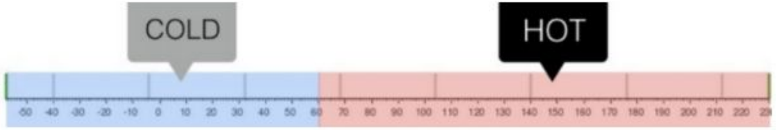
## Regression

What will be the temperature tomorrow?


Fahrenheit

## Classification

Will it be hot or cold tomorrow?

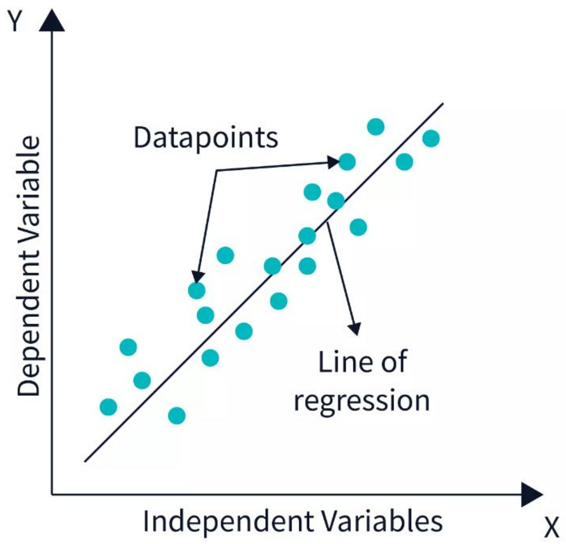



Fahrenheit

# Simple linear regression

$$y = a\,x + b$$

- y - dependent variable (test score)
- x - independent variable (hours studiet)
- a - slope / linear coef
- b - bias

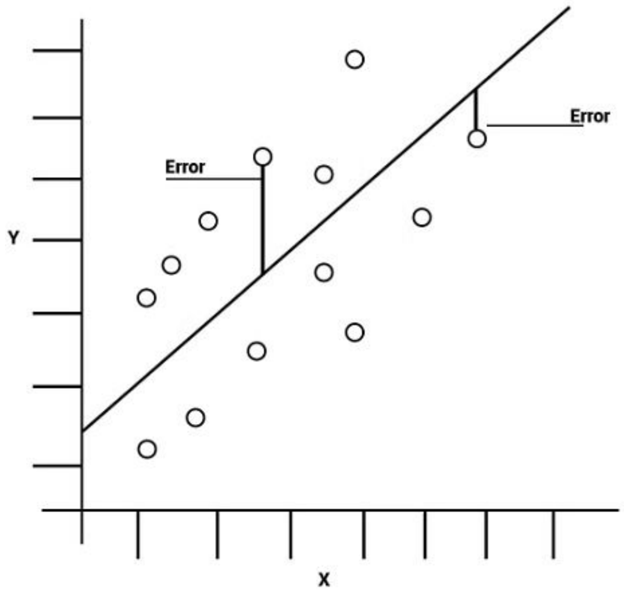


# Does the line fit?

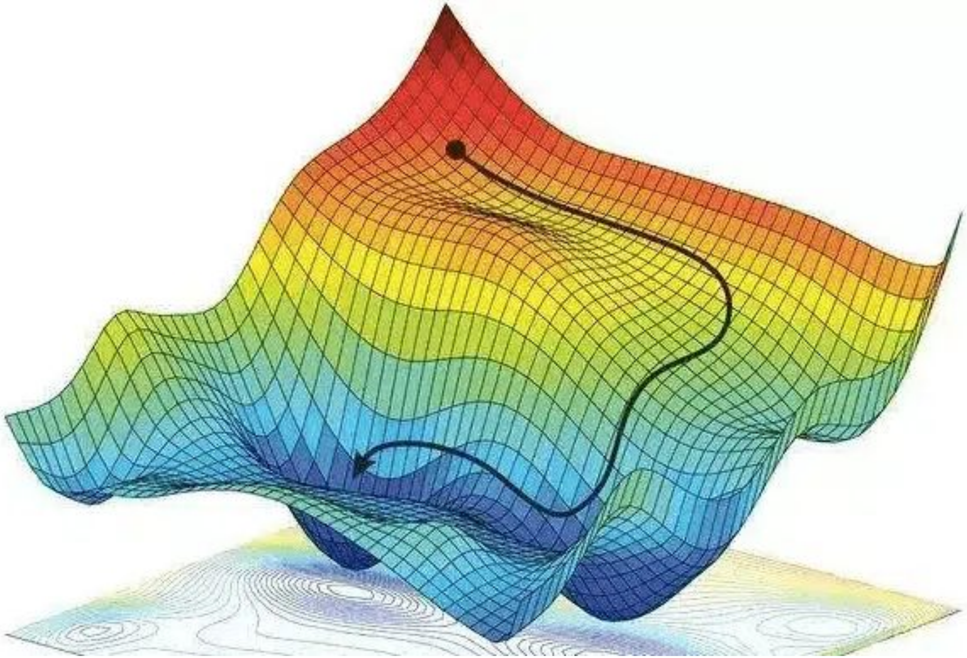
$$\mathrm{MSE} = \overbrace{\boxed{\frac{1}{n}\sum_{i=1}^{n}}}^{\text{Mean}}\Big(\overbrace{\boxed{Y_i - \hat{Y}_i}}^{\text{Error}}\Big)^{\overbrace{\boxed{2}}^{\text{Squared}}}$$

# Does the line fit?

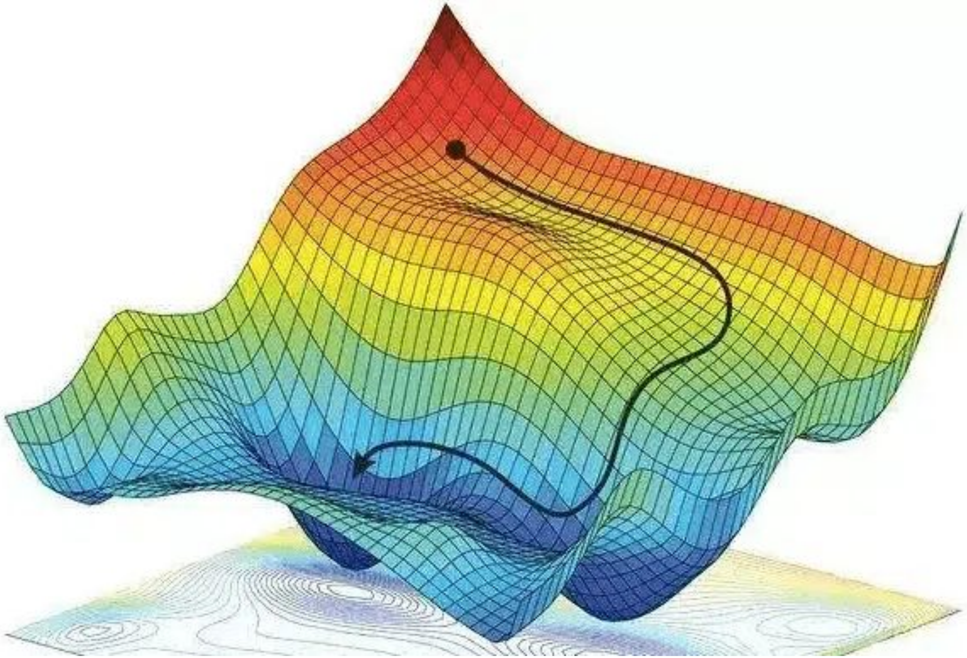
$$L(a,b) = \frac{1}{N}\sum_{i=1}^{N}(y_i - (ax_i + b))^2$$



# How to use the error?

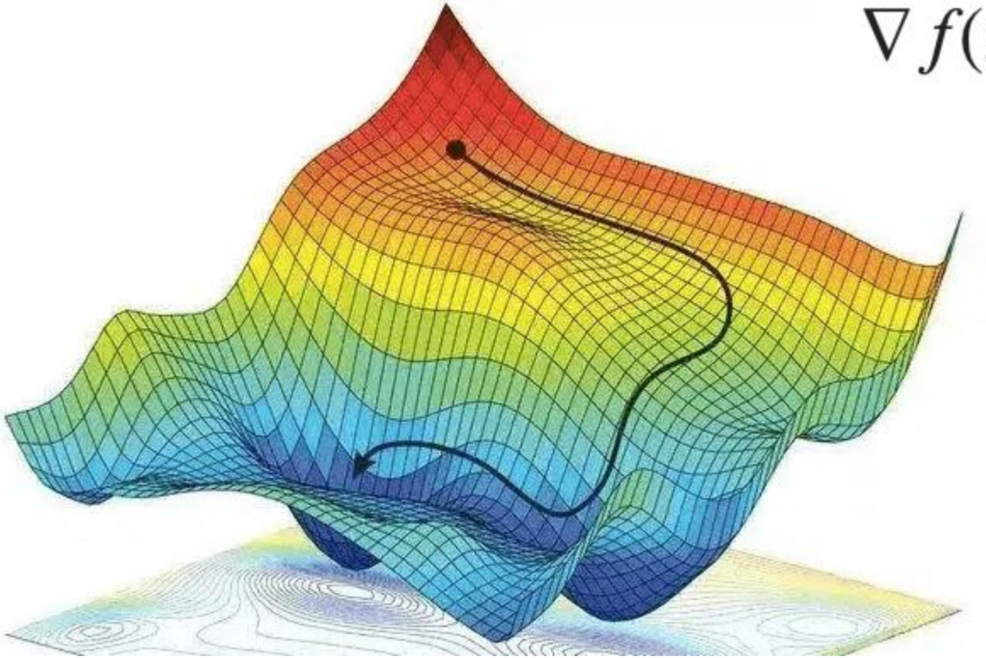


# How to use the error?



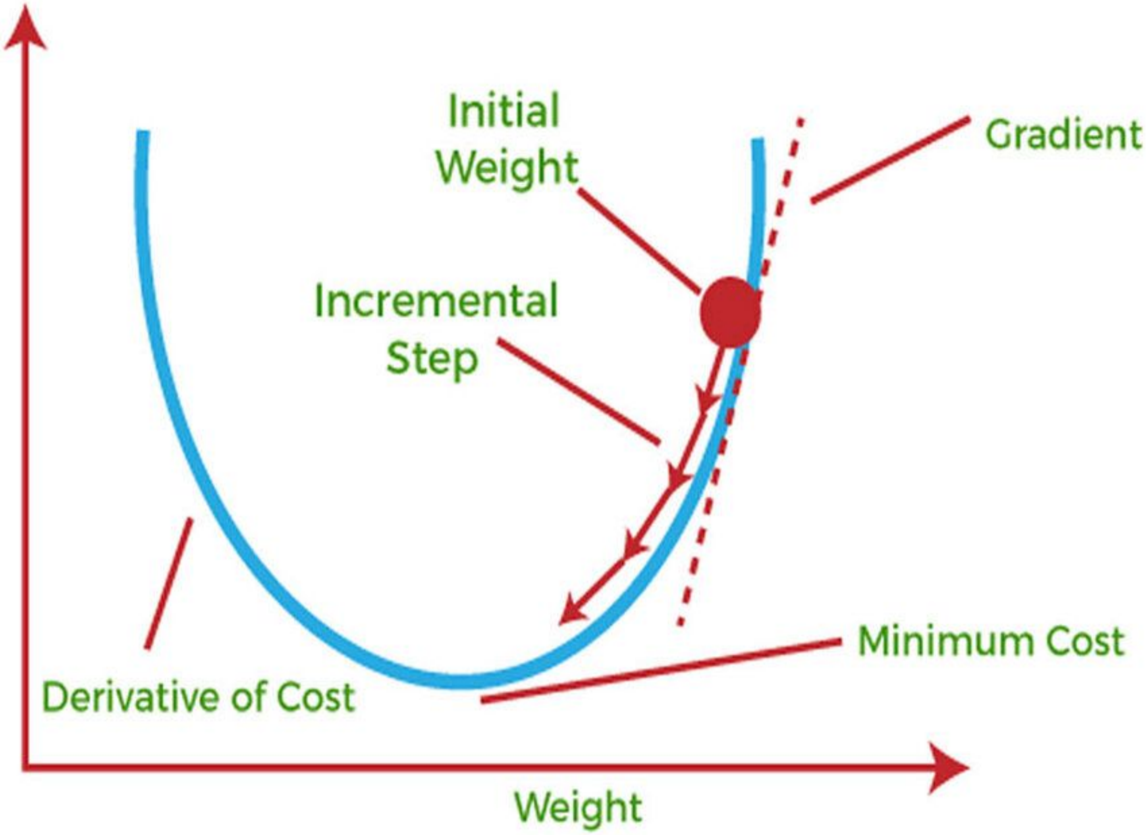
$$\nabla f = \left[ \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right]$$

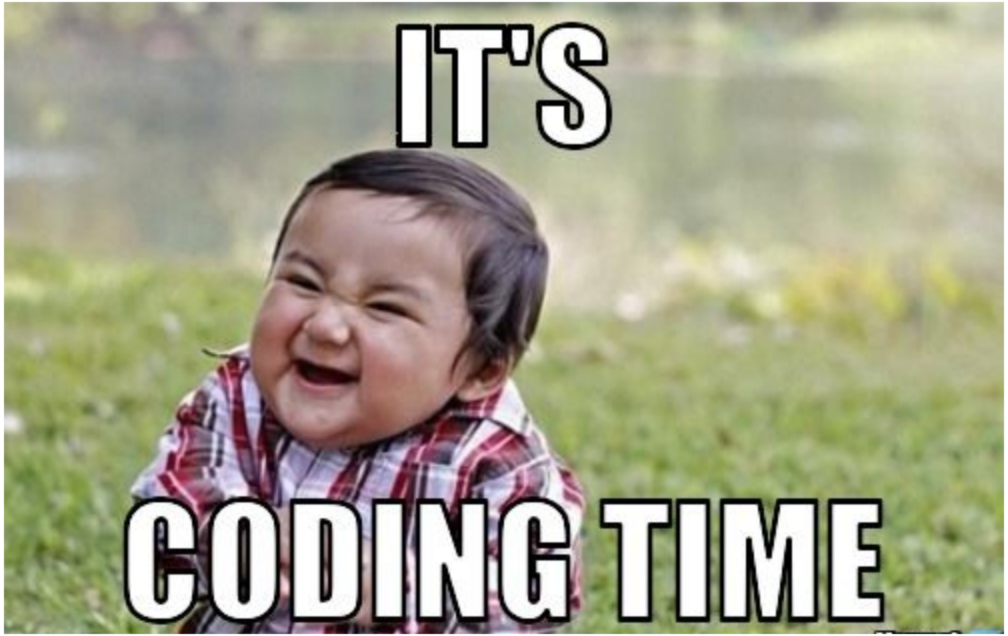
# How to use the error?



$$\nabla f(x, y, z) = [\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}]$$

# How to use the error?




IT'S
CODING TIME

# Math recap

$$\frac{d}{dx}(x^2) = 2x$$

# Math recap

$$\frac{d}{dx}(x^2) = 2x$$

chain rule

$$h(x) = (f \circ g)(x) = f(g(x))$$

$$h'(x) = f'(g(x)) \cdot g'(x)$$

# Math recap

$$\frac{d}{dx}(x^2) = 2x$$

$$h(x) = (f \circ g)(x) = f(g(x))$$

$$h'(x) = f'(g(x)) \cdot g'(x)$$

$$L(a, b) = \frac{1}{N} \sum_{i=1}^{N} (y_i - (ax_i + b))^2$$

# Math recap

$$\frac{d}{dx}(x^2) = 2x$$

$$h(x) = (f \circ g)(x) = f(g(x))$$

$$h'(x) = f'(g(x)) \cdot g'(x)$$

$$L(a,b) = \frac{1}{N}\sum_{i=1}^{N}(y_i - (ax_i + b))^2$$

partial derivatives

$$\frac{\partial L}{\partial a} = \frac{1}{N}\sum_{i=1}^{N} 2(y_i - (ax_i + b)) \cdot (-x_i) = -\frac{2}{N}\sum_{i=1}^{N} x_i(y_i - (ax_i + b))$$

# Math recap

$$\frac{d}{dx}(x^2) = 2x$$

$$h(x) = (f \circ g)(x) = f(g(x))$$

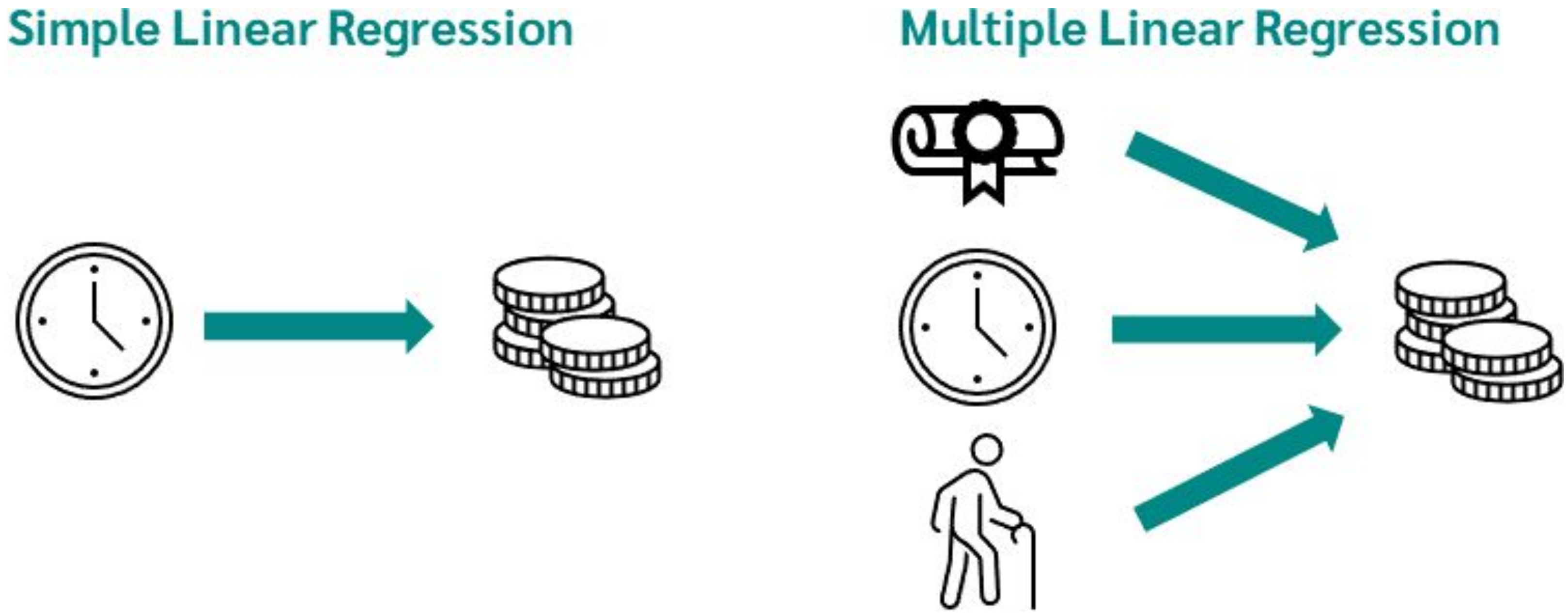
$$h'(x) = f'(g(x)) \cdot g'(x)$$

$$L(a, b) = \frac{1}{N} \sum_{i=1}^{N} (y_i - (ax_i + b))^2$$

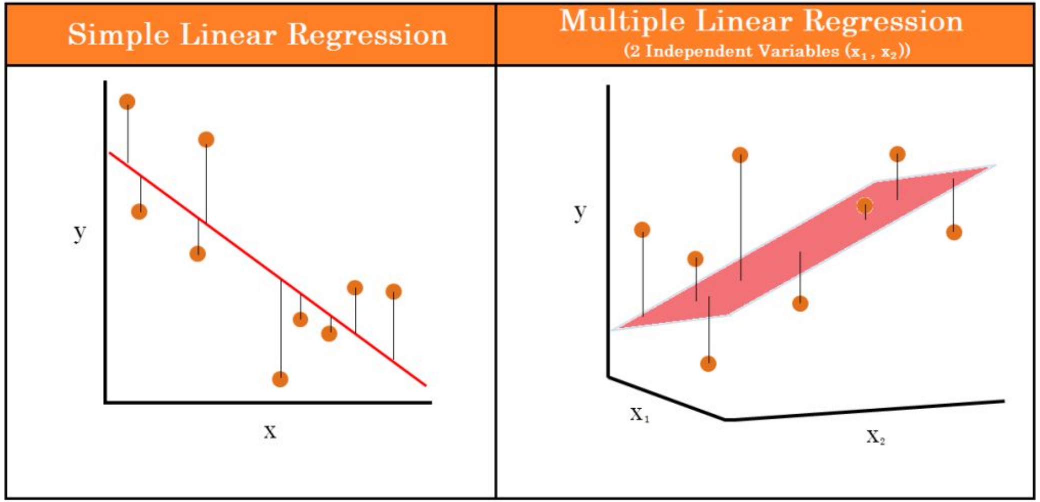
$$\frac{\partial L}{\partial a} = \frac{1}{N} \sum_{i=1}^{N} 2(y_i - (ax_i + b)) \cdot (-x_i) = -\frac{2}{N} \sum_{i=1}^{N} x_i(y_i - (ax_i + b))$$

$$\frac{\partial L}{\partial b} = \frac{1}{N} \sum_{i=1}^{N} 2(y_i - (ax_i + b)) \cdot (-1) = -\frac{2}{N} \sum_{i=1}^{N} (y_i - (ax_i + b))$$

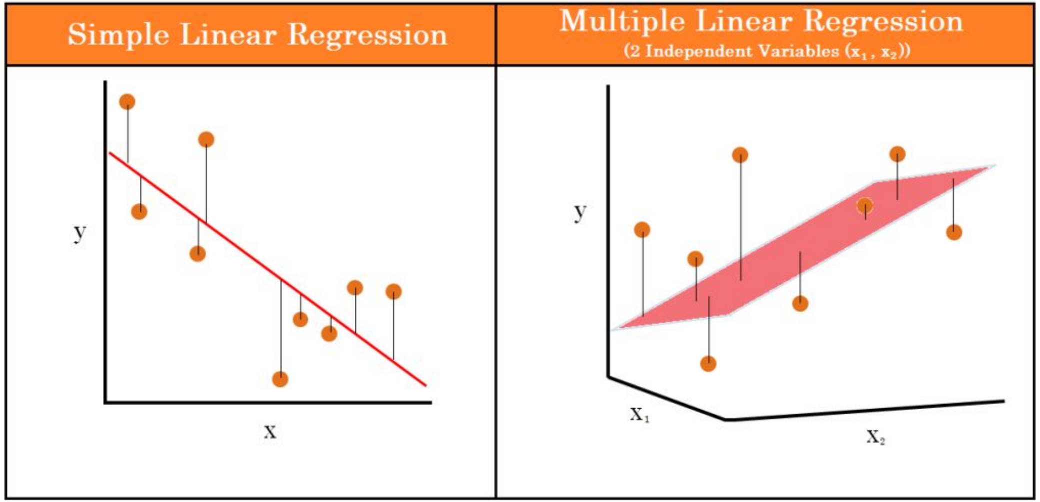
# Multiple linear regression



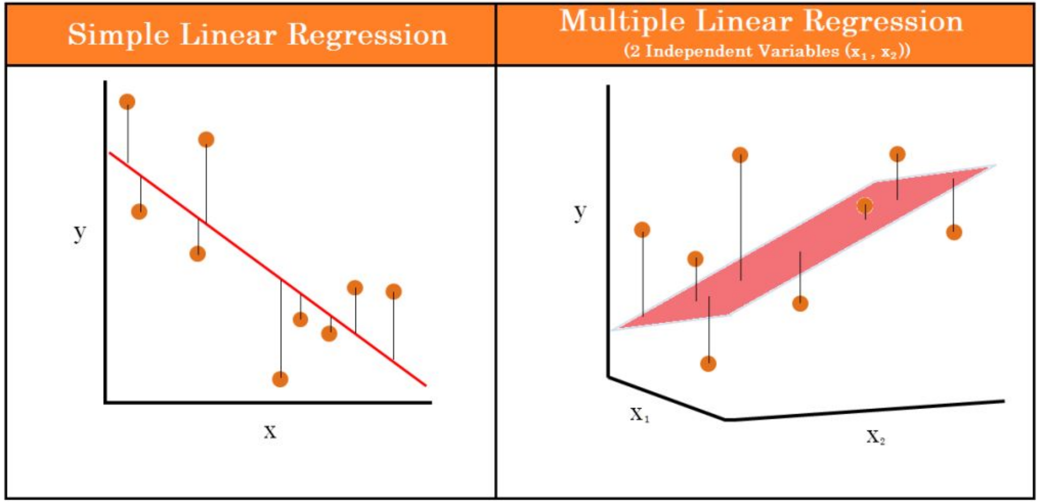
# Multiple linear regression



# Multiple linear regression



# Multiple linear regression




$$y = b_0 + b_1{}^*x_1$$

Dependent variable (DV)

Independent variables (IVs)

$$y = b_0 + b_1{}^*x_1 + b_2{}^*x_2 + \dots + b_n{}^*x_n$$

# Let's code

# Lasso regression

**least absolute shrinkage and selection operator**

# Lasso regression

**least absolute shrinkage and selection operator**

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

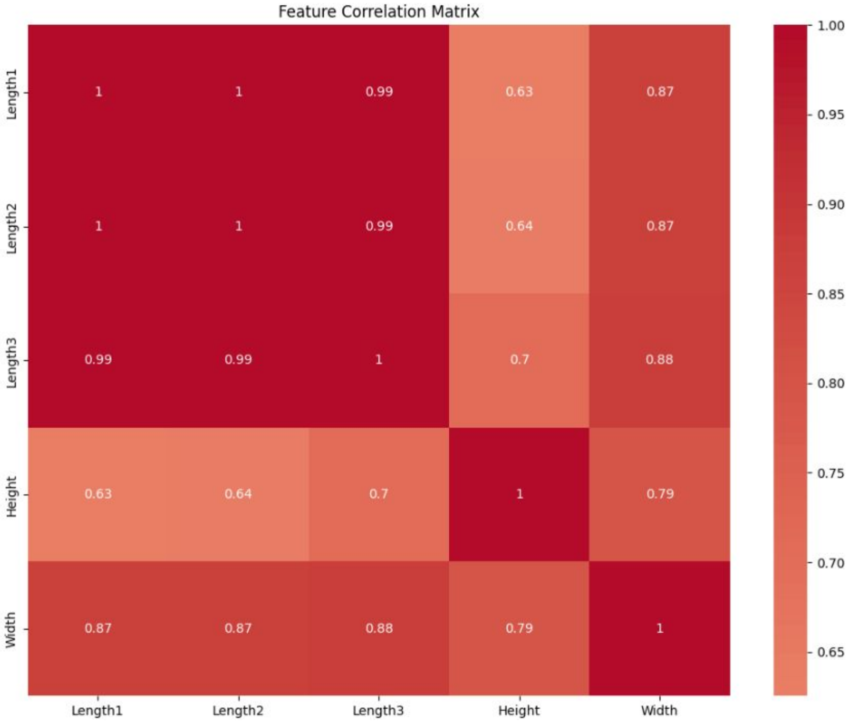
# Lasso regression

**least absolute shrinkage and selection operator**

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

LASSO Part

# Feature selection

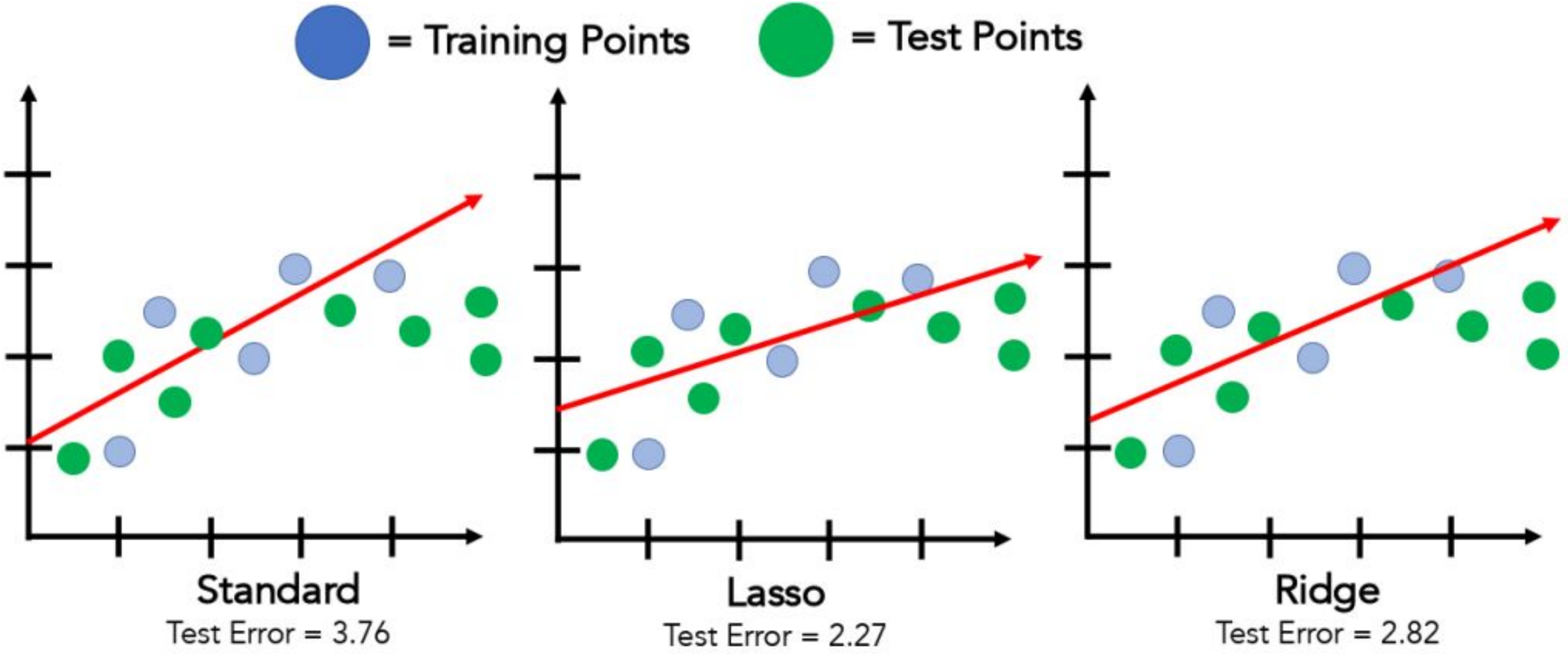



# Let's code

# Ridge regression

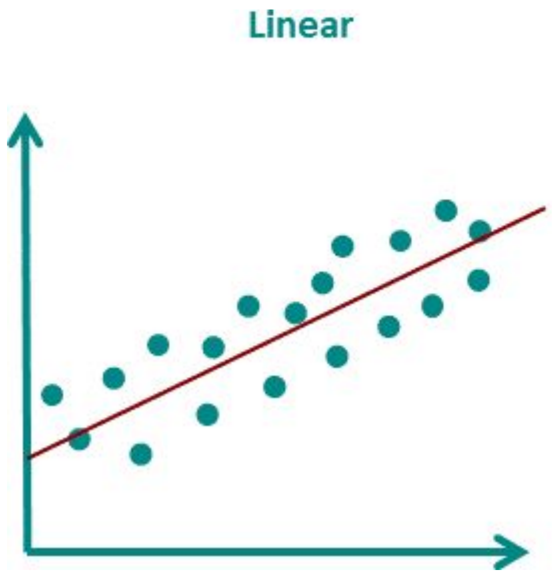
$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

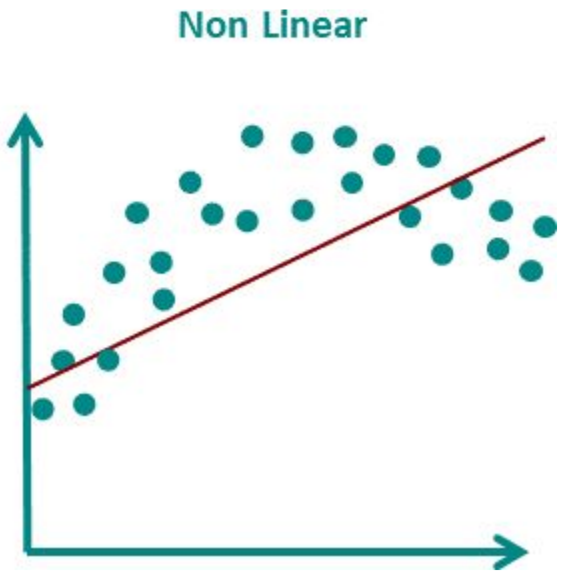
# Ridge regression



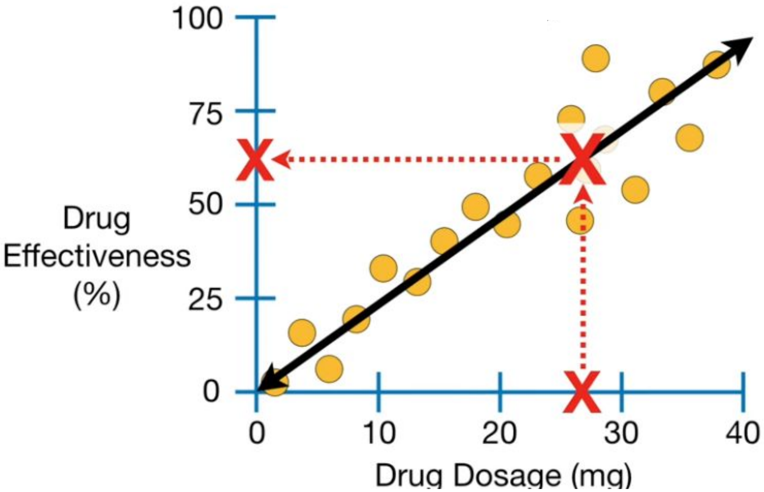

*Our Data was actually "parabolic" but we couldn't tell from the small training sample.*

# More complex data

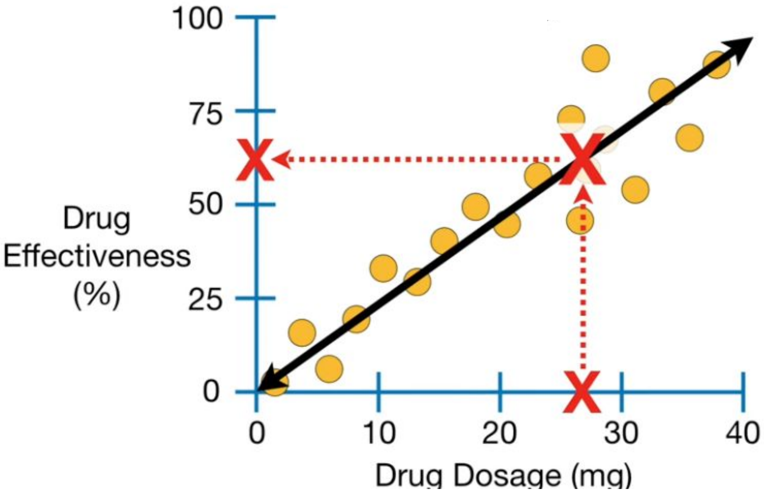


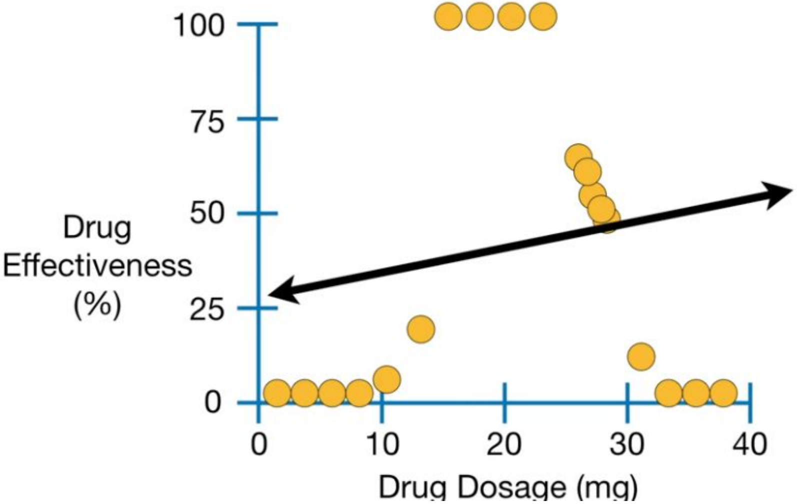



# Regression trees motivation

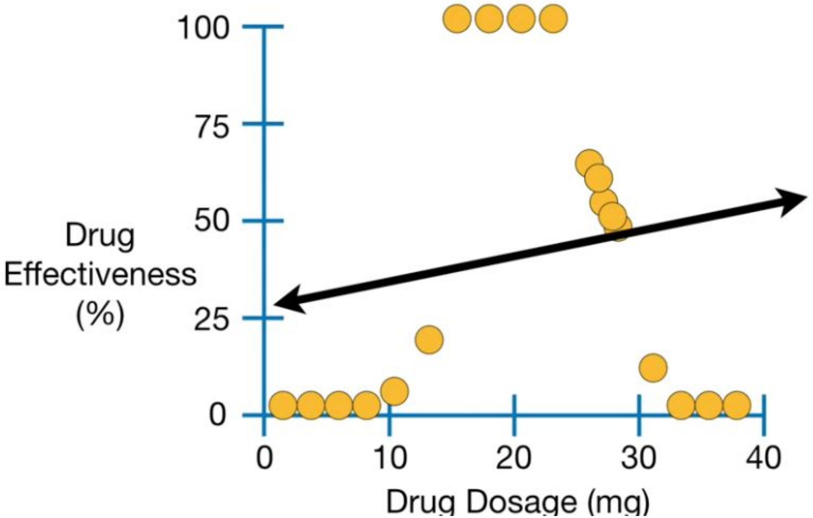


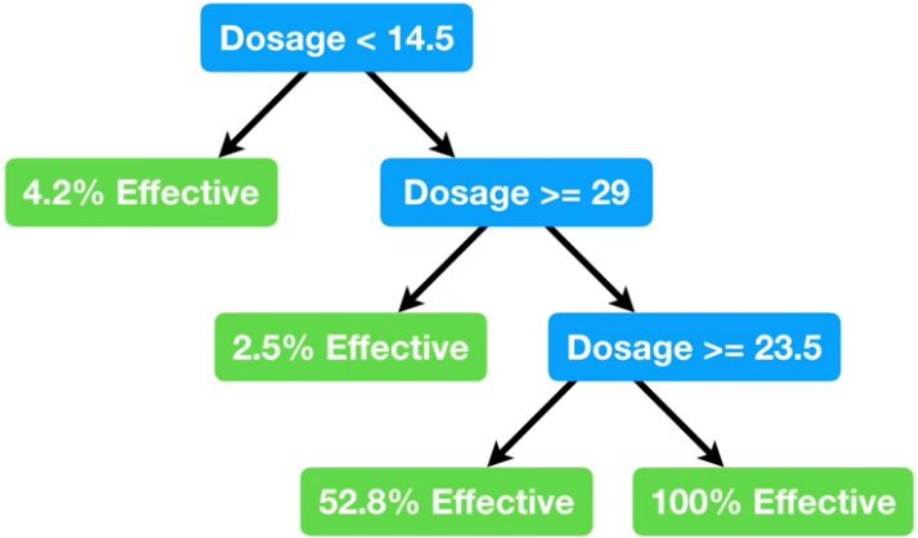
# Regression trees motivation



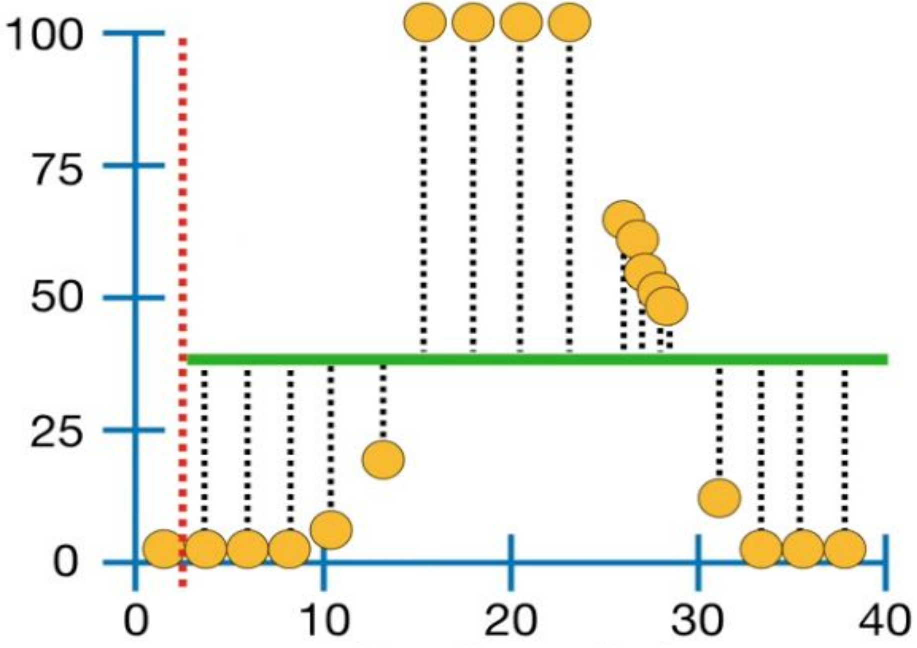



# The trees way

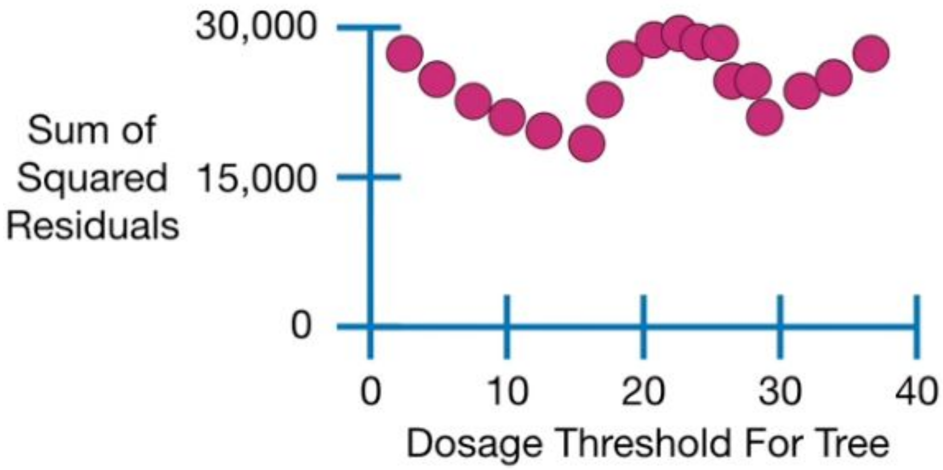


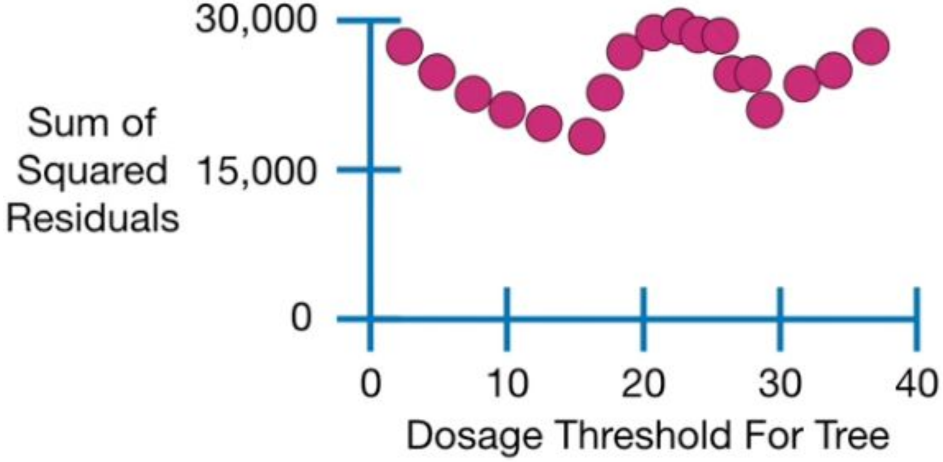
# Squared residuals

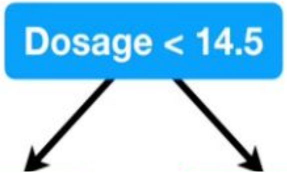


# Choosing a node

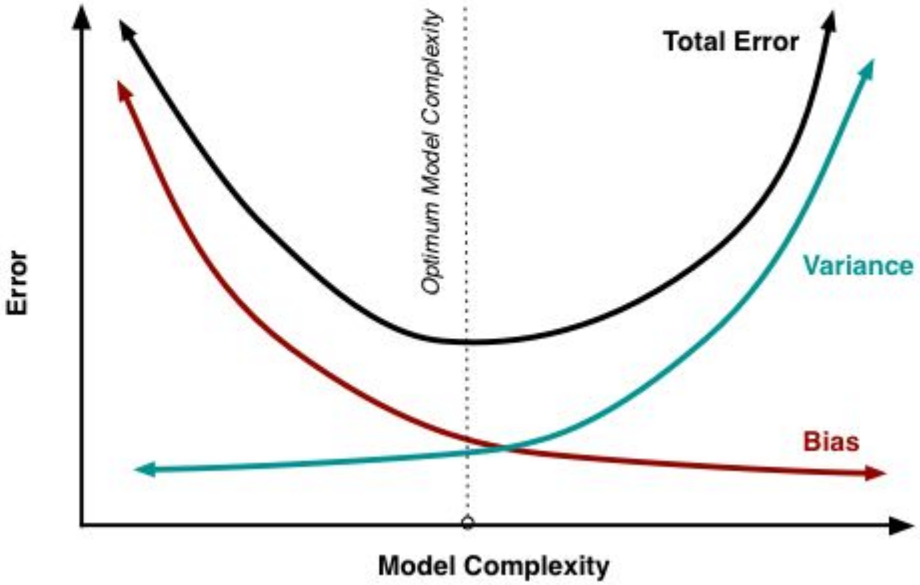


# Choosing a node

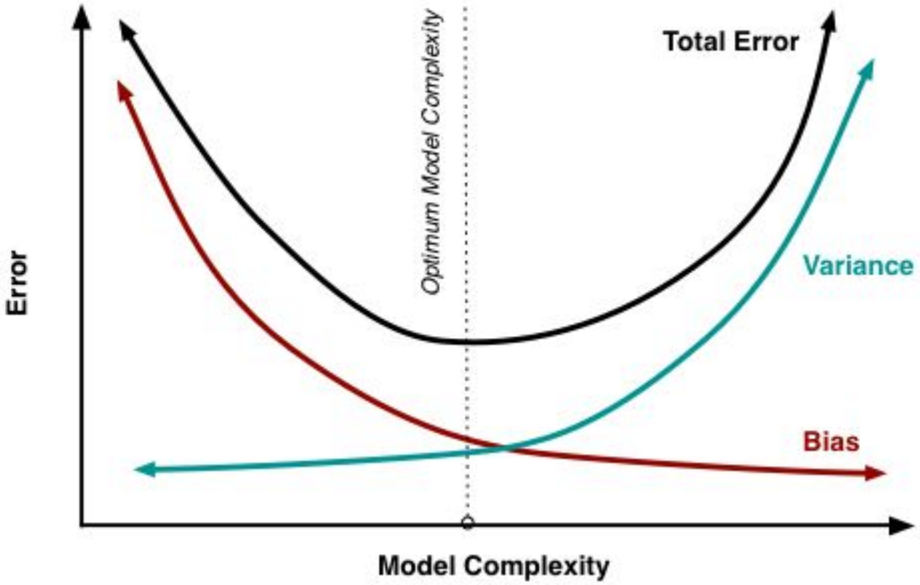


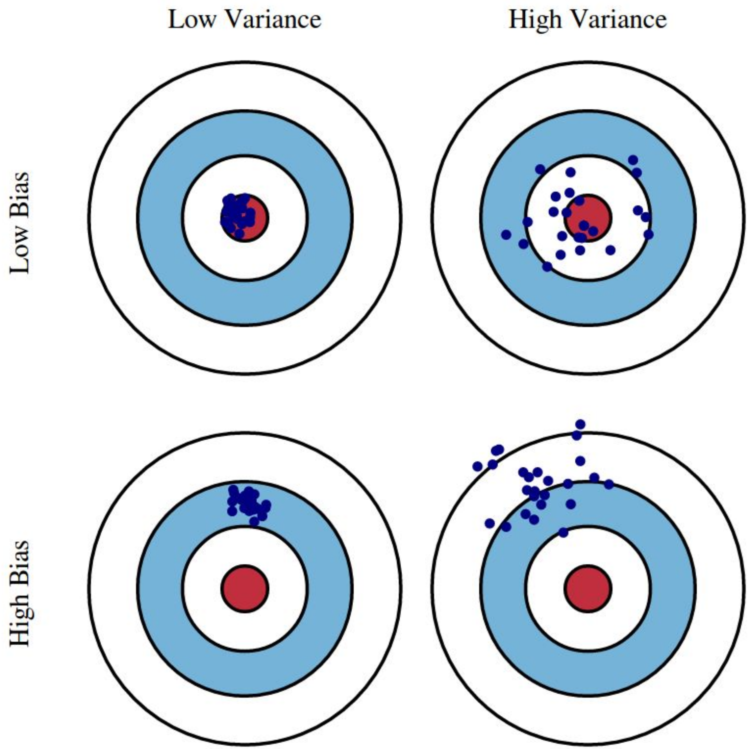



# Bias - variance



# Bias - variance



# Let's code


THANK YOU!
YOU ROCK!