

HUDK 4050: CORE METHODS IN EDM

In the news

M

**Why becoming a data scientist is NOT
actually easier than you think**

**I Built A Bot To Apply To Thousands Of Jobs
At Once—Here's What I Learned**

**IBM's Quest To Design The "New
Helvetica"**

**VR Gives Student-Teachers a
Taste of the Classroom**

A University at Buffalo virtual reality training program lets
educators deepen their pre-service experiences.



VB

**An AI god will emerge by 2042 and write its own bible.
Will you worship it?**

**Help! This Edtech Company Says It Uses AI. (What
Does That Mean? What Should I Ask?)**



**When Students Can't Get
Broadband, Career Success
Proves Elusive**

Research finds that students in rural Florida lag behind in access
to connectivity and educational opportunities.



BBC

100 Women: Where do women outnumber men in science?

edscoop

**25 innovators and innovations
that changed education over the
last 25 years**

The Washington Post
Democracy Dies in Darkness

**Report finds fewer new
international students on U.S.
college campuses**

HUFFPOST

abc NEWS

**SEAL candidate injured in study-session slapping
incident**

QUARTZ
Africa

**Digital technology can help reinvent basic
education in Africa**

**Alcohol Can Help You Speak A Foreign
Language More Fluently**

NIBLETZ

EdTech Everywhere | Startups Everywhere Else

**The #1 Problem in EdTech? It's Not
What You Think!**

Events

Event	Date	Time	Location	URL
KPMG Technology Career Forum	November 14	1:00pm	Online	https://app.brazenconnect.com/events/X1r8O?utm_medium=Marketing&utm_source=Data+Science+Association#!eventLanding;eventCode=X1r8O
Data Science Case Study with R	November 15	11:00am	Online	http://info.rstudio.com/m000CS20a0jY3SajXWW0NN0
Cybernetics Conference	November 15-18	All day	Prime Produce 56 W 54th St	http://cybernetics.social/
TCLA Happy Hour	November 15	7:00pm	Hex Cafe 2871 Broadway	
Tableau Workshop Columbia DS Society	November 14	1:30-3:30 pm	Lerner Hall Broadway Room	

Opportunities

Letters to Betsy (volunteer)

email: laura.davis@nyu.edu

Data & Society
Office Manager

<https://datasociety.net/blog/2017/10/30/now-hiring-office-manager/>

Prediction

Machine Learning

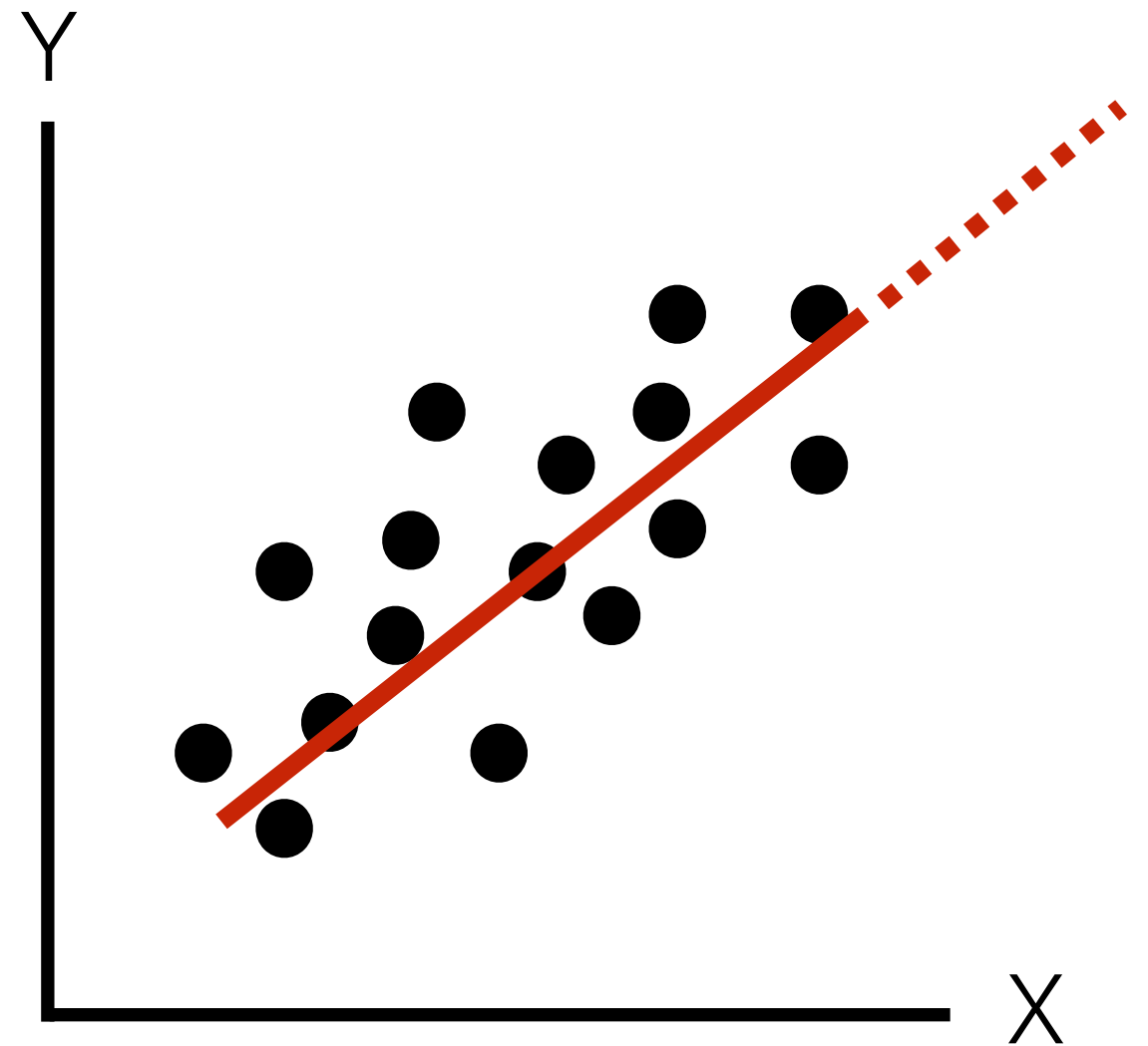
- Samuel built a computer program that could play checkers
- Recognized when it made a mistake and avoided that mistake again (“learning” through prediction)
- Built a tree of all possible moves for a given board
- Maximized a function that described the probability of winning
- Within three weeks it beat Samuel
- In 1956 it beat the world checkers champion



Arthur Samuel, 1952

Prediction

- Cuts to the ❤️ of the difference between machine learning and ed statistics
- Characterize data vs. predict the future



Terminology

Supervised Learning: Techniques used to learn the relationship between independent attributes and a designated dependent attribute (the label). (Have labelled data available that the machine can learn from)

For example: Have images labelled as dog, cat, etc, machine must learn the labels

Unsupervised Learning: Learning techniques that group instances without a pre-specified dependent attribute.

For example: Clustering algorithms

Terminology

Classification: Mapping an unlabeled instance to a discrete class by a classifier

Example: Identify a student as likely to drop out or not based on demographic data

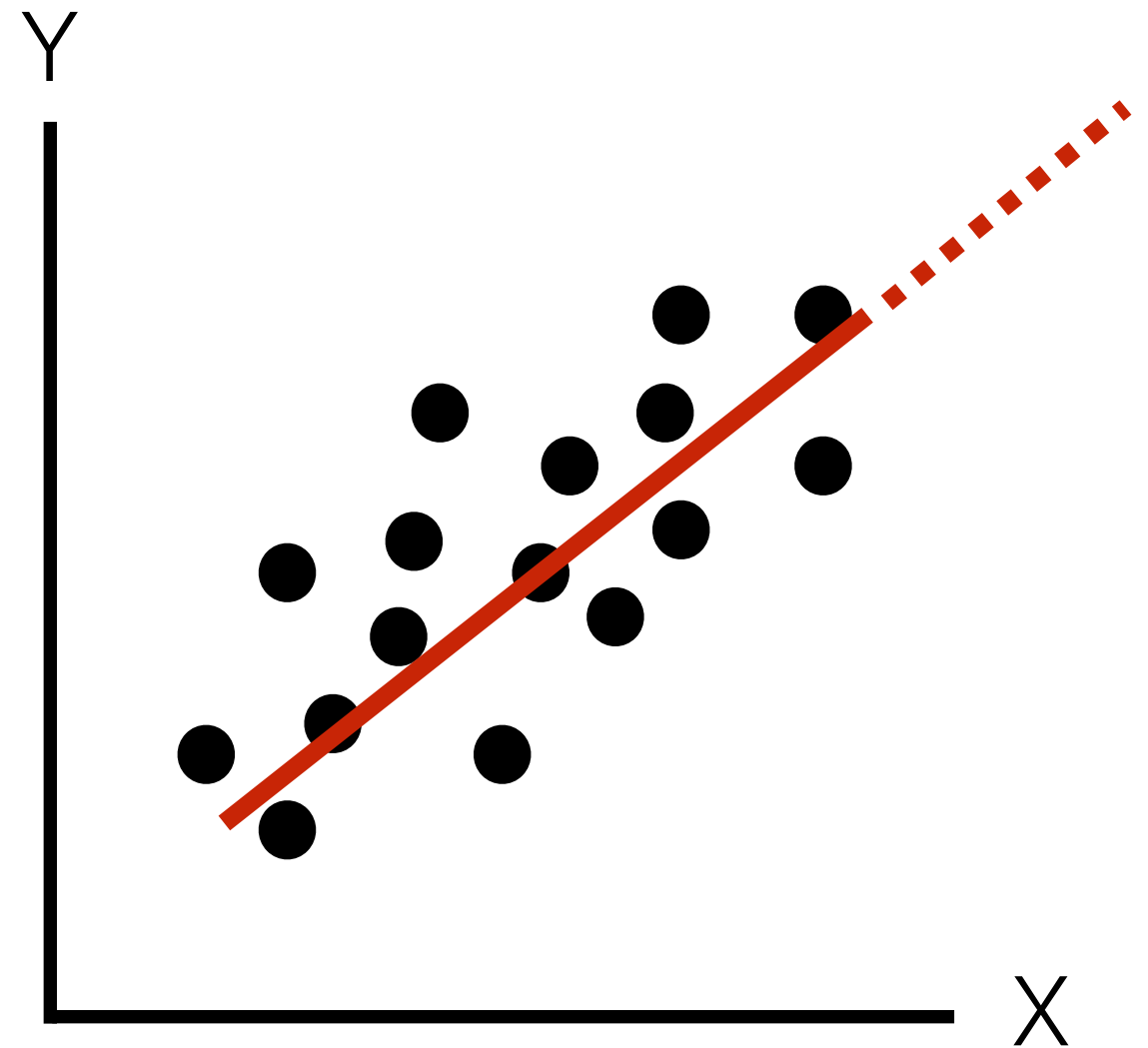
Regression (*as a form of classification*): Mapping from an unlabeled instance to a value within a continuous range

Example: Identify a student as having a math test score of 70 based on online assignment performance

Training Sets: Either supplied by a previous stage of the knowledge discovery process or from some external source

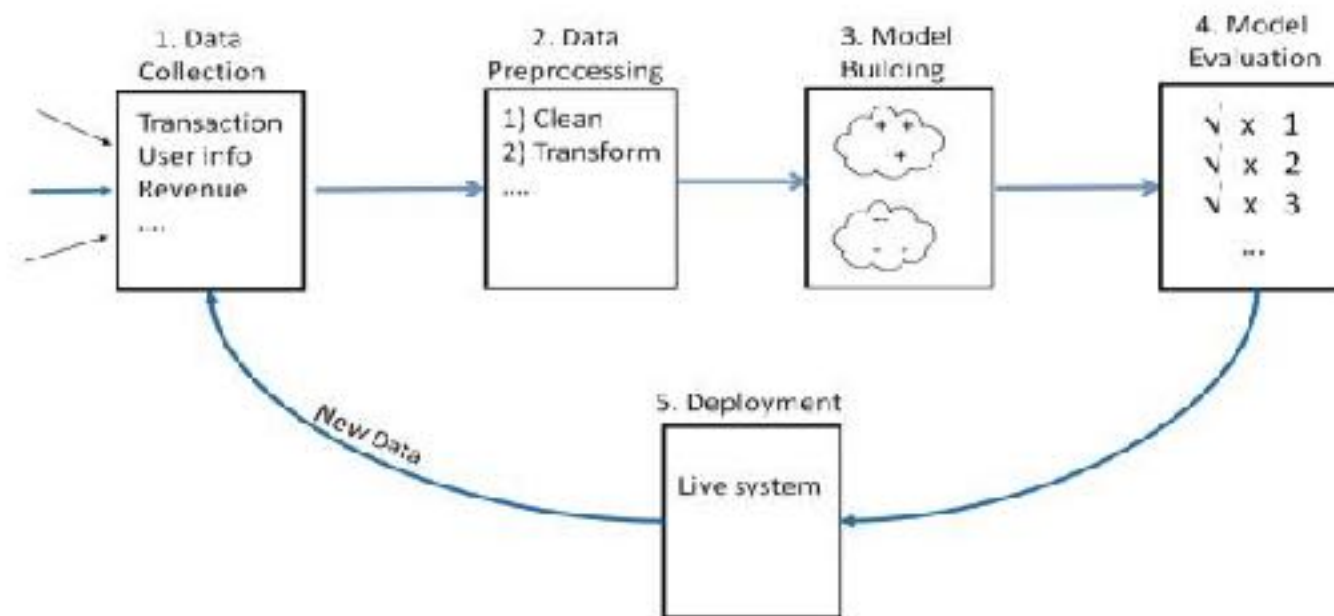
Regression

- In Ed Stat = OLS
Regression/Logistic
Regression (characterize)
- In ML = Mapping from
unlabeled instances to a
value within a continuous
range (future)



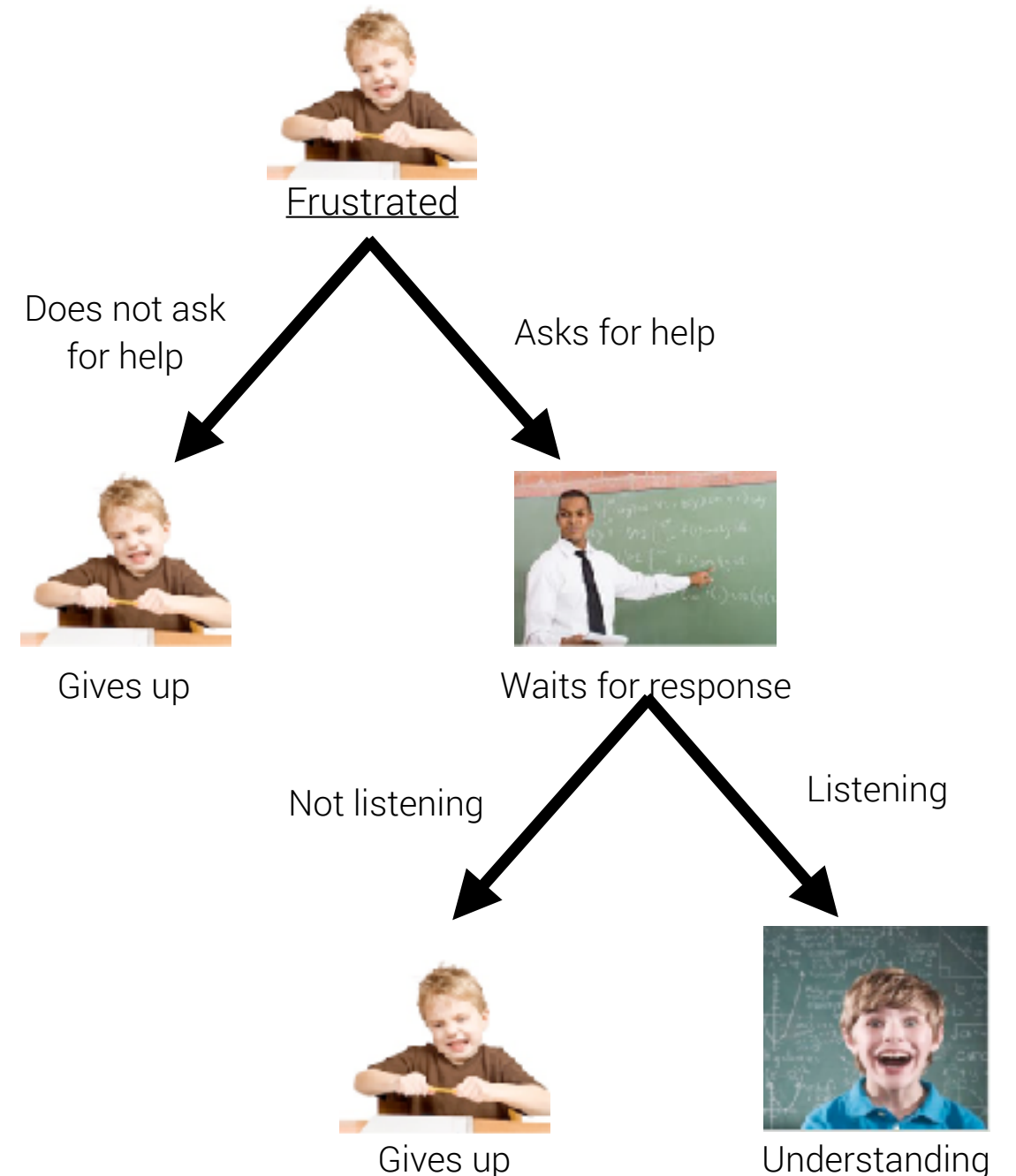
Machine Learning Framework

Machine Learning Process



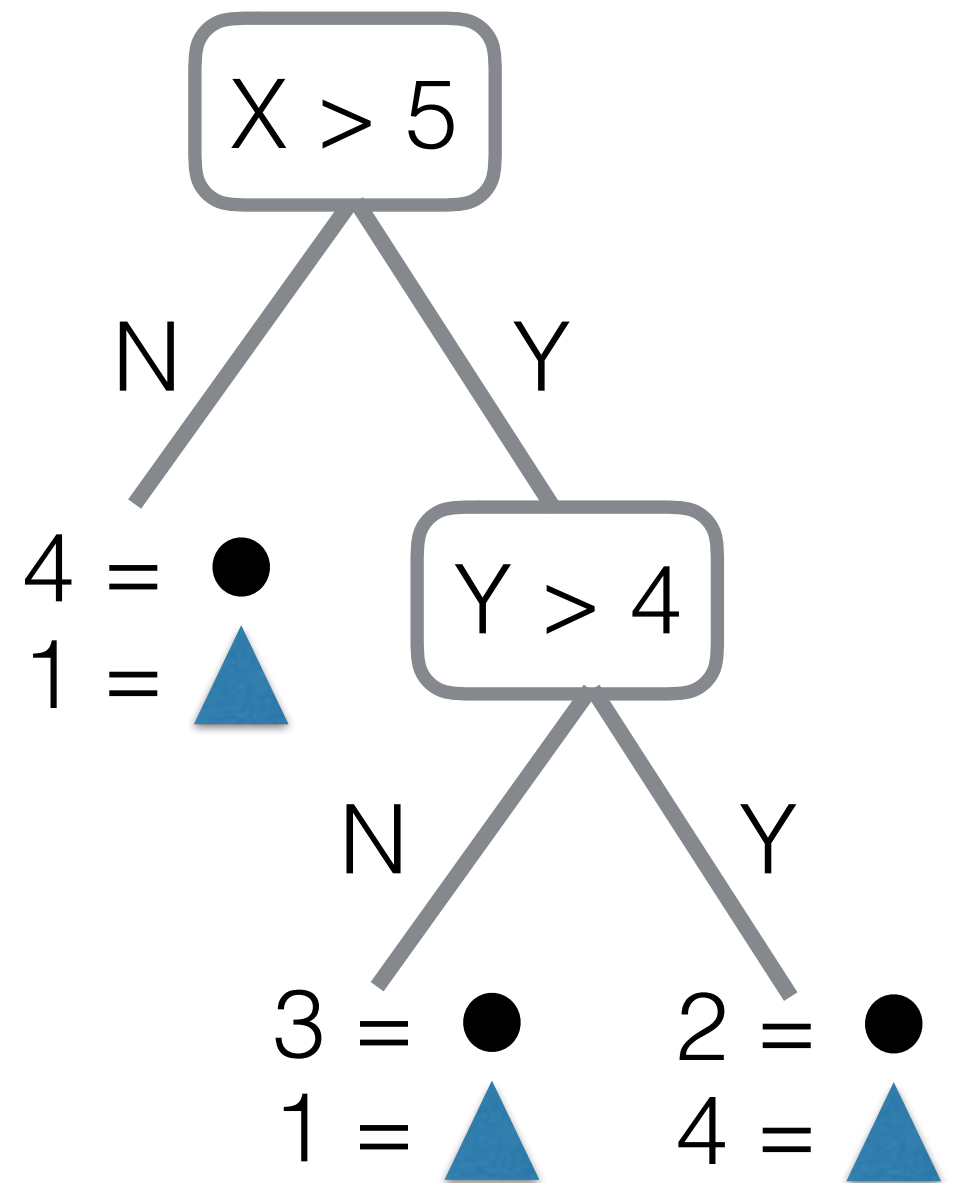
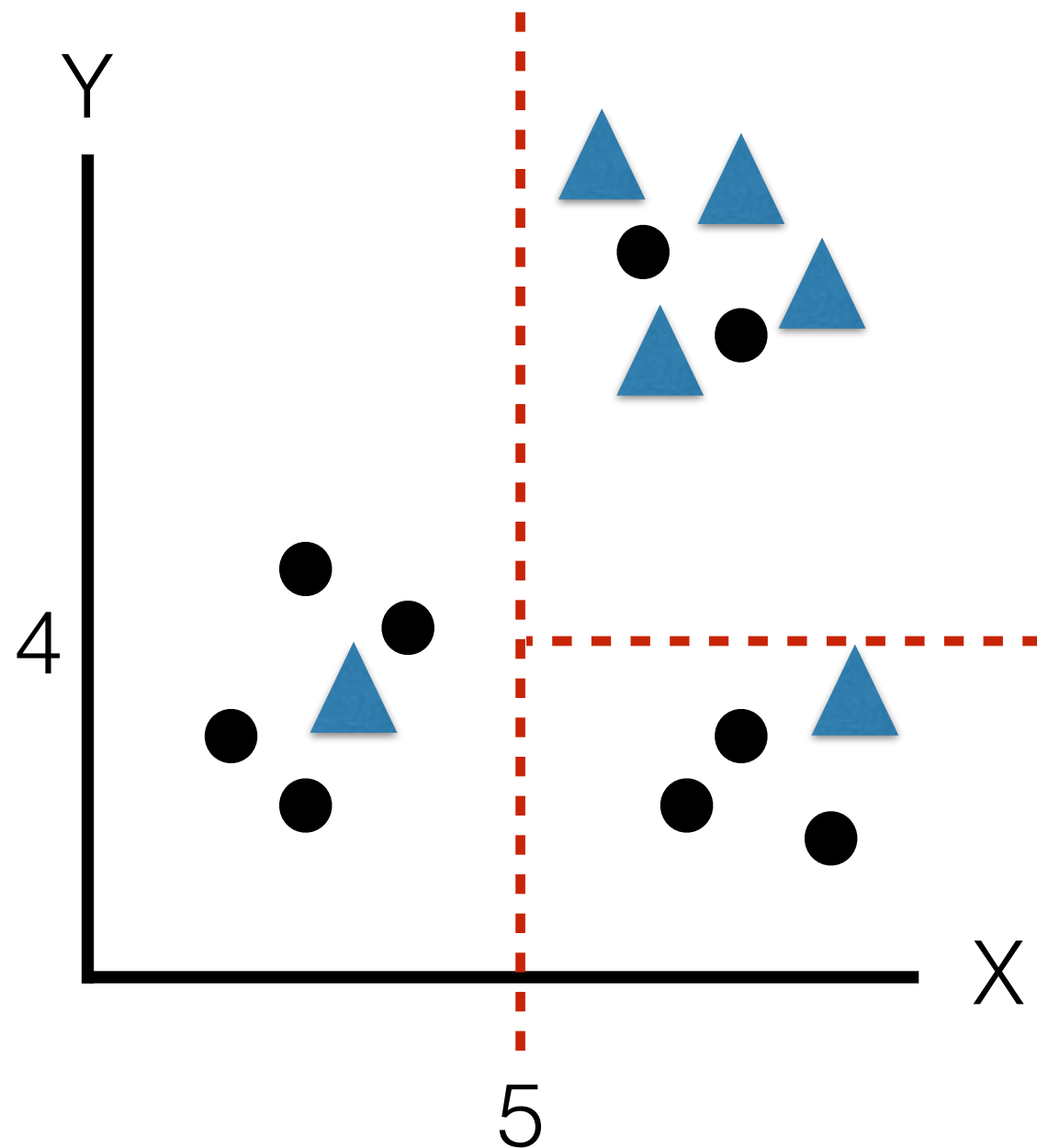
Classification Tree

- Decision tree
- Map observations (branches) onto classes (leaves)
- Tree describes the data but can be used classification
- EG: student states = leaves, student actions = branches



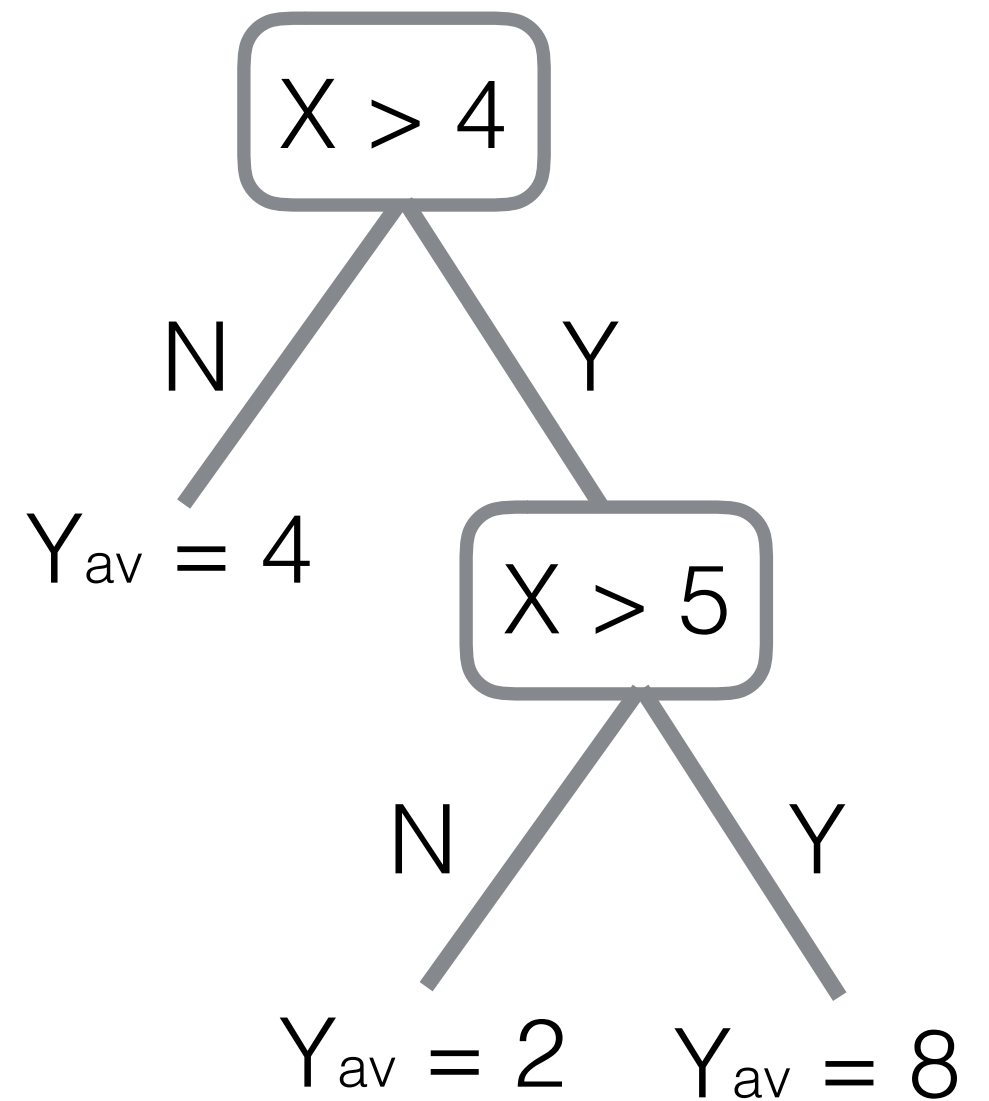
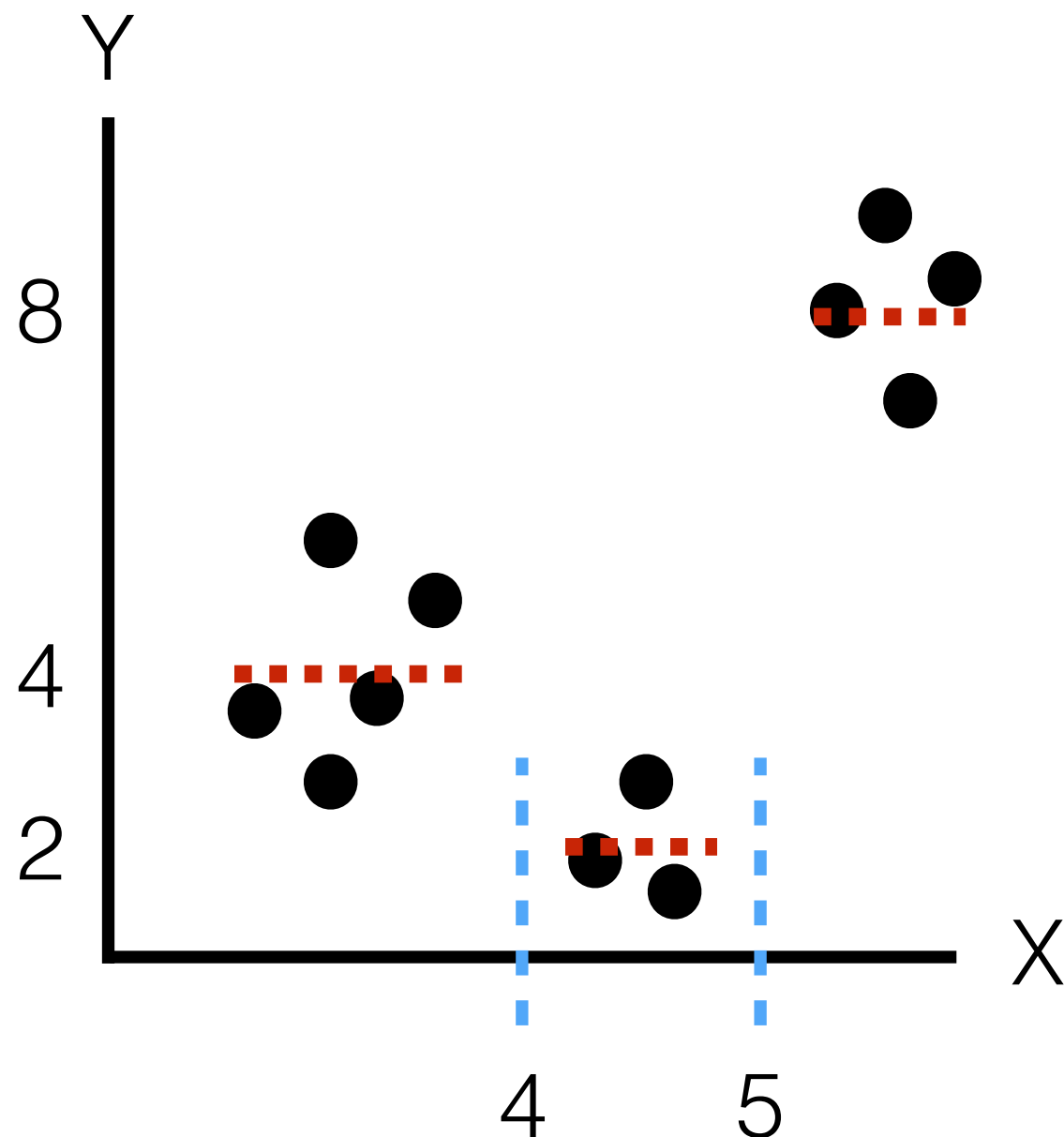
Binary Classification Tree

* Minimize the error



Binary Regression Tree

* Minimize the error



<https://classroom.github.com/a/tw-yDKgD>