

Unsupervised Learning

DIMENSIONALITY REDUCTION

PCA
t-SNE
Factor Analysis

CLUSTERING

AD, NLP may be solved with clustering.

Algorithms

KMeans
DBScan
Hierarchical

MACHINE LEARNING

Reinforcement Learning

AD, NLP, RS may be solved with reinforcement learning

Collaborative Filtering
Content Filtering
Neural Networks (DL)

Supervised Learning

REGRESSION

• Continuous Target
[TSA, AD may be regression probs.]

Linear Regression
GLM
Polynomial Reg.
SVR
Decision Tree Regr.
Neural Networks

CLASSIFICATION

- Discrete, Categorical Target

[TSA, NLP, AD may be classification problems]

Logistic Regression
Decision Tree
Random Forest
SVC
KNN
Neural Networks

SUB-METHODOLOGIES

TSA: time series analysis

AD: anomaly detection

NLP: natural language processing

RS: recommendation system

Predict
Continuous
Variable

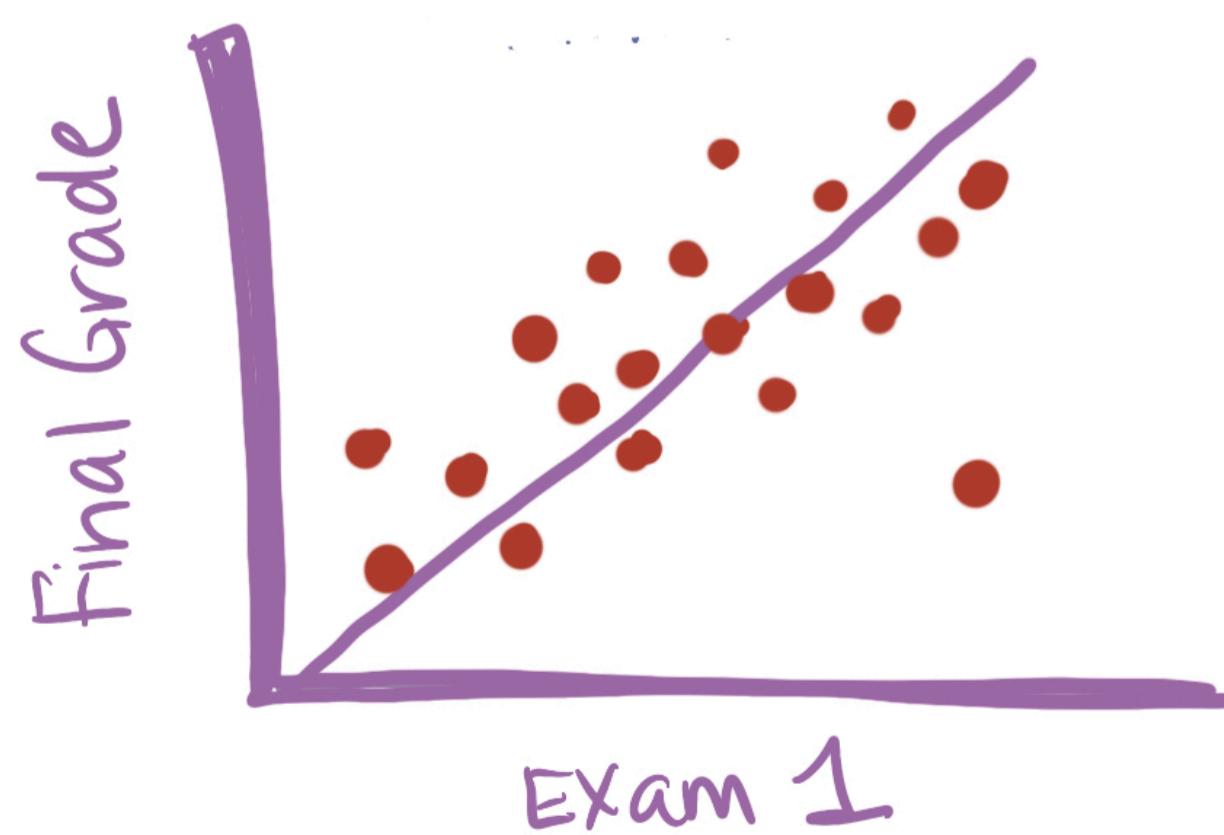
REGRESSION

VS. CLASSIFICATION

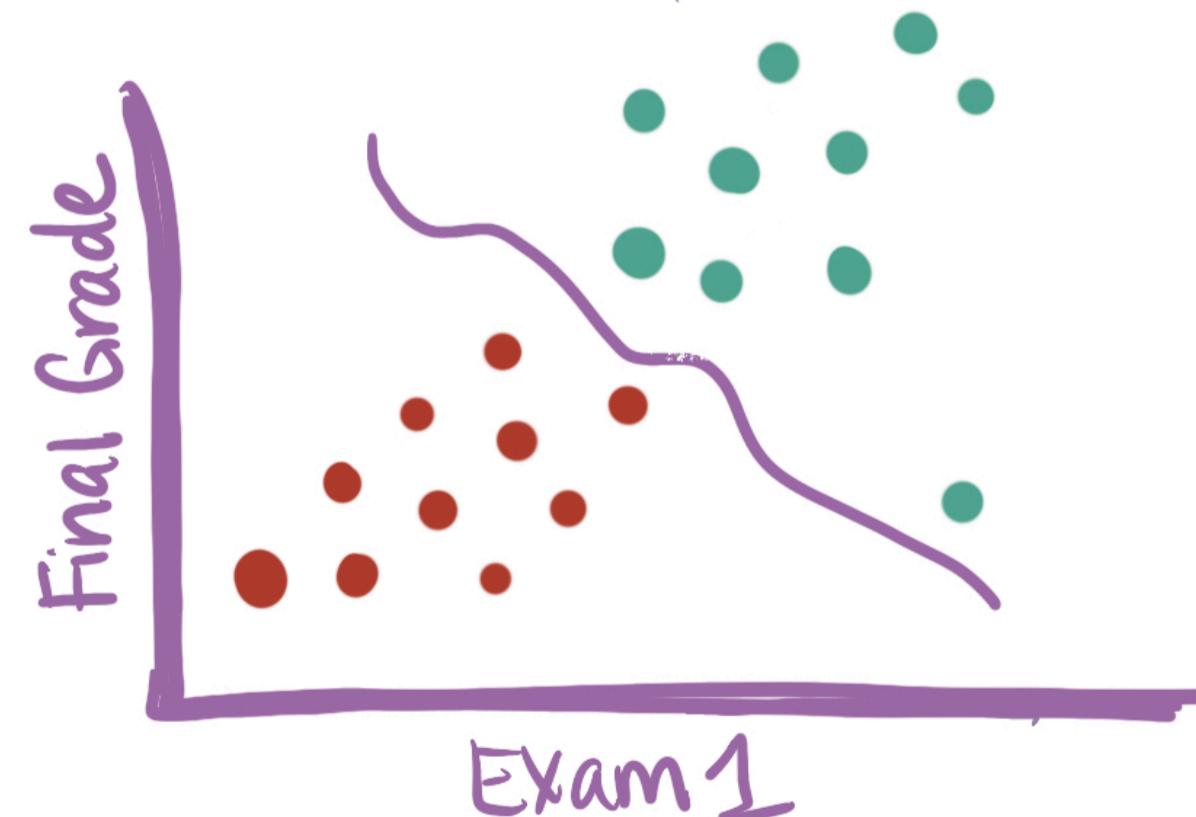
Predict
Categorical
Variable

How will each Student perform
in the course?

Regression: Predict
Final Grade



Classification: Predict
Pass or Fail



REGRESSION types

SIMPLE

$$y = b_0 + bx$$

$y = \text{SAT}$

b_0

$x = \text{gpa}$

$$\text{SAT} = 400 + 500 \cdot \text{gpa}$$

$b_1 = 500$

$b_0 = 400$

UNIVARIATE

MULTIPLE

$$y = b_0 + b_1x_1 + \dots + b_nx_n$$

$y = \text{SAT}$

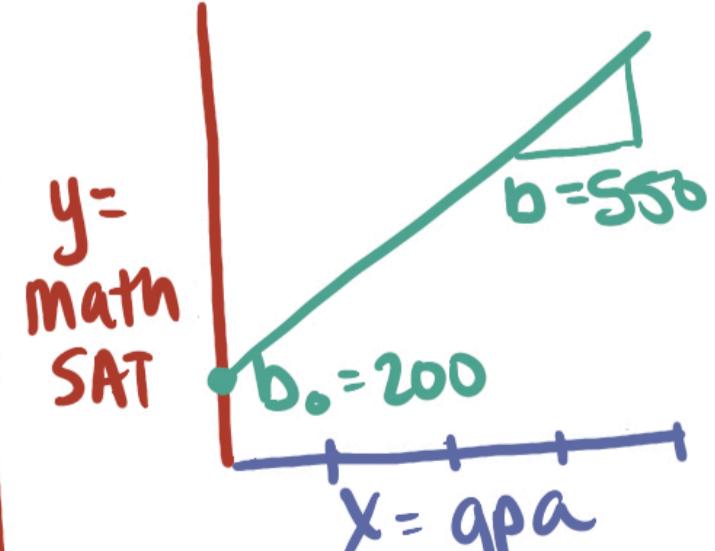
$x_1 = \text{INCOME}$

$x_1 = \text{GPA}$

$$\text{SAT} = 400 + 500 \cdot \text{GPA} + .005 \cdot \text{INCOME}$$

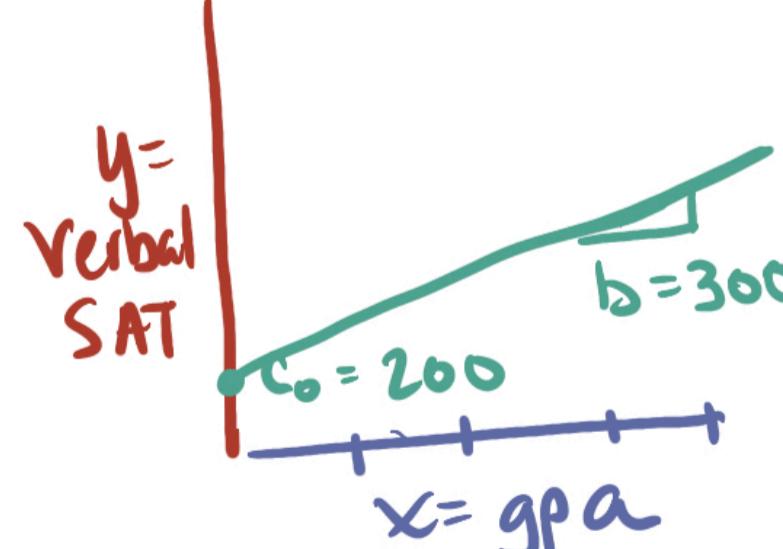
$$y_1 = b_0 + bx$$

$$y_2 = c_0 + cx$$



$\text{Math SAT} =$

$$200 + 550 \cdot \text{gpa}$$



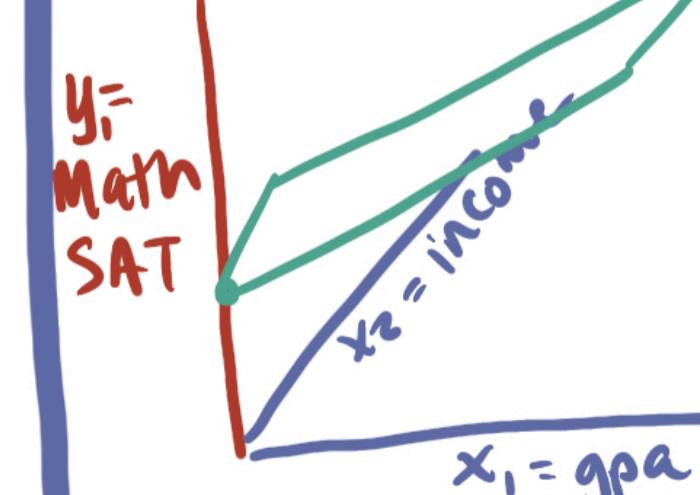
$\text{Verbal SAT} =$

$$200 + 300 \cdot \text{gpa}$$

MULTIVARIATE

$$y_1 = b_0 + b_1 + \dots + b_n$$

$$y_2 = c_0 + c_1 + \dots + c_n$$



$\text{Math SAT} =$

$$200 + 550 \cdot \text{gpa}$$

$$+ .003 \cdot \text{income}$$



$\text{Verbal SAT} =$

$$200 + 250 \cdot \text{gpa}$$

$$+ .002 \cdot \text{income}$$