

Regression



Golem Bootcamp - XI 2024 - Wojciech Zarzecki

Supervised learning

Machine Learning Algorithms

Supervised Learning Algorithms

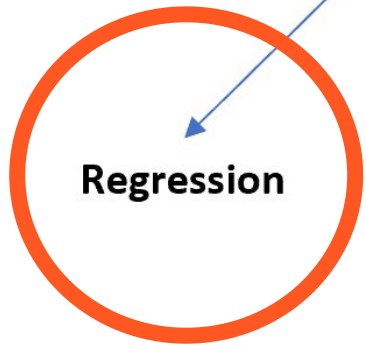
Unsupervised Learning Algorithms

Regression

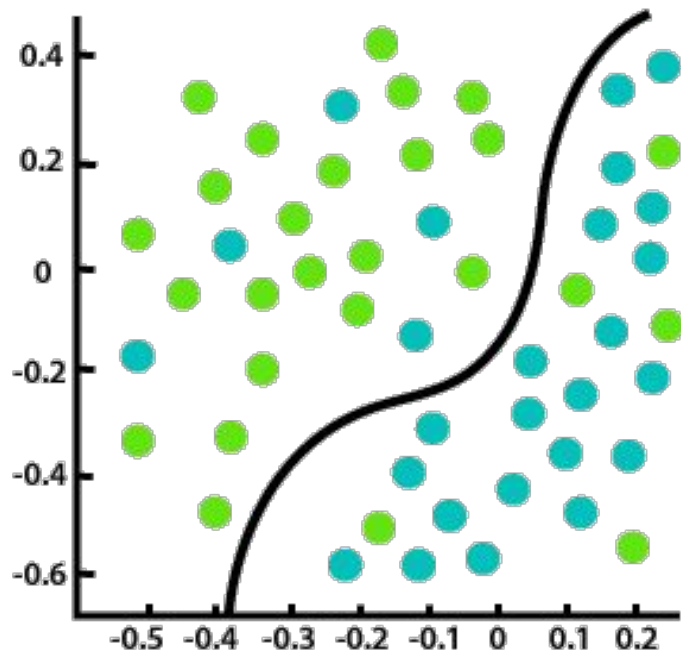
Classification

Clustering

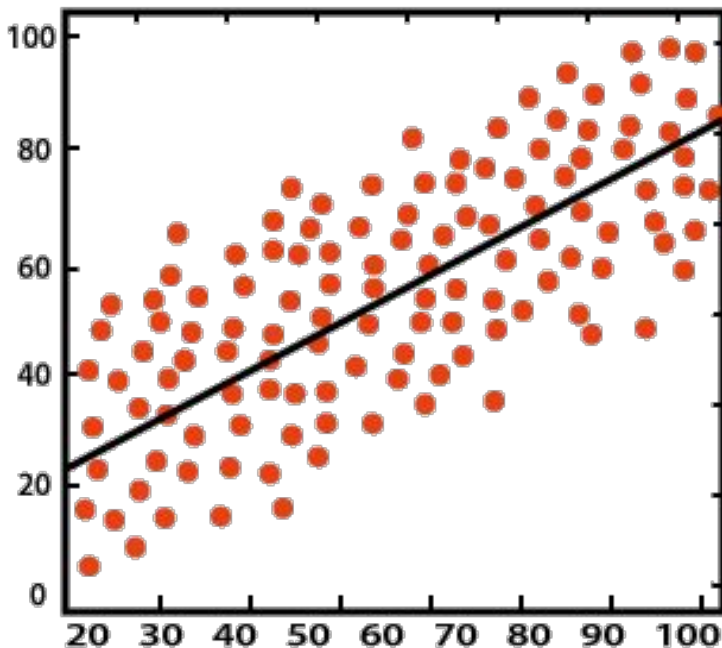
Association



Regression vs classification



Classification



Regression

Regression



What will be the temperature tomorrow?

84°



Fahrenheit

Classification



Will it be hot or cold tomorrow?

COLD

HOT

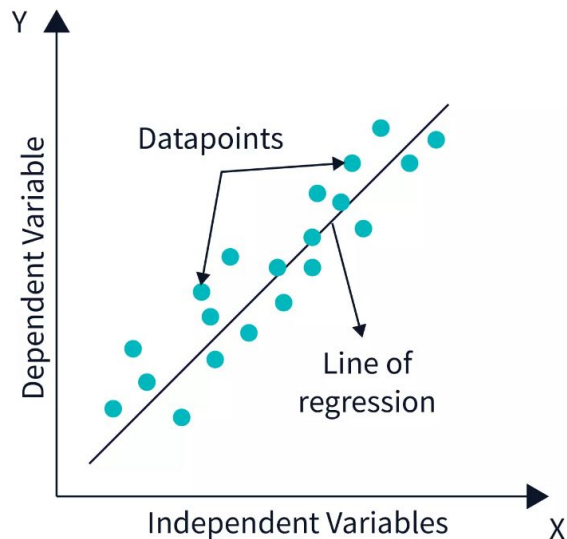


Fahrenheit

Simple linear regression

$$y = a x + b$$

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

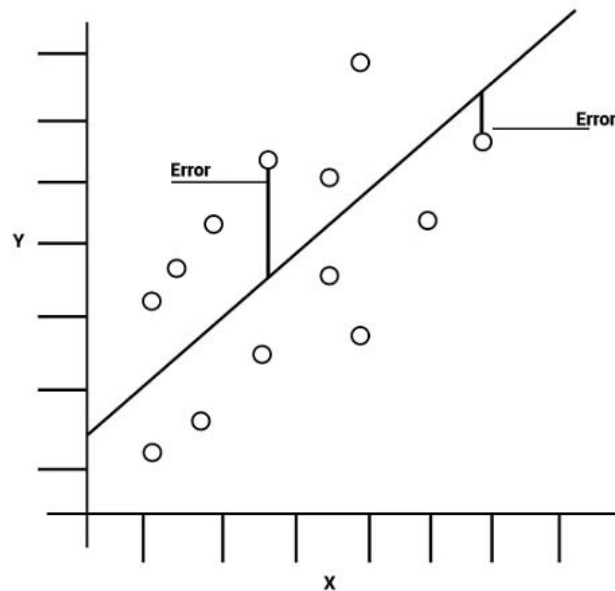


Does the line fit?

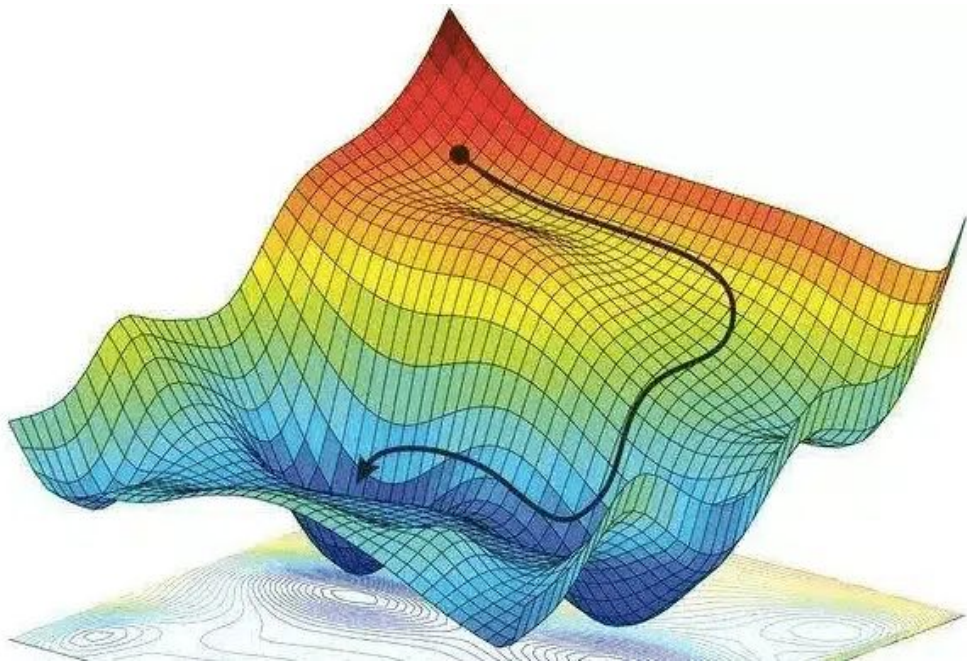
$$\text{MSE} = \overset{\text{Mean}}{\frac{1}{n} \sum_{i=1}^n} \left(\overset{\text{Error}}{Y_i - \hat{Y}_i} \right) \overset{\text{Squared}}{^2}$$

Does the line fit?

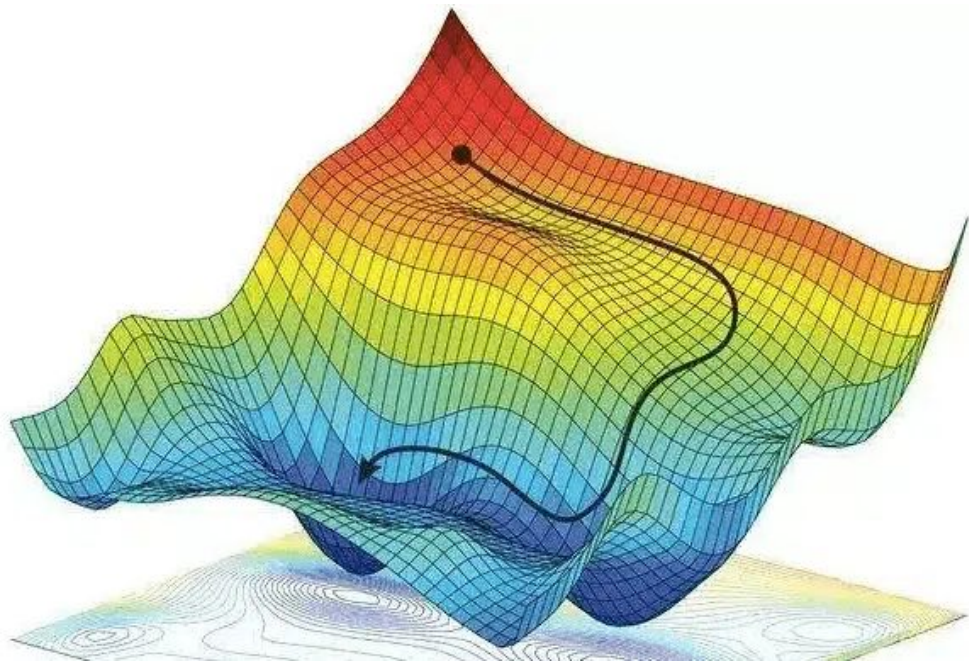
$$L(a, b) = \frac{1}{N} \sum_{i=1}^N (y_i - (ax_i + b))^2$$



How to use the error?



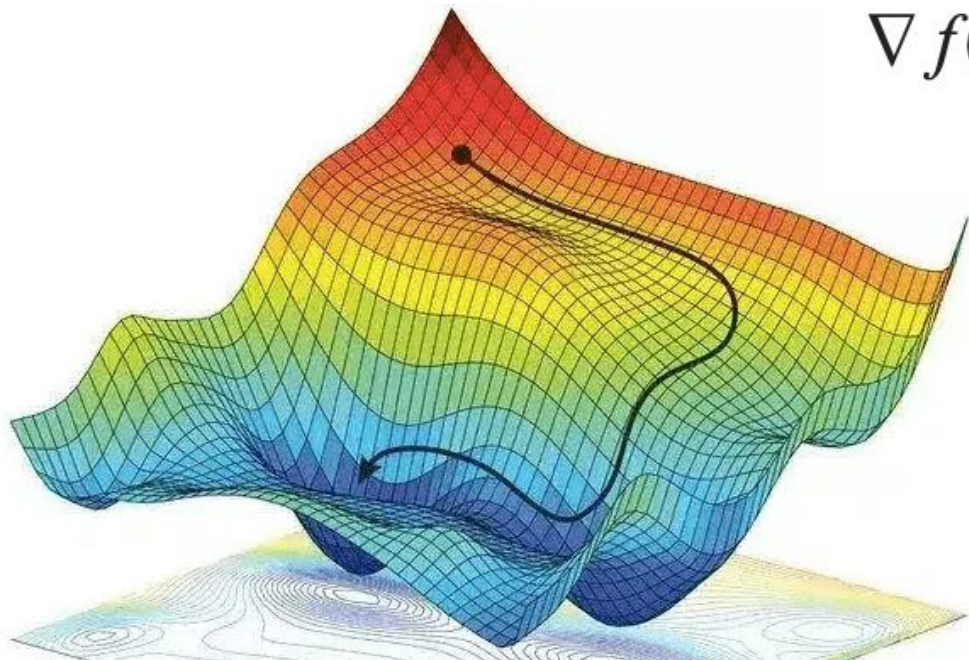
How to use the error?



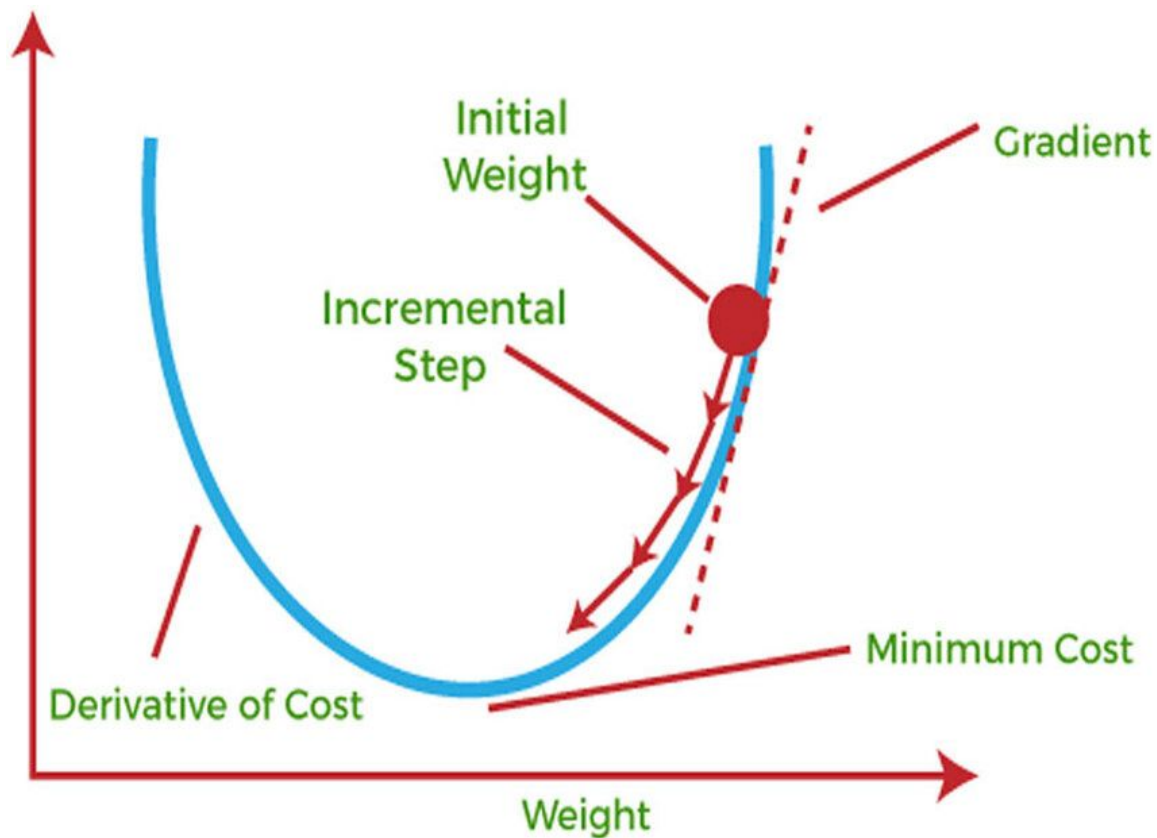
$$\nabla f = \left[\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right]$$

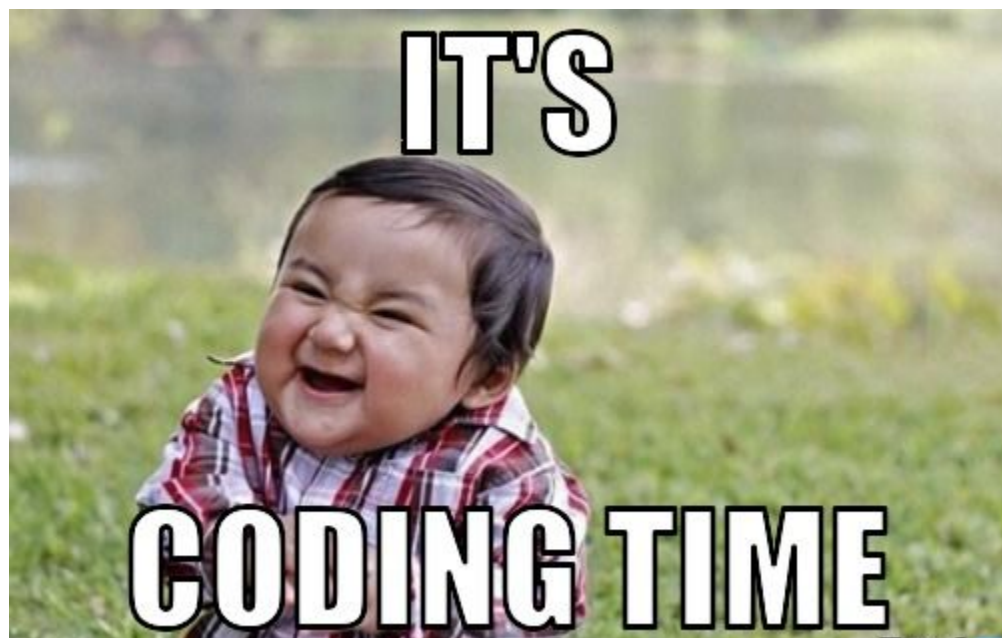
How to use the error?

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



How to use the error?





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

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

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

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

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

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

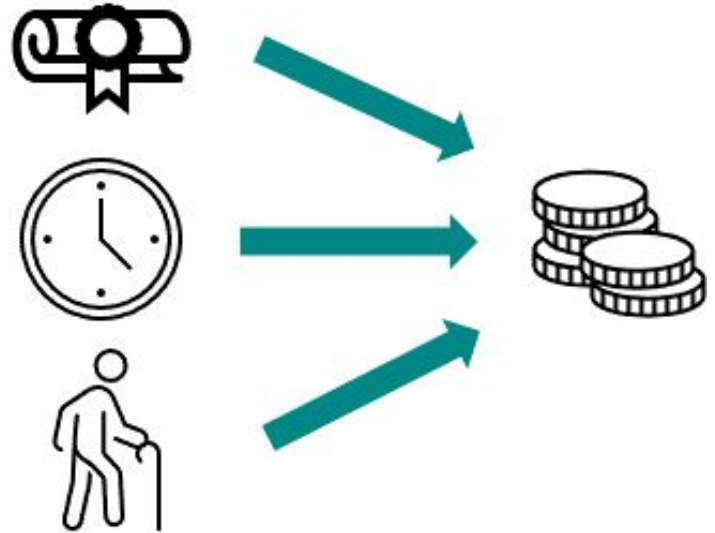
$$\frac{\partial L}{\partial b} = \frac{1}{N} \sum_{i=1}^N 2(y_i - (ax_i + b)) \cdot (-1) = -\frac{2}{N} \sum_{i=1}^N (y_i - (ax_i + b))$$

Multiple linear regression

Simple Linear Regression

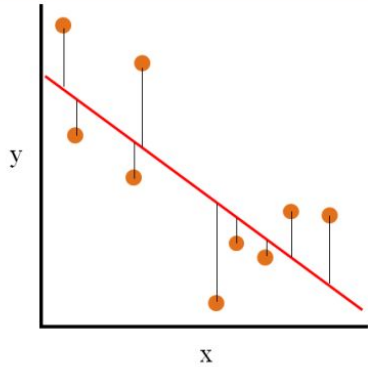


Multiple Linear Regression

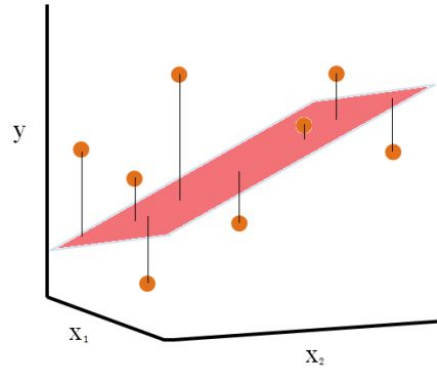


Multiple linear regression

Simple Linear Regression

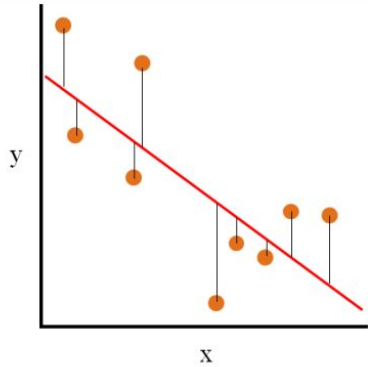


Multiple Linear Regression
(2 Independent Variables (x_1, x_2))

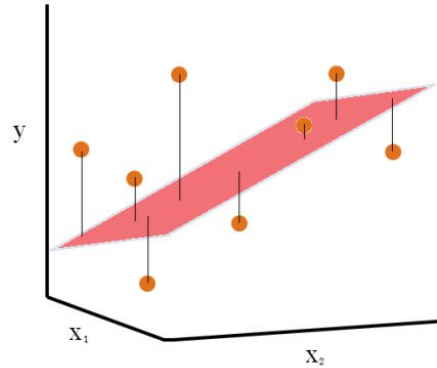


Multiple linear regression

Simple Linear Regression

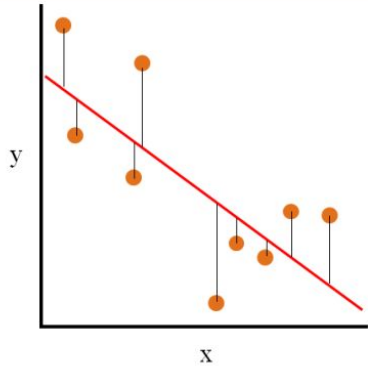


Multiple Linear Regression
(2 Independent Variables (x_1, x_2))



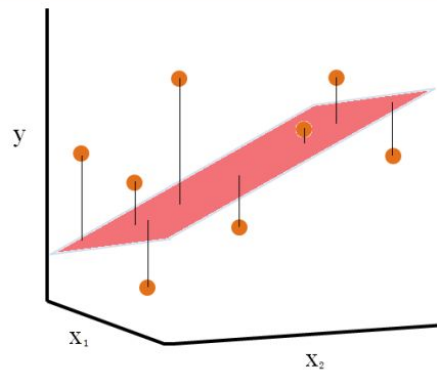
Multiple linear regression

Simple Linear Regression



Multiple Linear Regression

(2 Independent Variables (x_1, x_2))



$$y = b_0 + b_1 * x_1$$

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

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

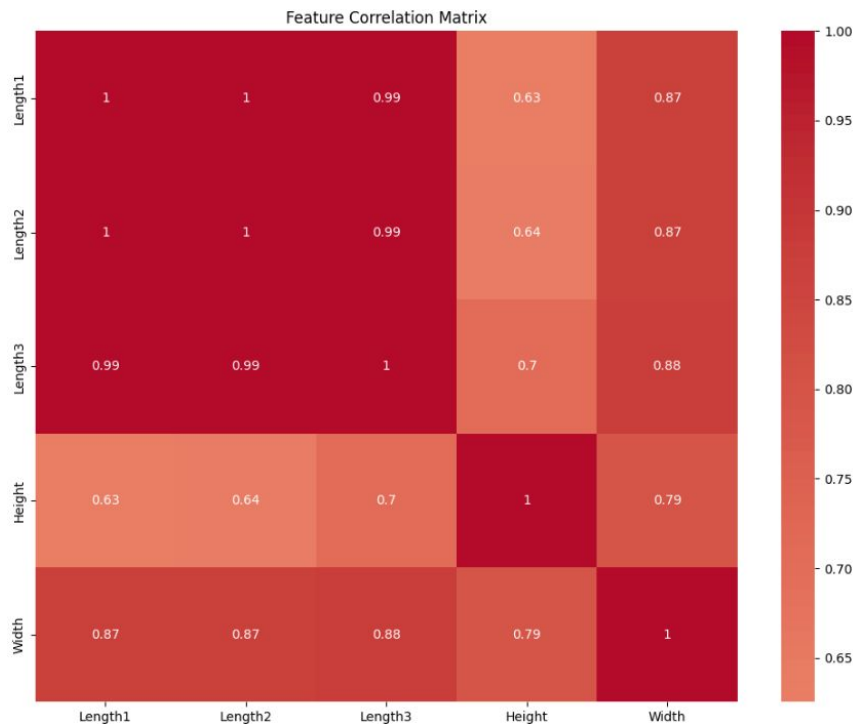
Lasso regression

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

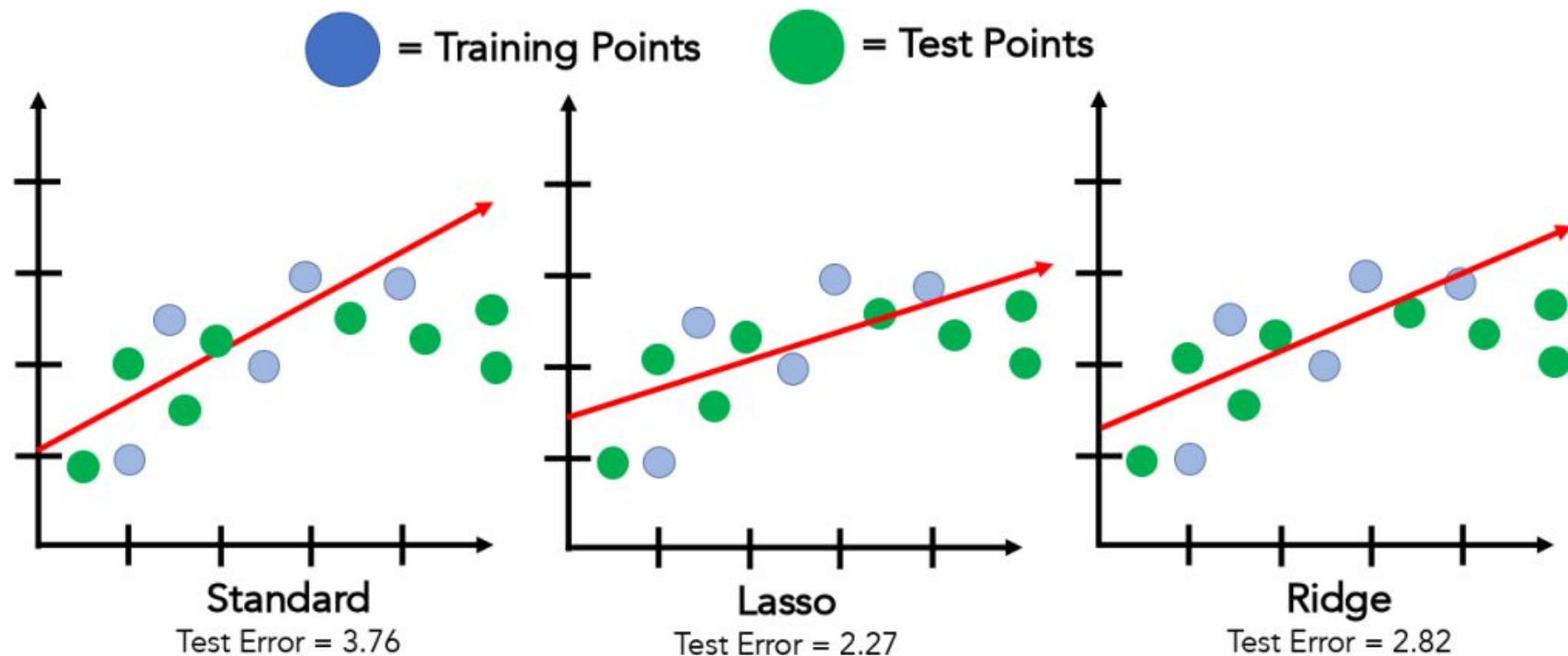


Let's code

Ridge regression

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

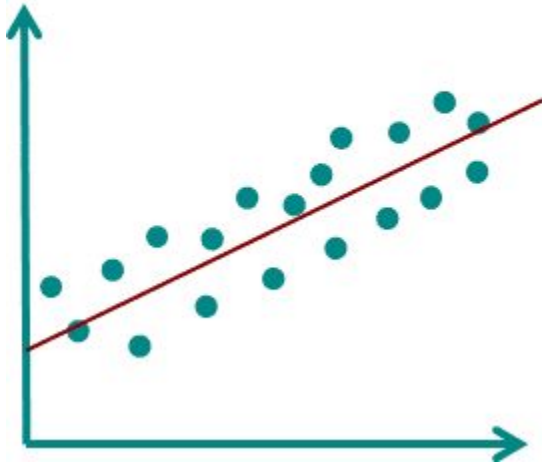
Ridge regression



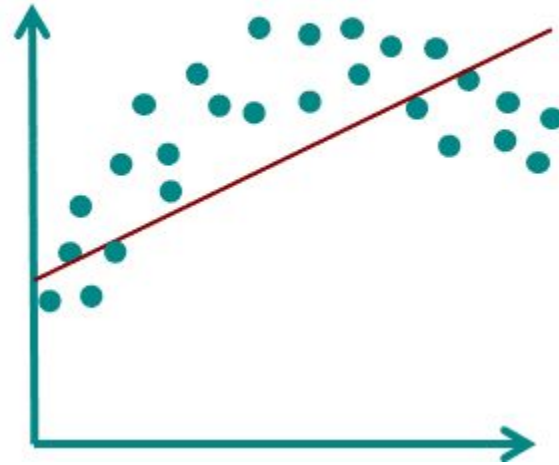
**Our Data was actually "parabolic" but we couldn't tell from the small training sample.*

More complex data

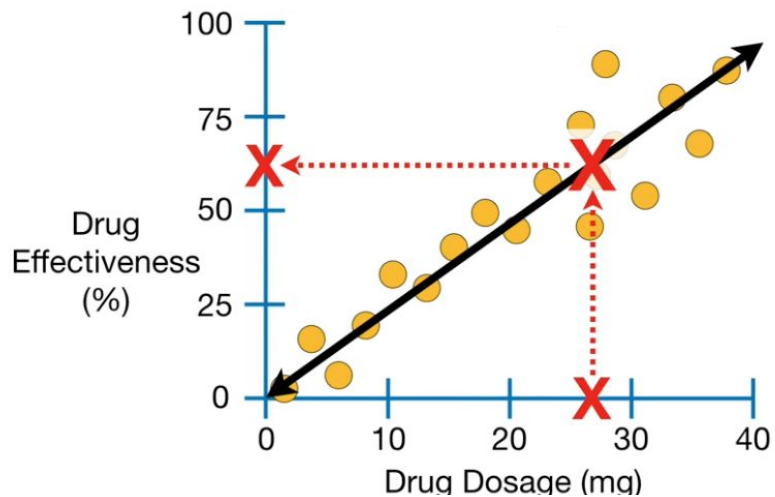
Linear



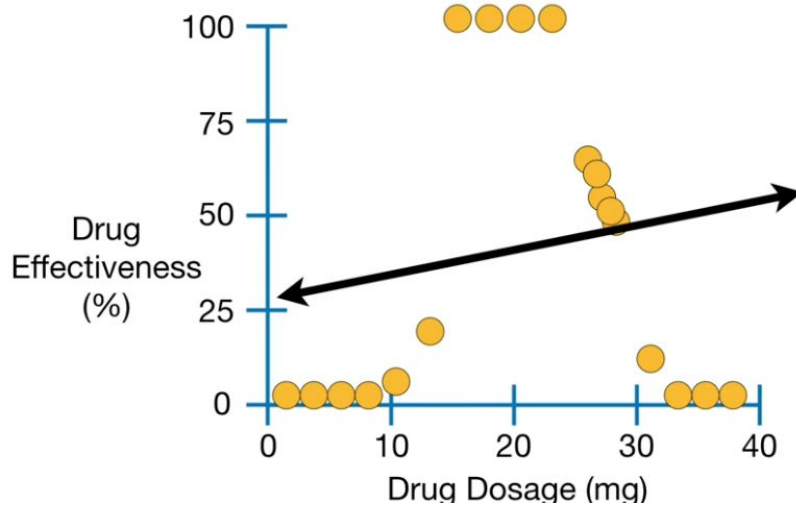
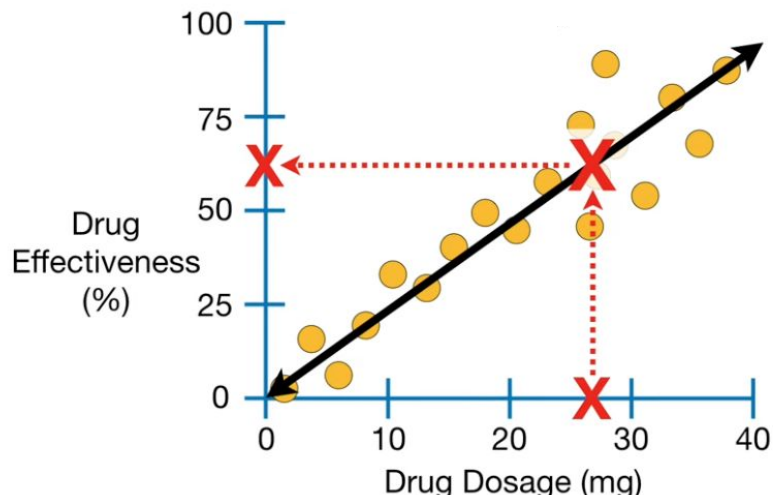
Non Linear



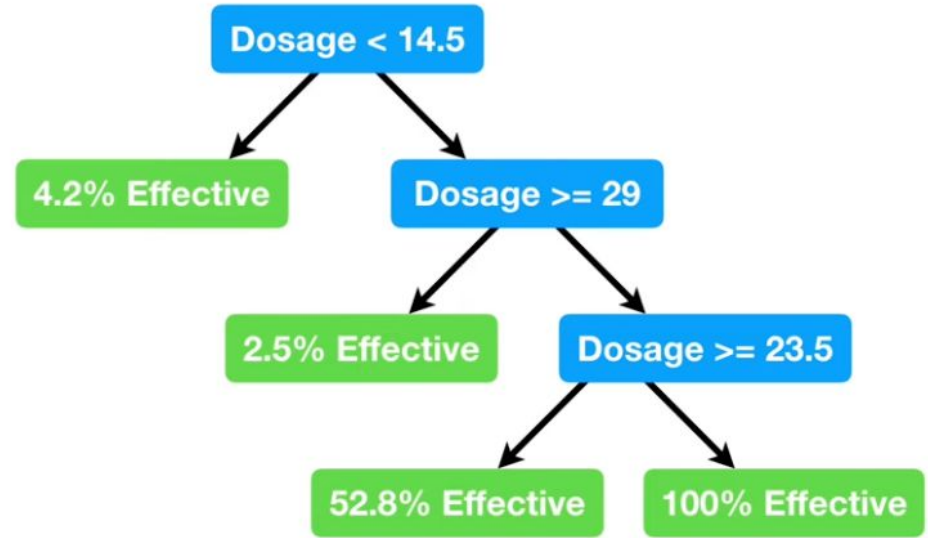
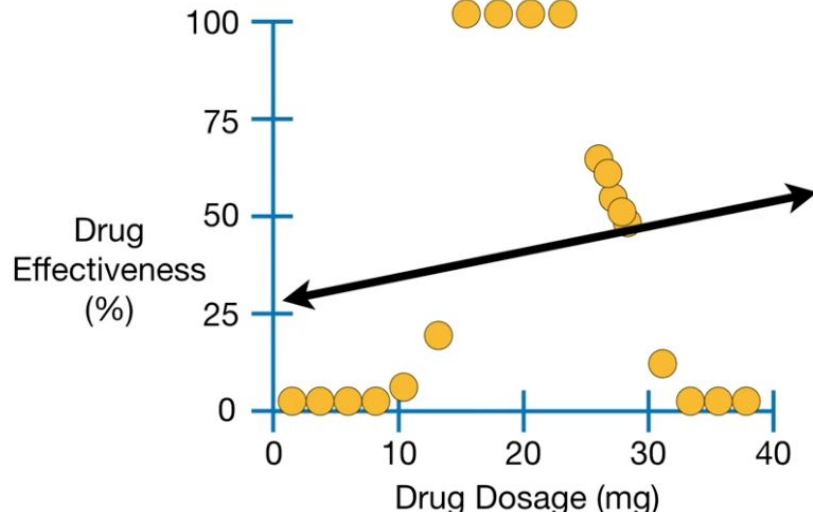
Regression trees motivation



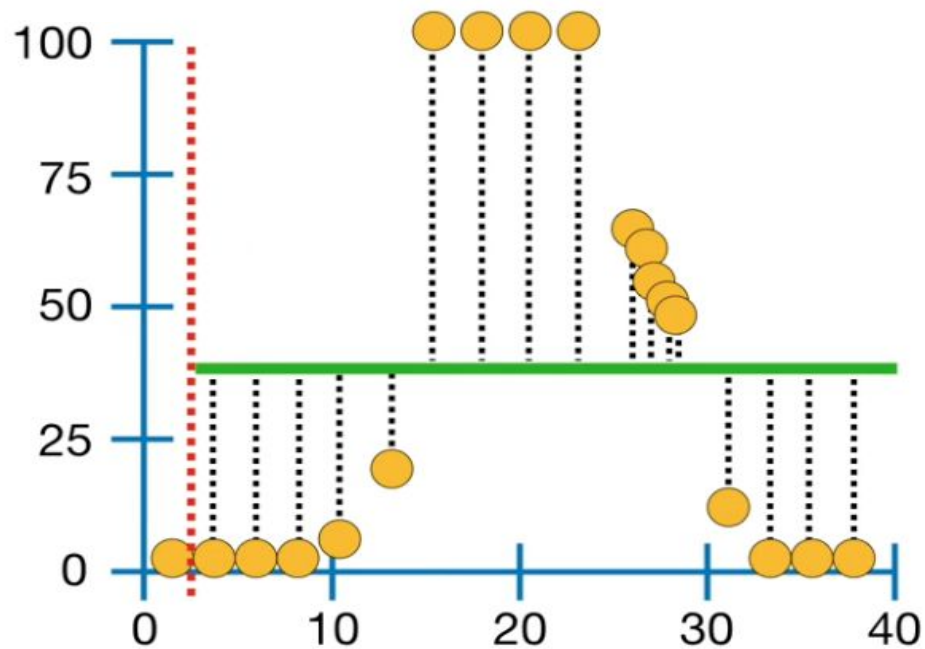
Regression trees motivation



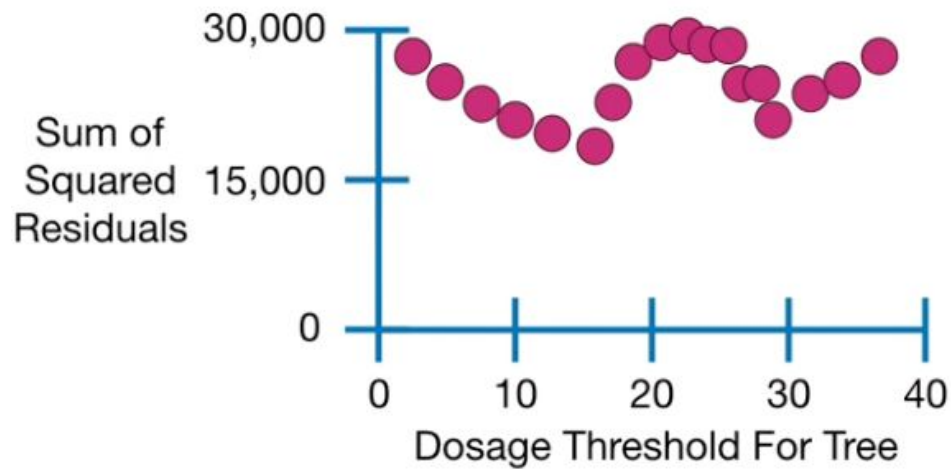
The trees way



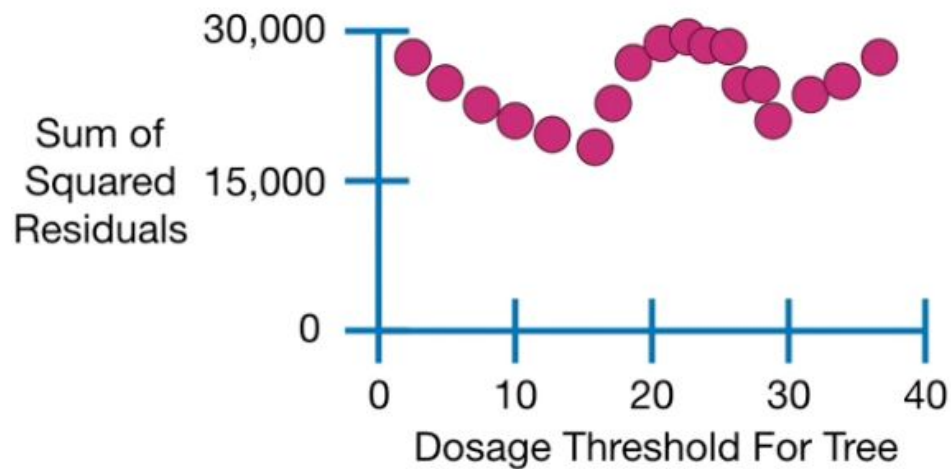
Squared residuals



Choosing a node



Choosing a node



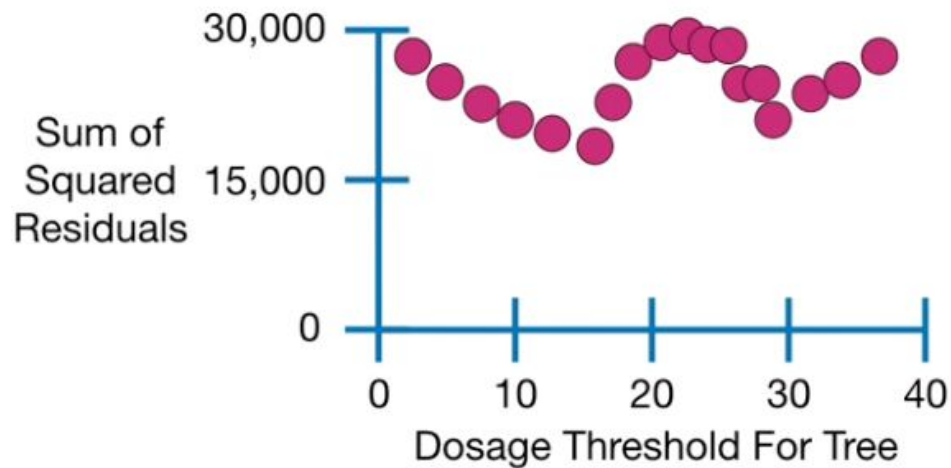
Dosage < 14.5



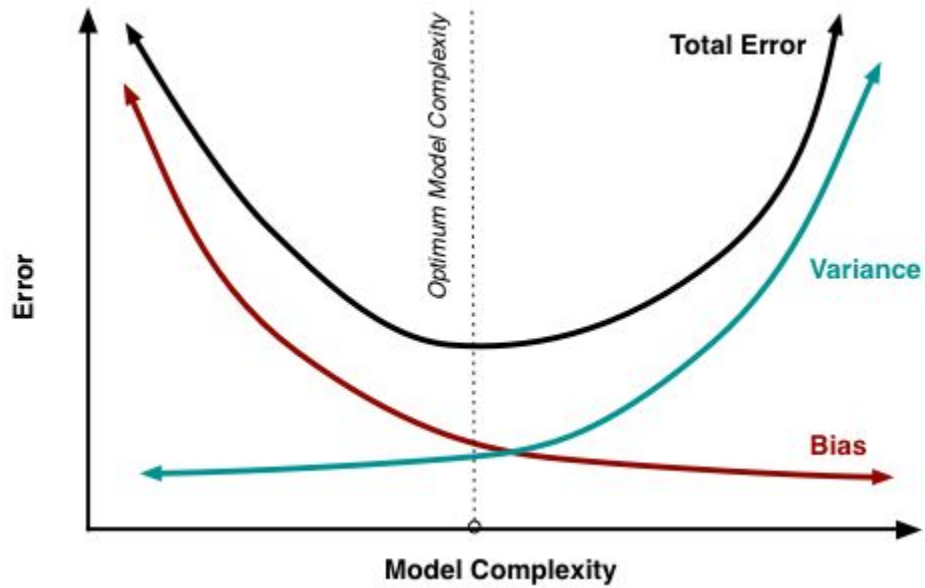
A diagram of a decision tree node. A blue rounded rectangle contains the text "Dosage < 14.5". Two black arrows originate from the bottom of the rectangle, pointing downwards and outwards to the left and right, indicating the branches of the tree.

Let's code

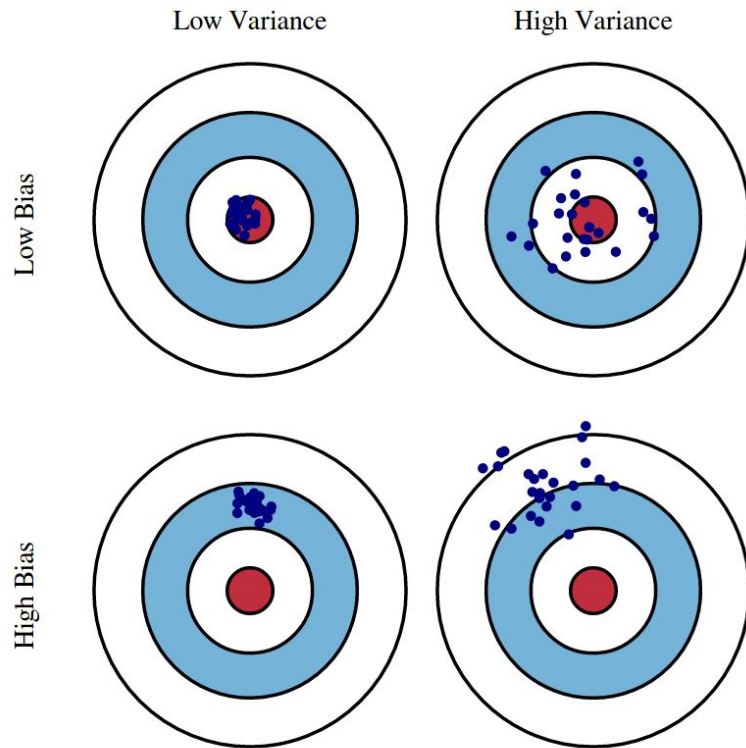
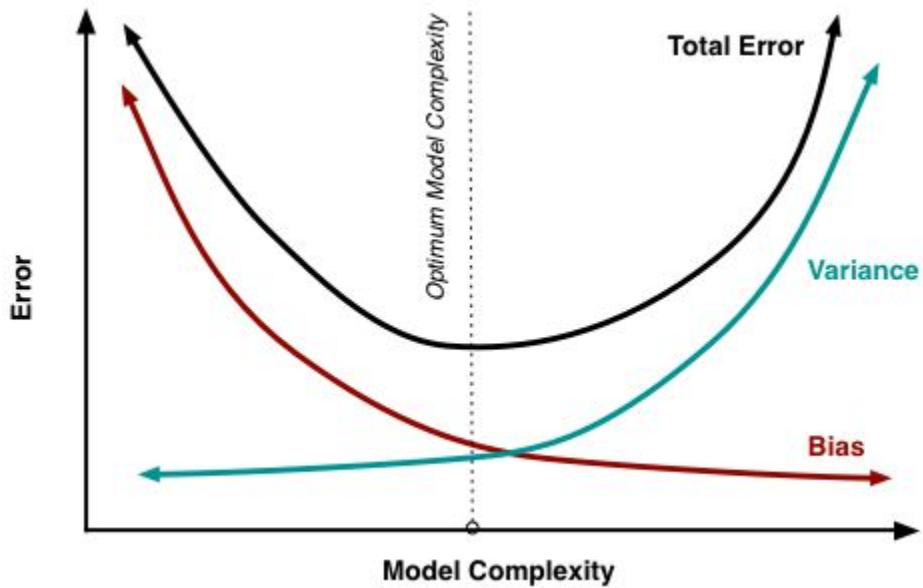
Possibility of overfitting



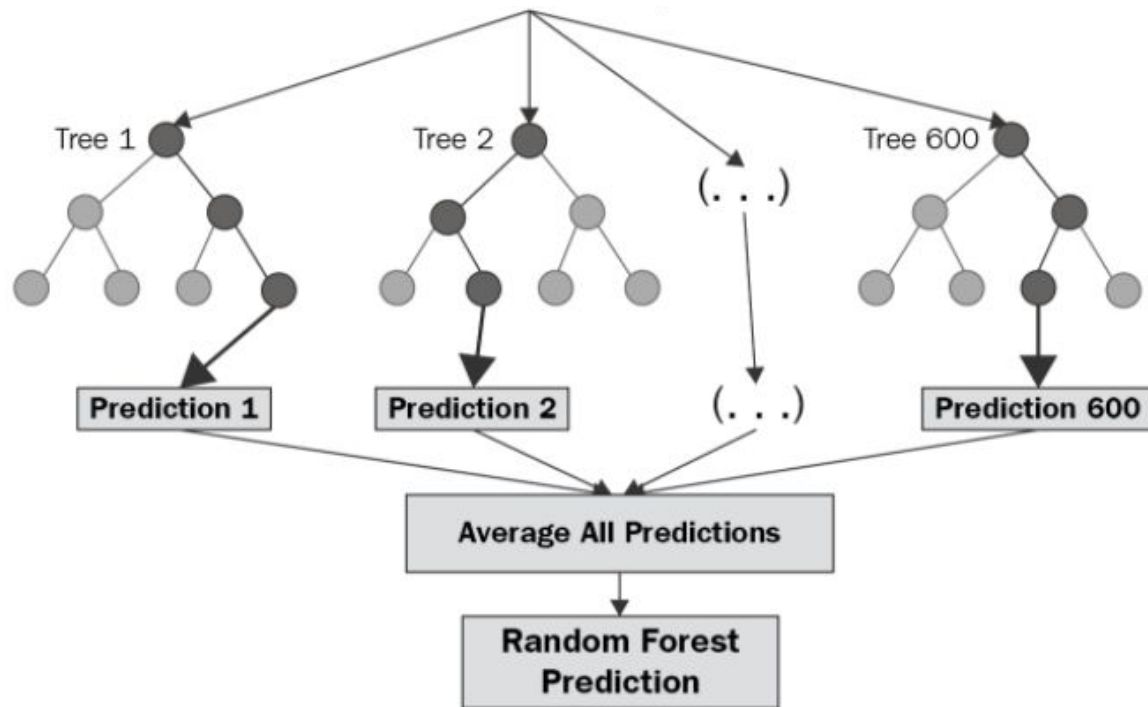
Bias - variance



Bias - variance

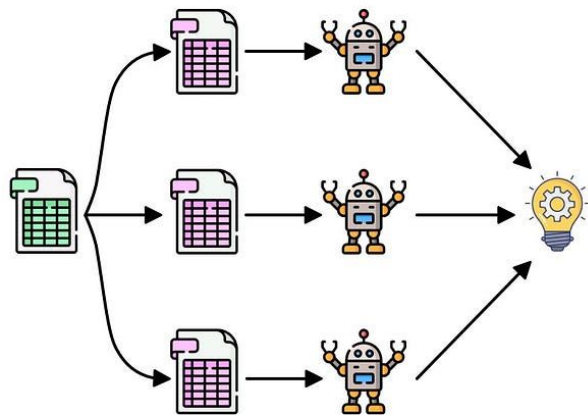


Random forest regressor



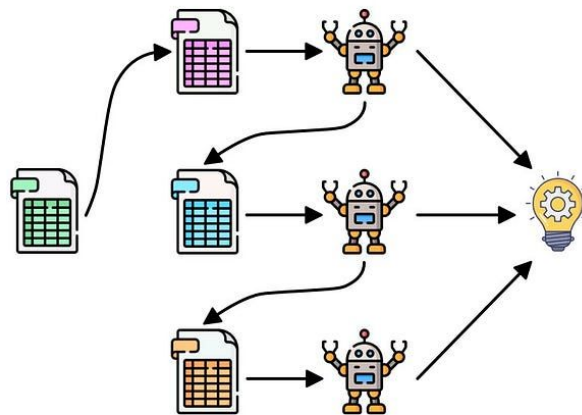
Bagging vs boosting

Bagging



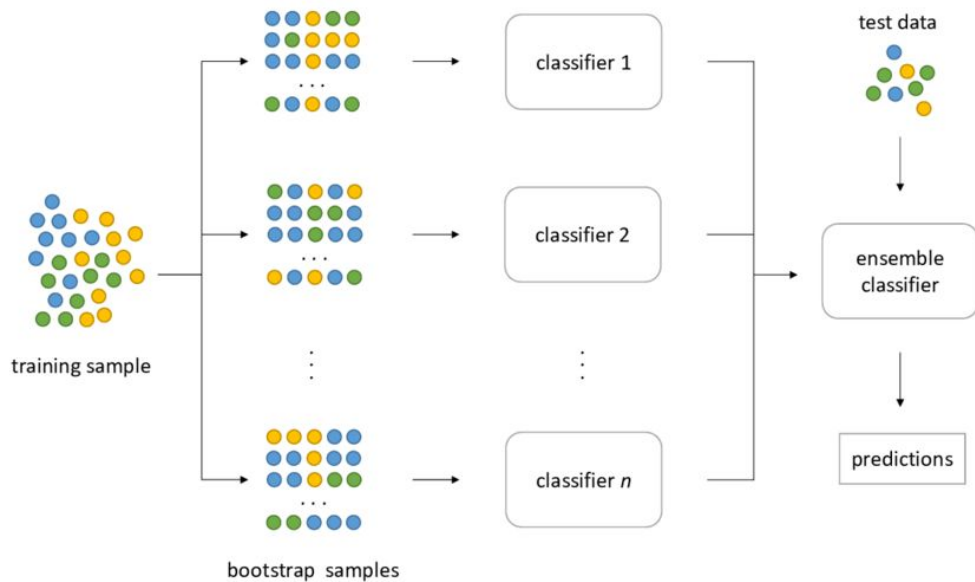
Parallel

Boosting

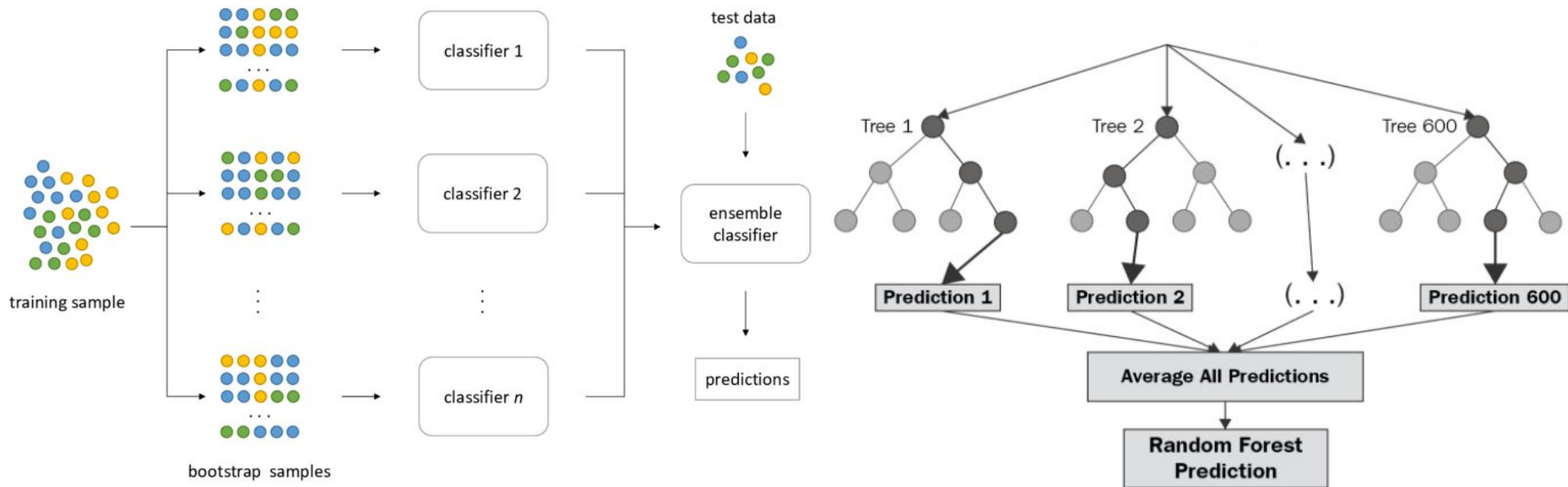


Sequential

Bagging vs bootstrapping



Bagging vs bootstrapping



Code example

A close-up photograph of a baby with light brown hair and blue eyes, looking directly at the camera with a grumpy or pouting expression. The baby is wearing a green and white long-sleeved shirt and is holding a small amount of sand in their right hand. The background is a blurred beach scene with sand and water.

THANK YOU!

YOU ROCK!