Course Announcements

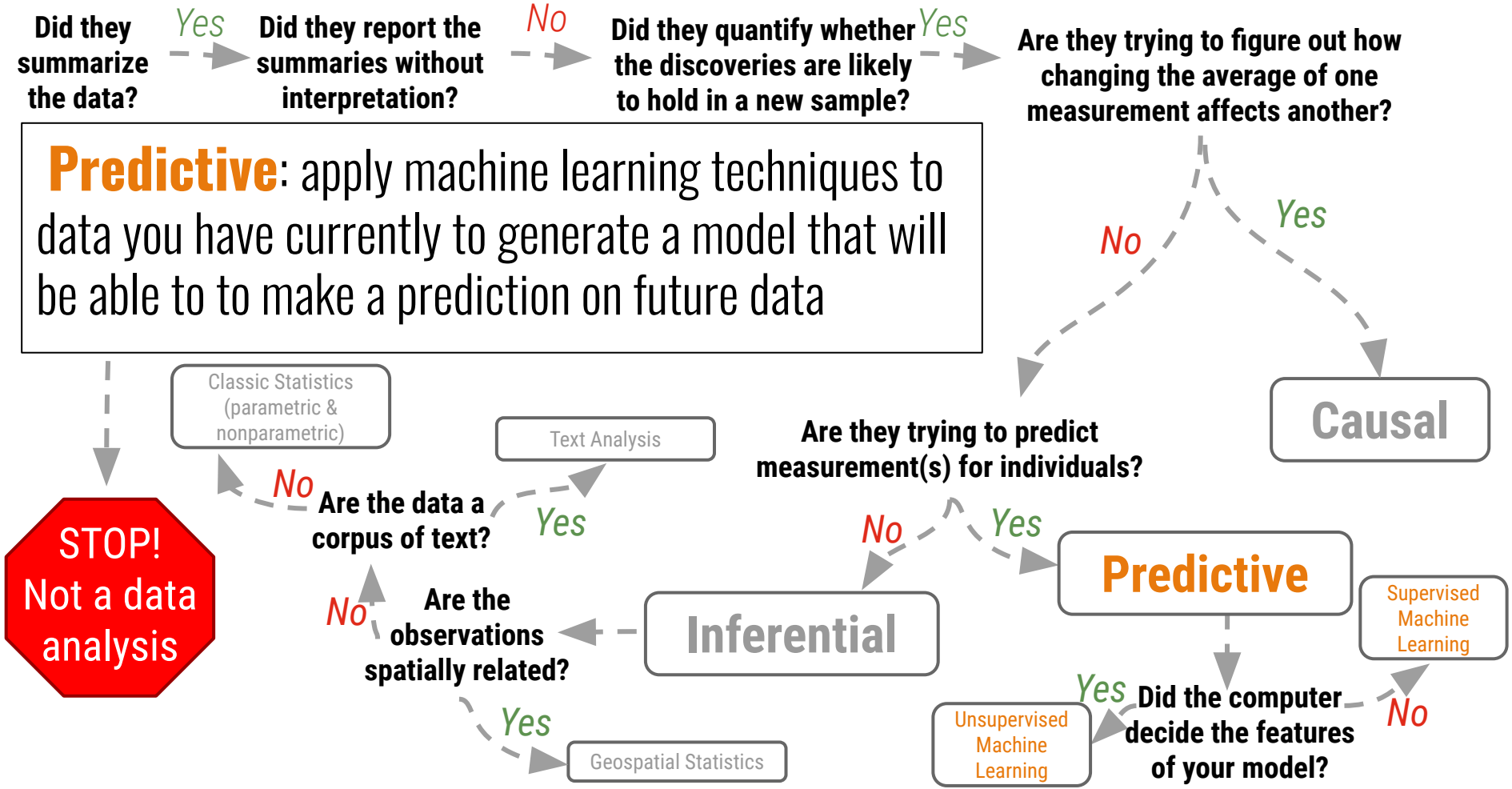
Due today (Friday; 11:59 PM):

- D6
- Q6
- A3
- Weekly Project Survey

Machine Learning

Shannon E. Ellis, Ph.D
UC San Diego

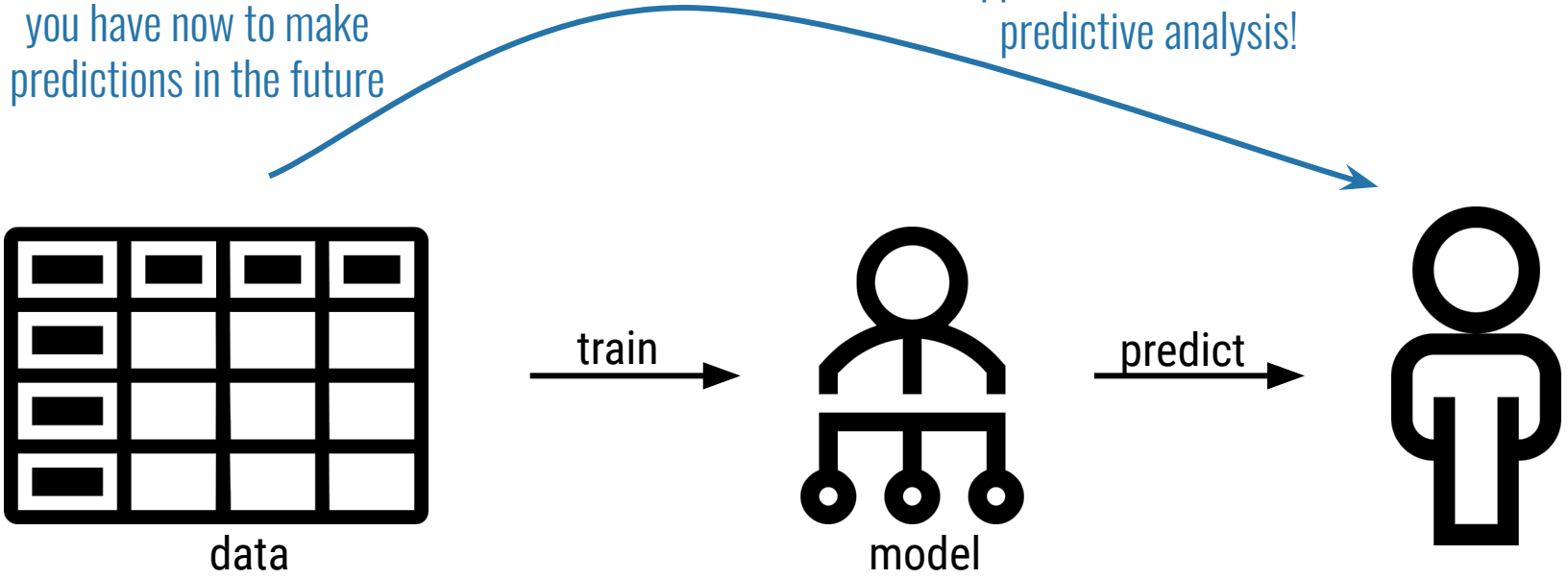
Department of Cognitive Science
sellis@ucsd.edu


Did they summarize the data?
Yes
Did they report the summaries without interpretation?
No
Did they quantify whether the discoveries are likely to hold in a new sample?
Yes
Are they trying to figure out how changing the average of one measurement affects another?
Predictive: apply machine learning techniques to data you have currently to generate a model that will be able to to make a prediction on future data
No
Yes
Causal
Are they trying to predict measurement(s) for individuals?
No
Yes
Predictive
Supervised Machine Learning
Did the computer decide the features of your model?
Yes
Unsupervised Machine Learning
No
Inferential
Are the observations spatially related?
No
Yes
Geospatial Statistics
Are the data a corpus of text?
No
Yes
Text Analysis
Classic Statistics (parametric & nonparametric)
STOP! Not a data analysis

- **Problem:** Detecting whether credit card charges are fraudulent.
- **Data science question:** Can we use the time of the charge, the location of the charge, and the price of the charge to predict whether that charge is fraudulent or not?
- **Type of analysis:** Predictive analysis

predictive analysis uses data you have now to make predictions in the future

machine learning approaches are used for predictive analysis!



What is machine learning?

"Machine learning is the science of getting computers to act without being explicitly programmed"

- Andrew Ng, Stanford, ex-Google, chief scientist at Baidu, Coursera founder, Stanford Adjunct Faculty

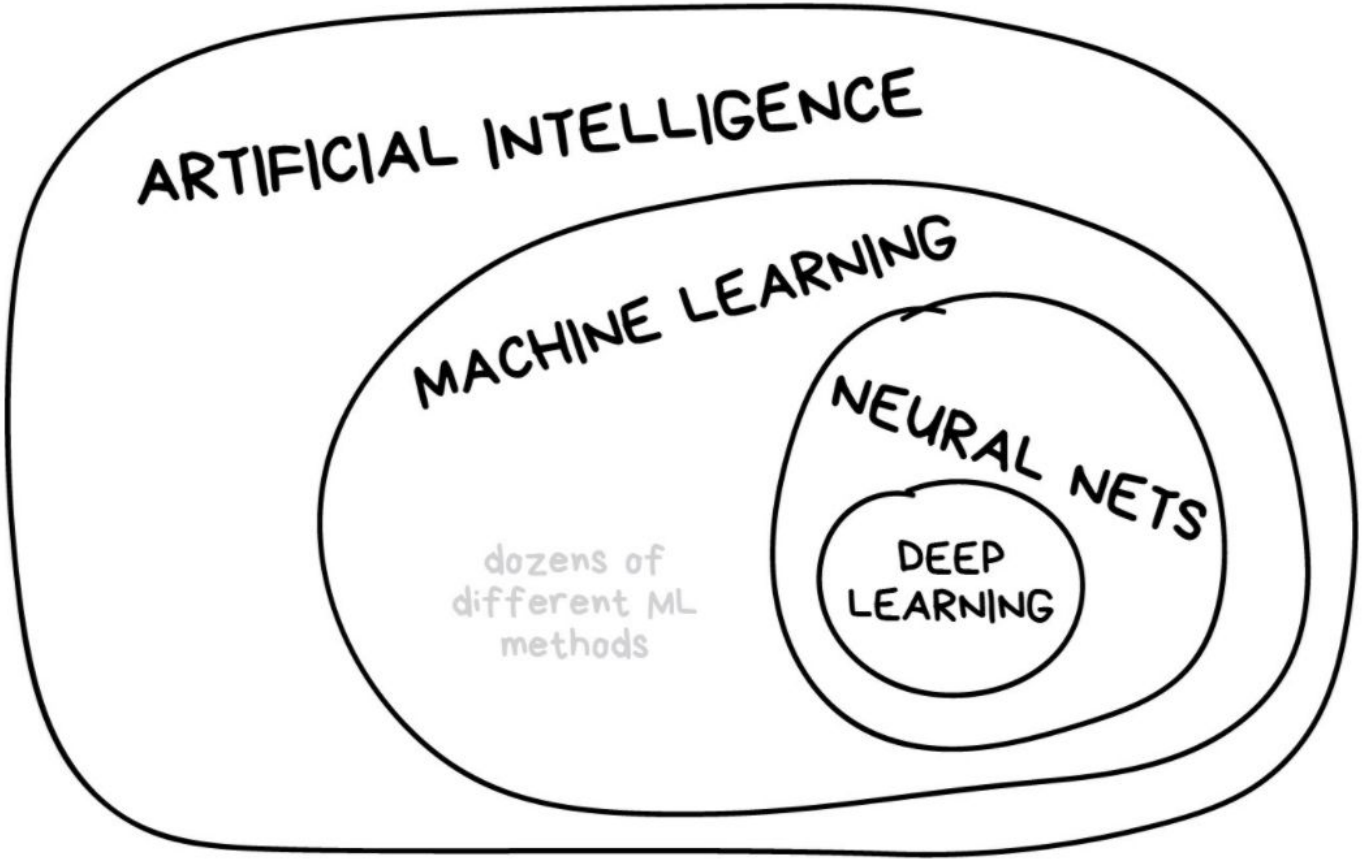


Image source: https://vas3k.com/blog/machine_learning/

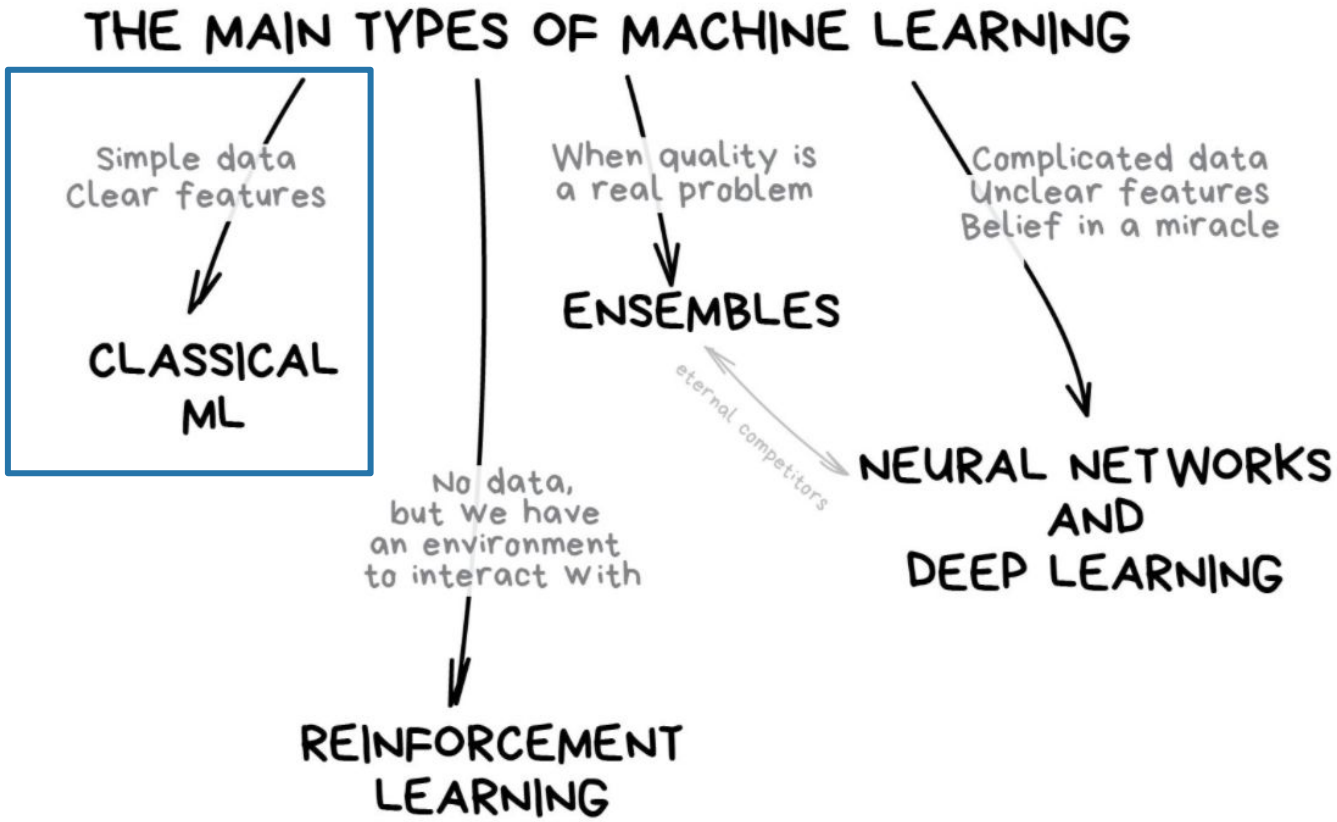


Image source: https://vas3k.com/blog/machine_learning/

Prediction Questions

Which of these questions is most appropriate for machine learning?

A How common is watching Sesame Street in the US?

B What is the effect of watching Sesame Street on children's brains?

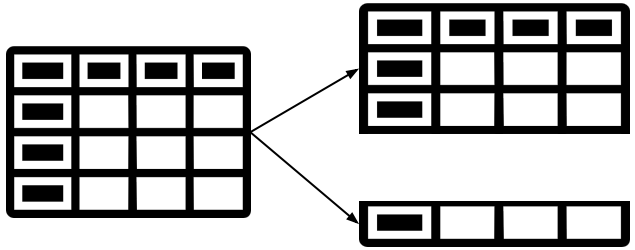
C What is the relationship between early childhood educational programming and success in elementary school?

D Can we use information about one's early childhood to predict their success in elementary school?

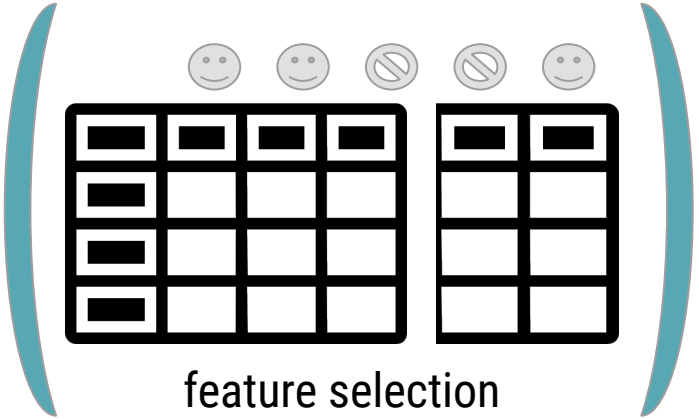
E How does Sesame Street cause an increase in educational attainment?

Machine Learning Generalizations

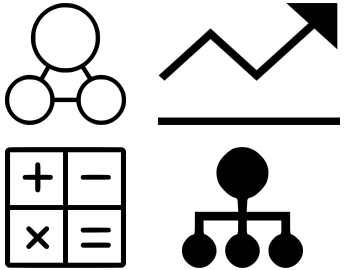
Basic Steps to Prediction



data
partitioning

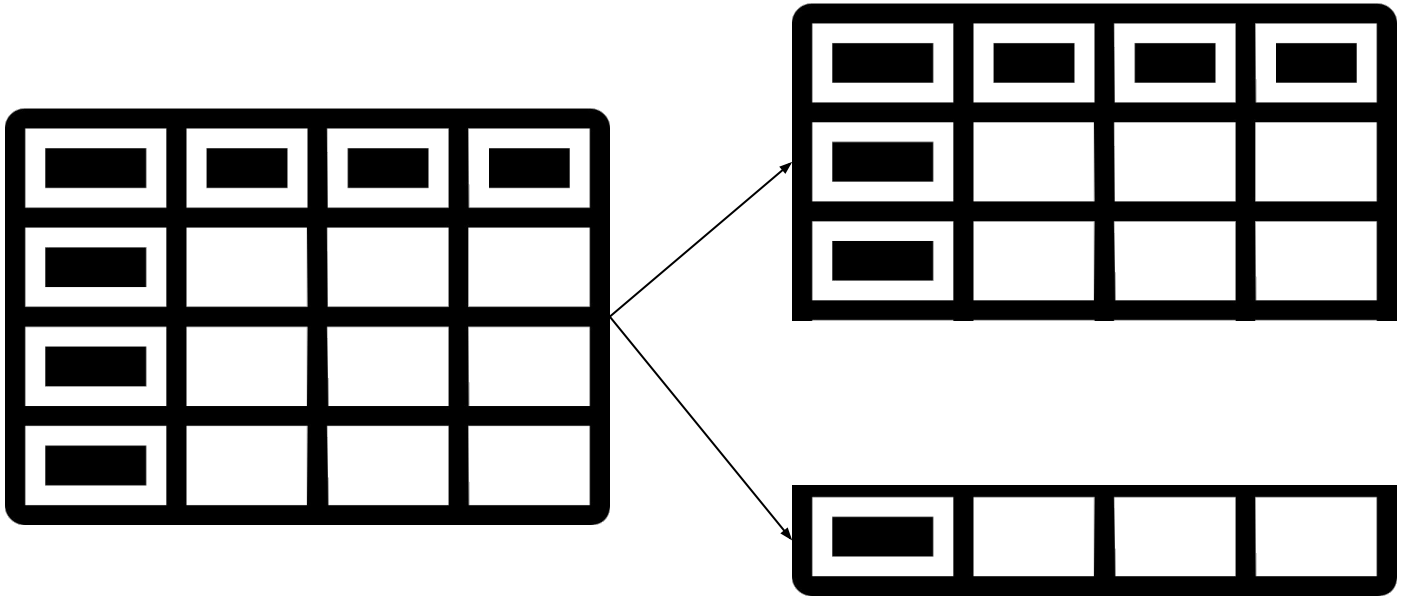


feature selection

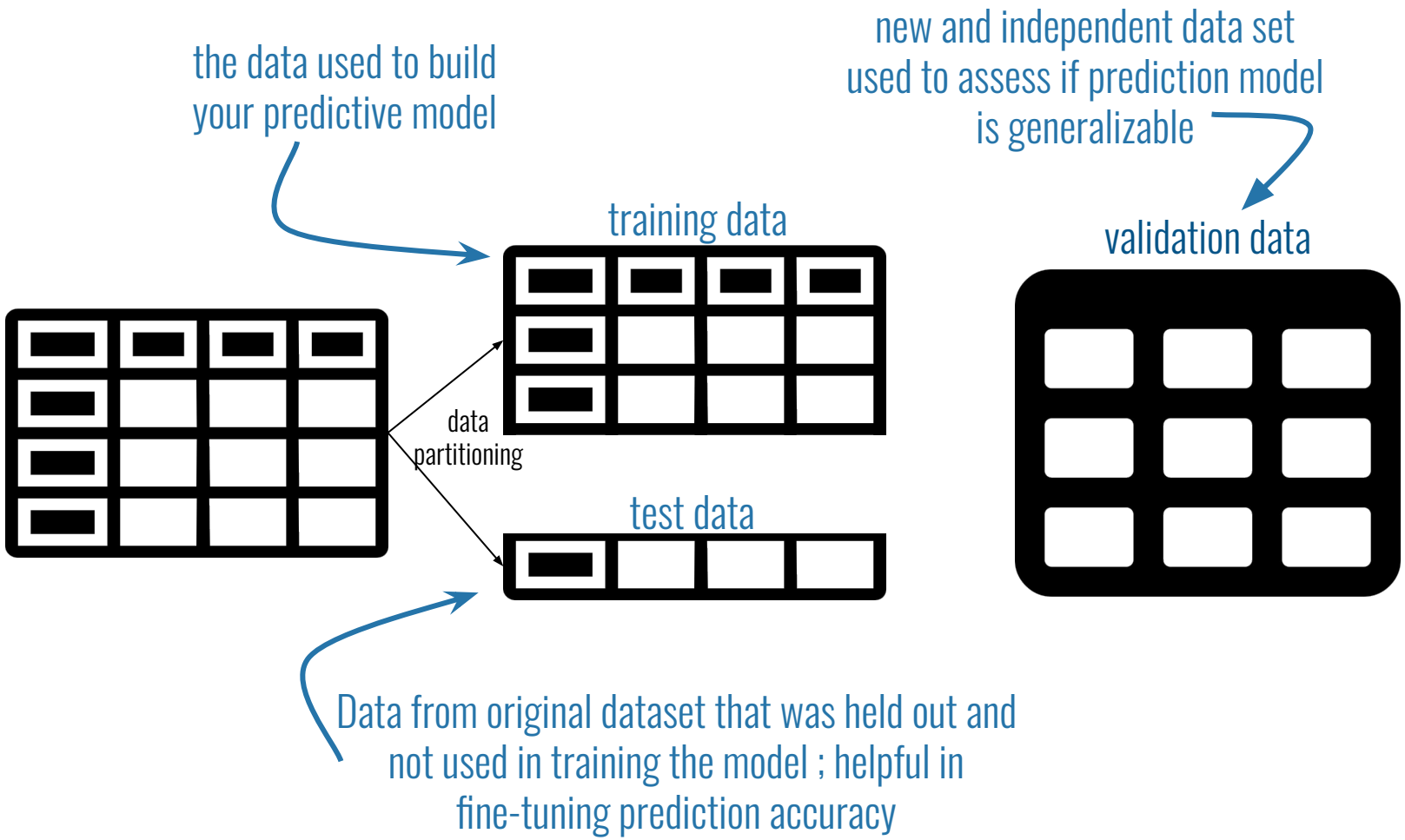


model selection

model assessment

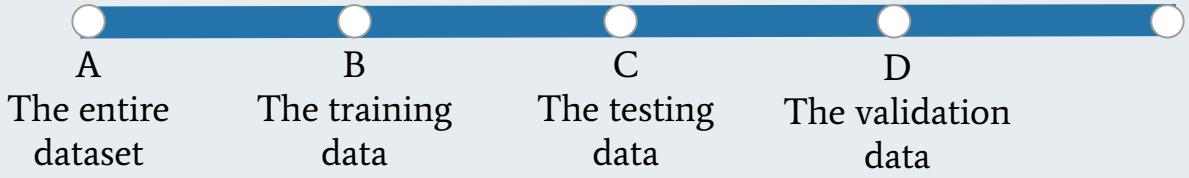


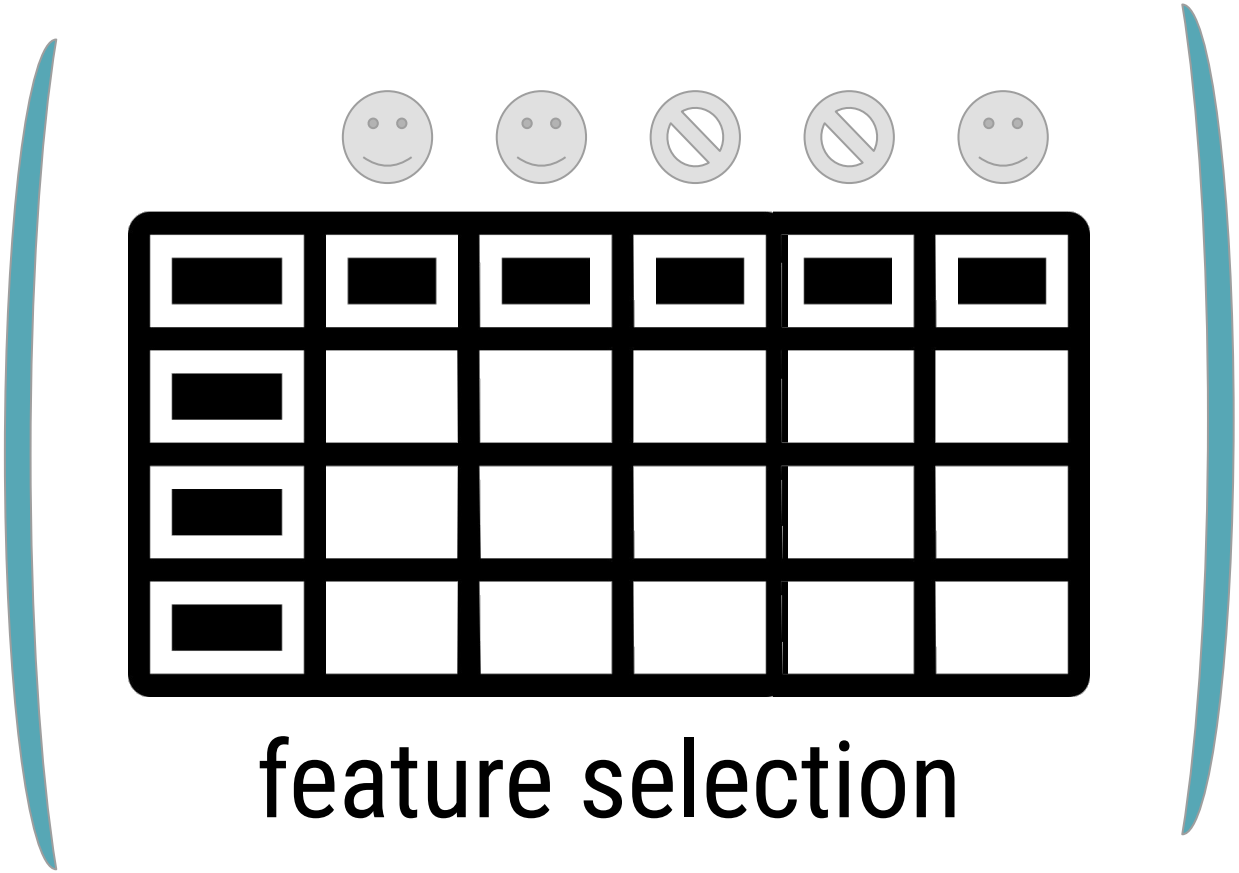
data partitioning


the data used to build your predictive model
new and independent data set used to assess if prediction model is generalizable
training data
validation data
data partitioning
test data
Data from original dataset that was held out and not used in training the model ; helpful in fine-tuning prediction accuracy

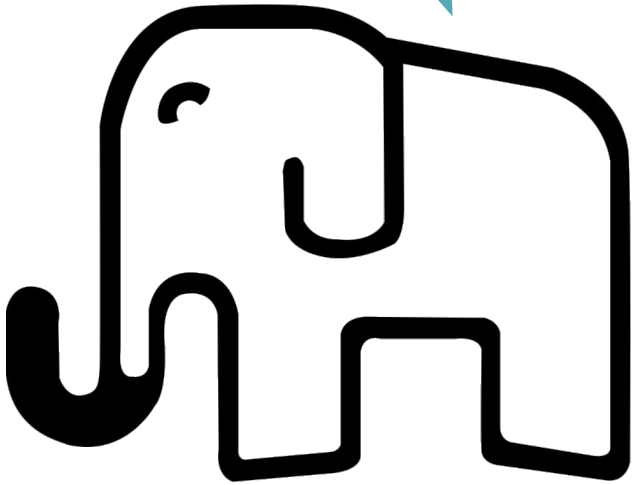
Data Partitioning

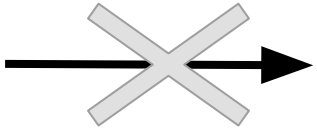
What portion of the data are typically used for generating the model?

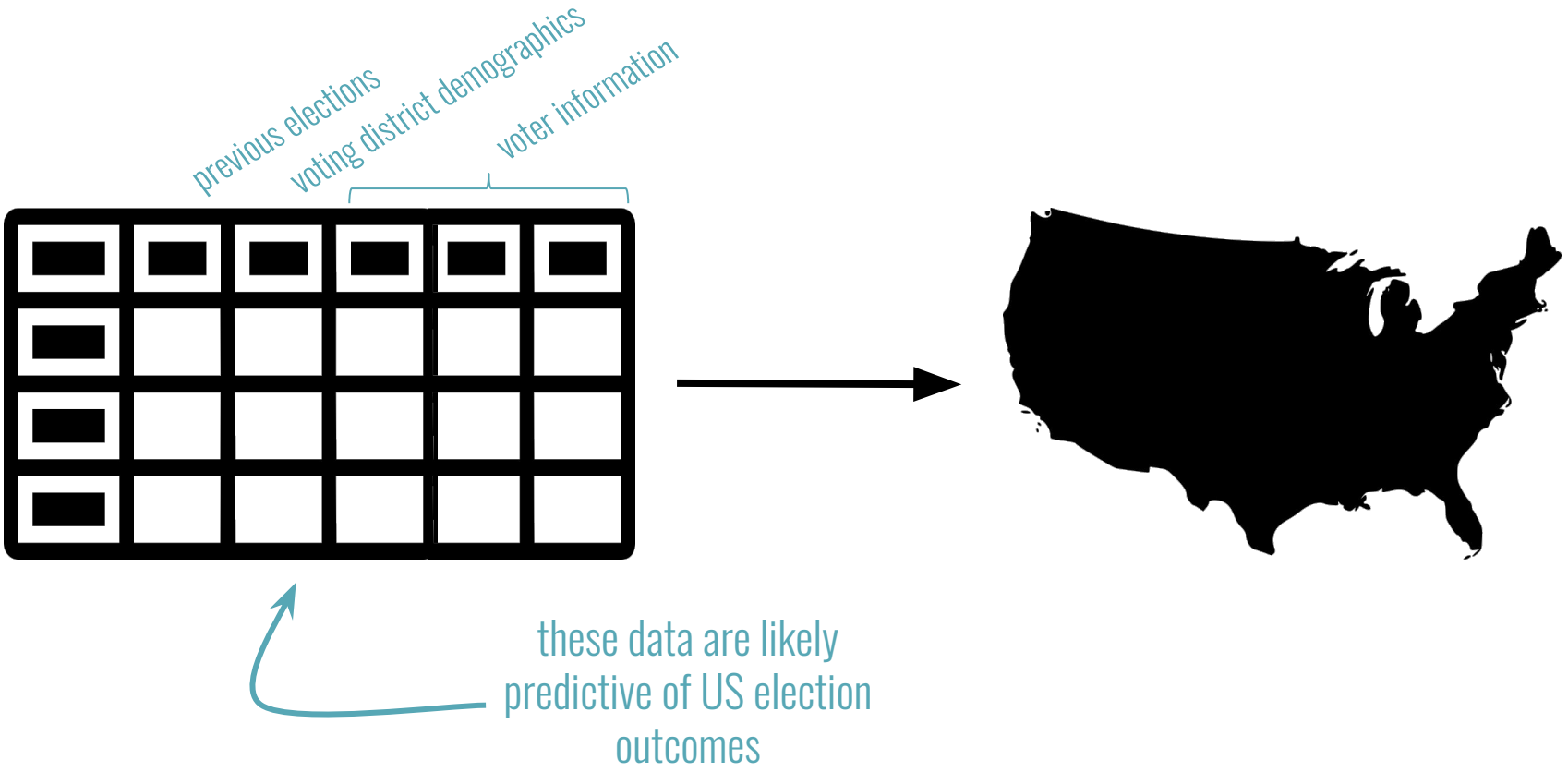


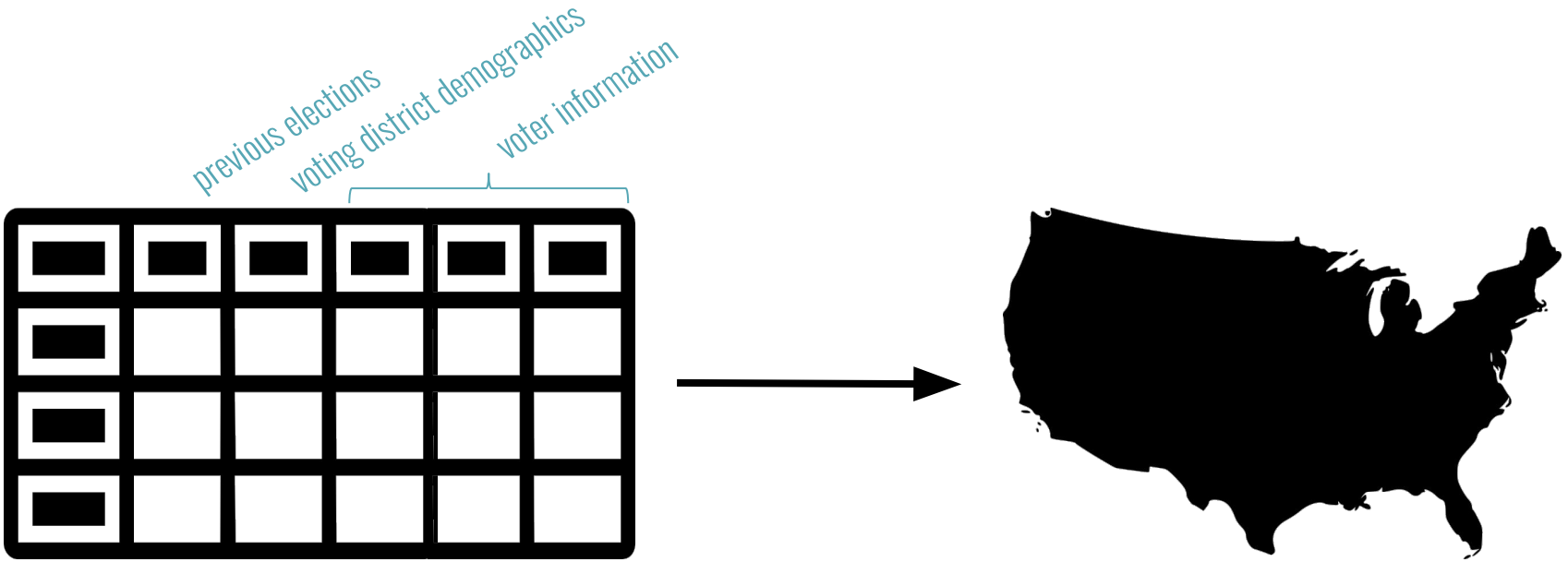

feature selection

elephant height data are likely
not predictive of US elections

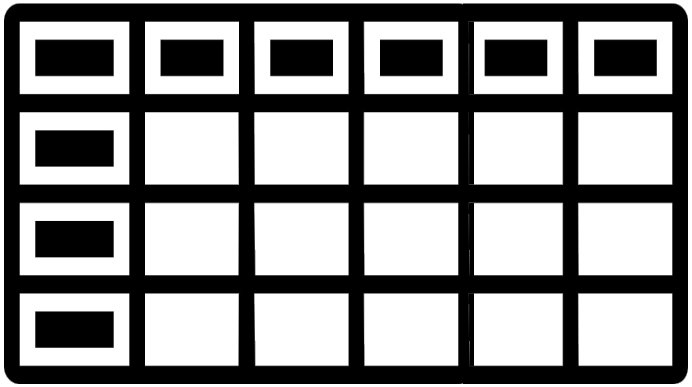





previous elections
voting district demographics
voter information
these data are likely predictive of US election outcomes



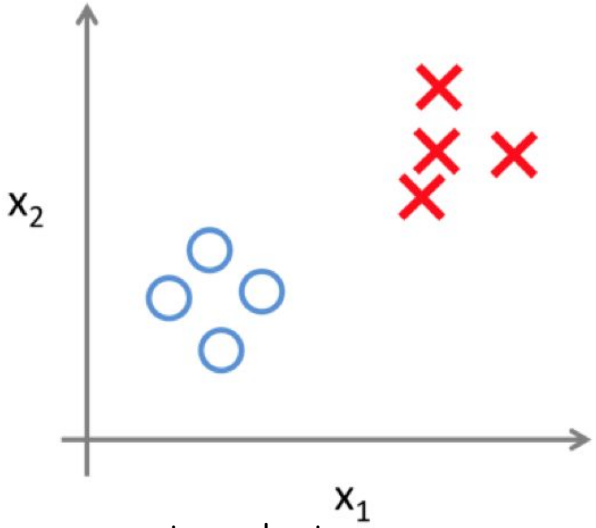
feature selection determines which variables are most predictive and includes them in the model



variables that can be used for accurate prediction exploit the relationship between the variables but do NOT mean that one causes the other

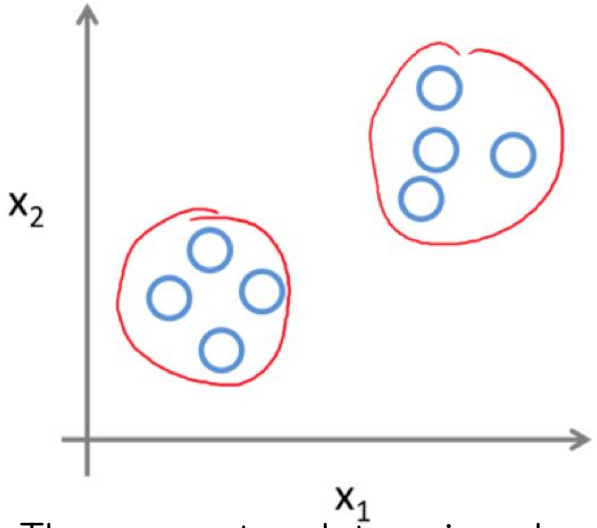
To modes of machine learning

Supervised Learning



You tell the computer what features to use to classify the observations

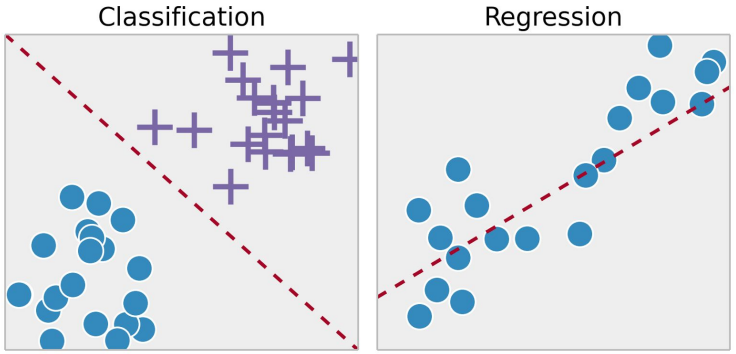
Unsupervised Learning



The computer determines how to classify based on properties within the data

Approaches to machine learning

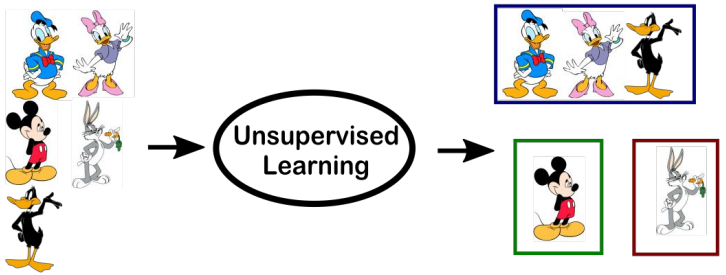
Supervised Learning



categorical variables

continuous variables

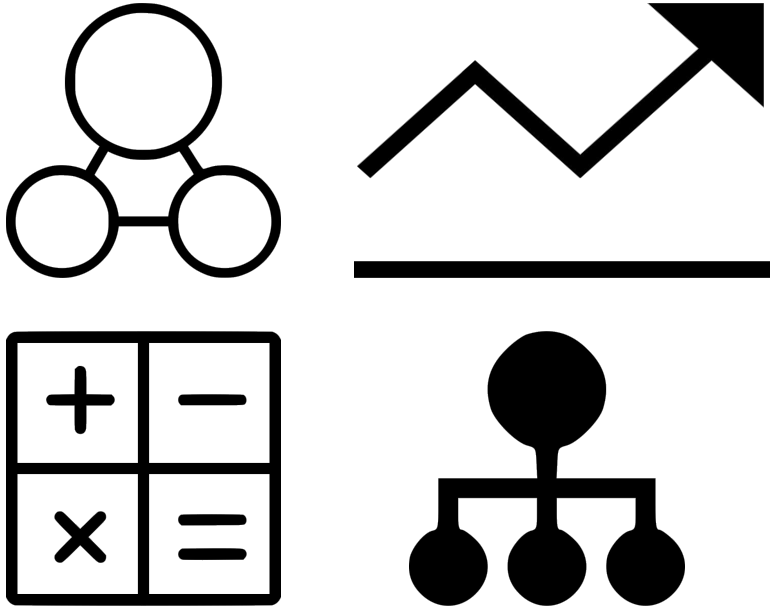
Unsupervised Learning



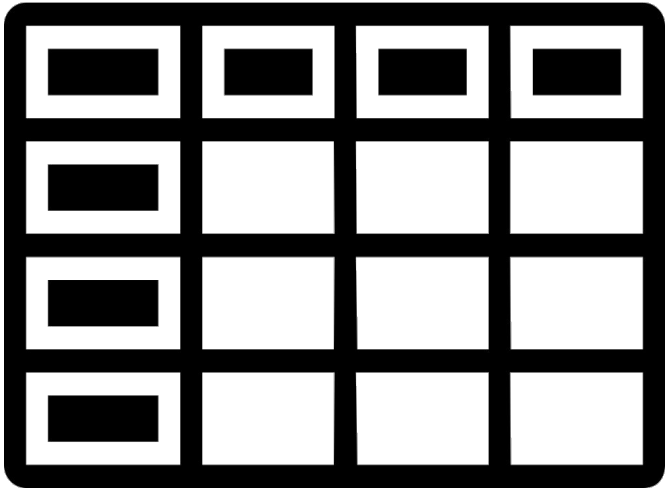
Clustering (categorical)
& dimensionality reduction (continuous)

can automatically identify structure in data

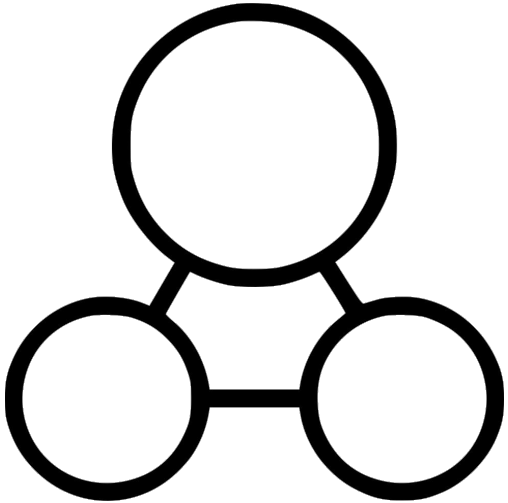
Image source: https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d



model selection



big
datasets



simple
models

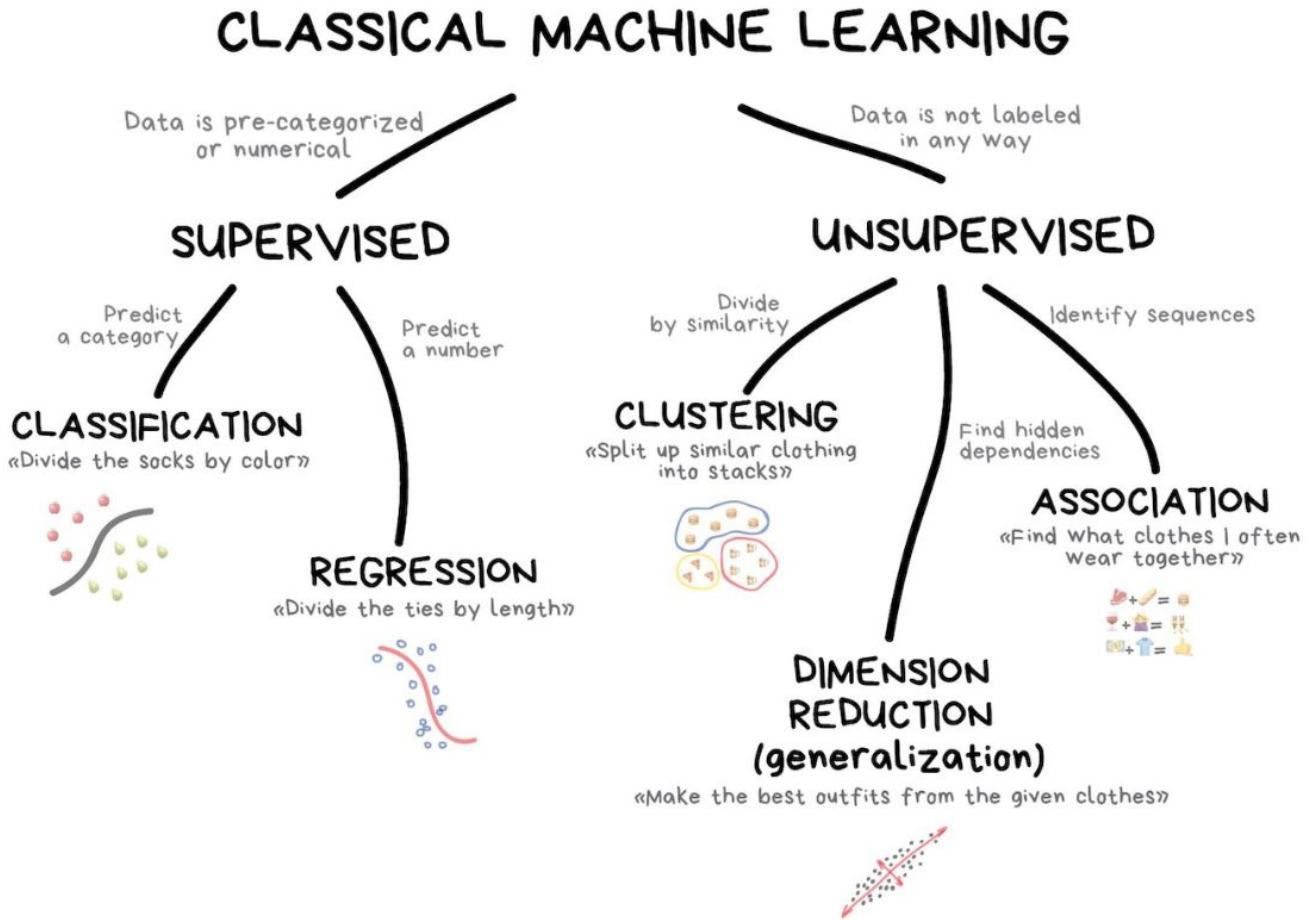


Image source: https://vas3k.com/blog/machine_learning/

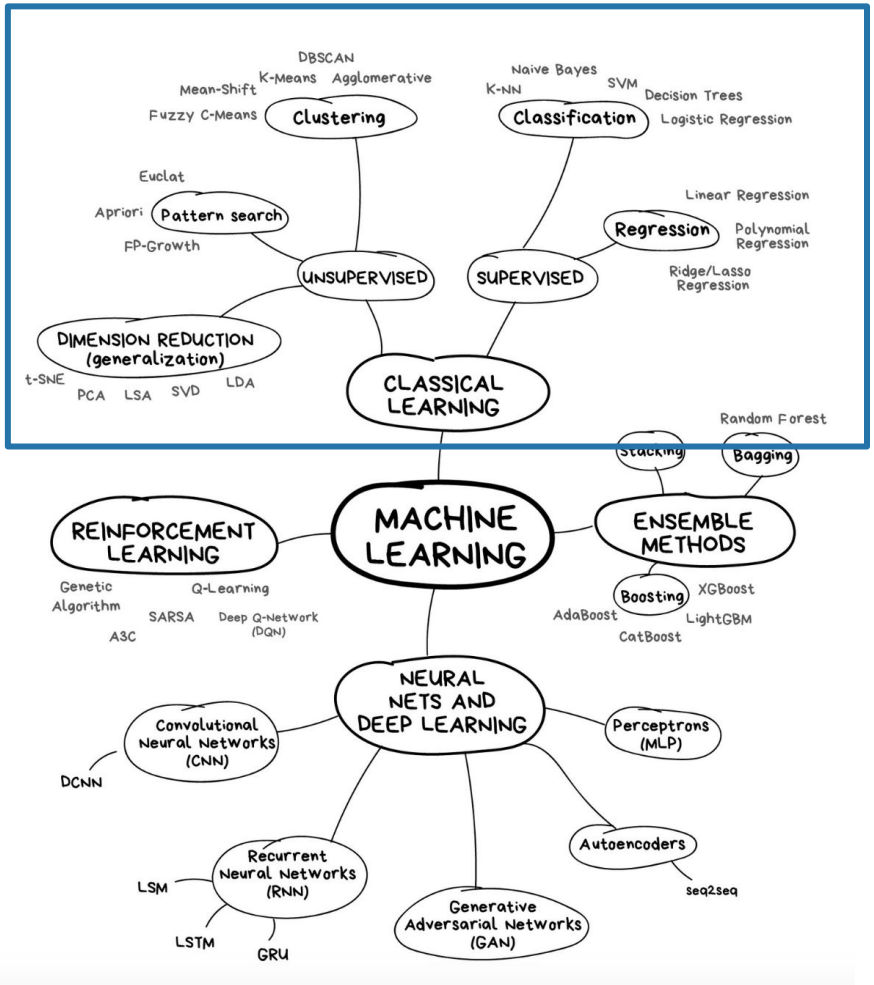
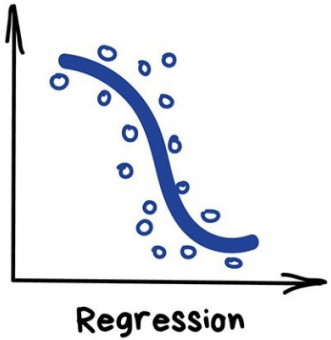


Image source: https://vas3k.com/blog/machine_learning/

Regression

"Draw a line through these dots. Yep, that's the machine learning"



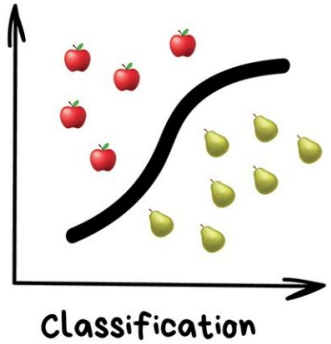
Today this is used for:

- Stock price forecasts
- Demand and sales volume analysis
- Medical diagnosis
- Any number-time correlations

Popular algorithms are Linear and Polynomial regressions.

Classification

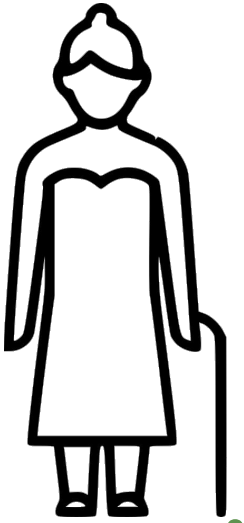
"Splits objects based at one of the attributes known beforehand. Separate socks by based on color, documents based on language, music by genre"



Today used for:
– Spam filtering
– Language detection
– A search of similar documents
– Sentiment analysis
– Recognition of handwritten characters and numbers
– Fraud detection

Popular algorithms: Naive Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbours, Support Vector Machine

Image source: https://vas3k.com/blog/machine_learning/



Regression:

predicting continuous variables
(i.e. Age)

continuous variable prediction

Classification:

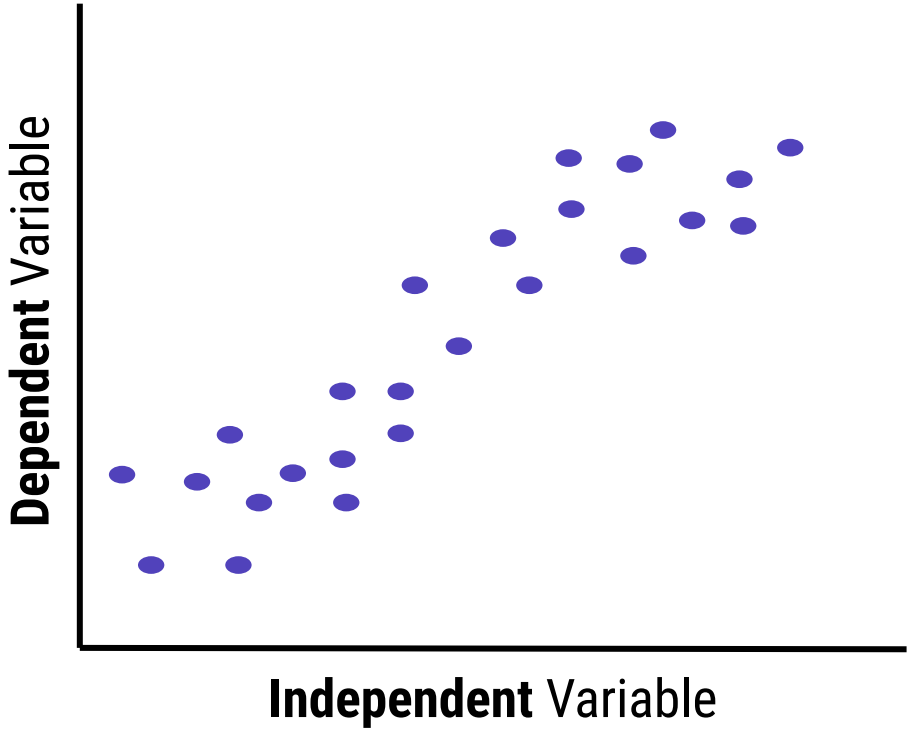
predicting categorical variables
(i.e. education level)

categorical variable prediction

Supervised Learning

regression

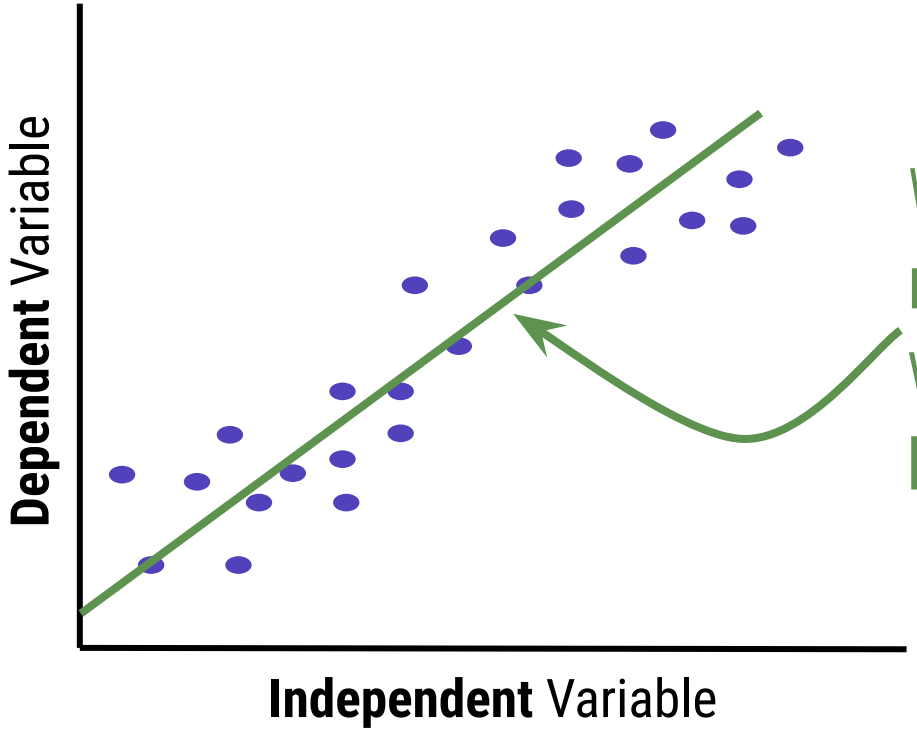
continuous variable prediction



Supervised Learning

regression

continuous variable prediction

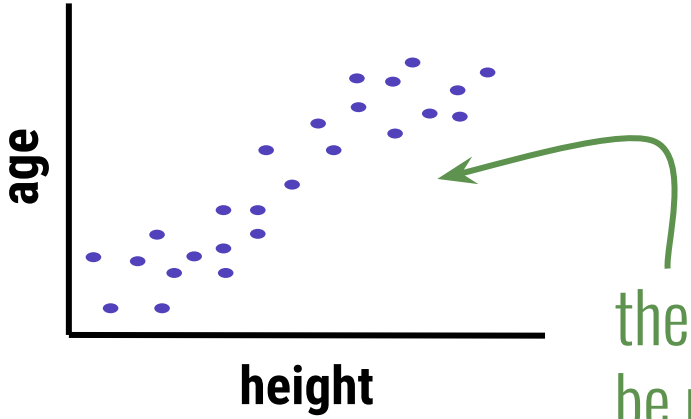


We'll use the linear relationship between variables to generate a **predictive model**

Supervised Learning

regression

continuous variable prediction

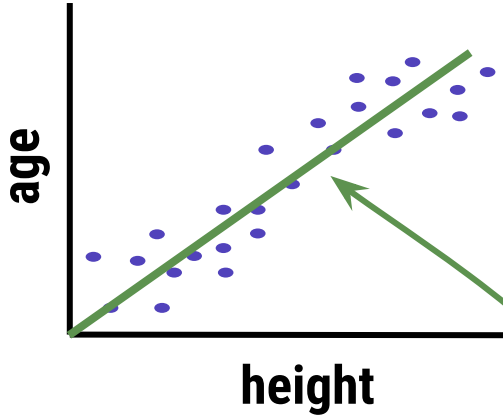


the training data will be used to build the predictive model

Supervised Learning

regression

continuous variable prediction

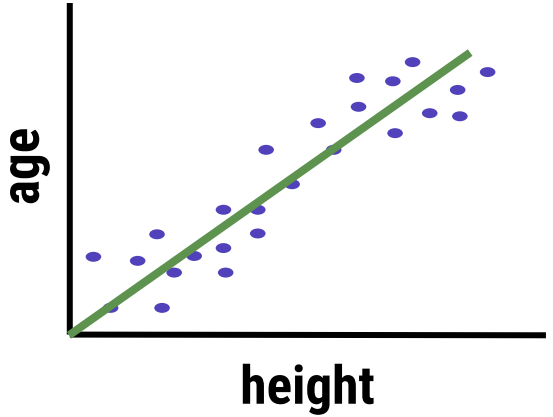


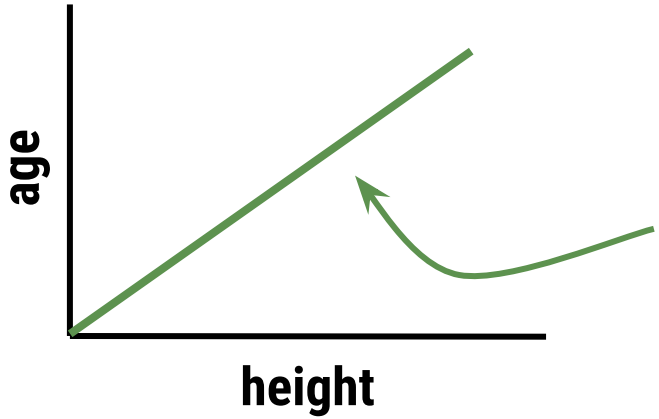
use linear regression to model the relationship

Supervised Learning

regression

continuous variable prediction



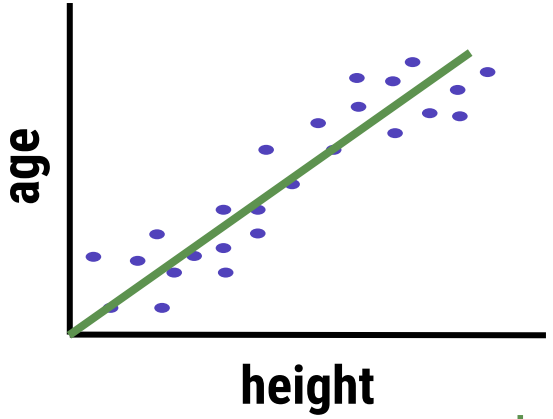


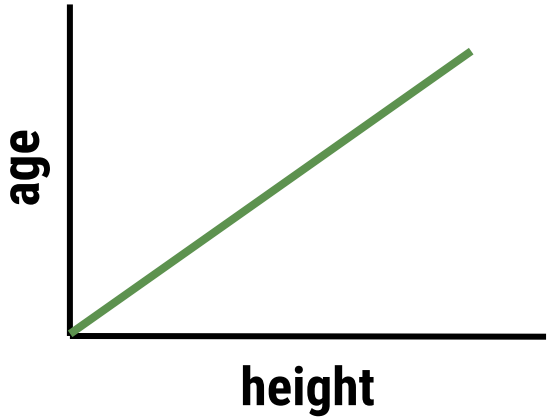
For prediction, the individual values in the training data are *not* important. We only need the model.

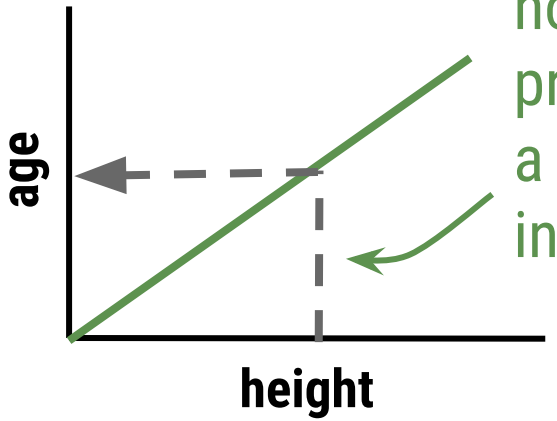
Supervised Learning

regression

continuous variable prediction





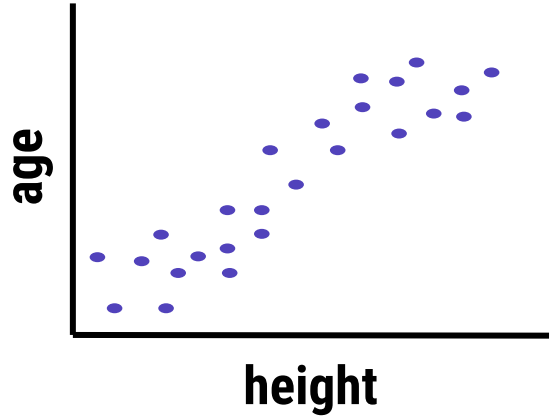


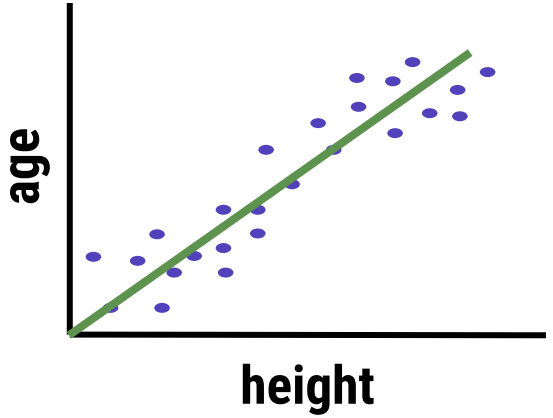
how we'll make predictions for a future individual

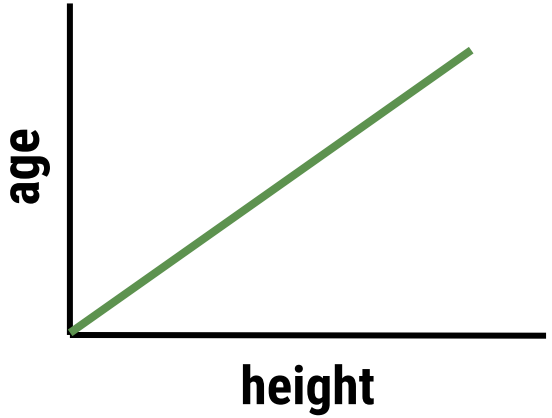
Supervised Learning

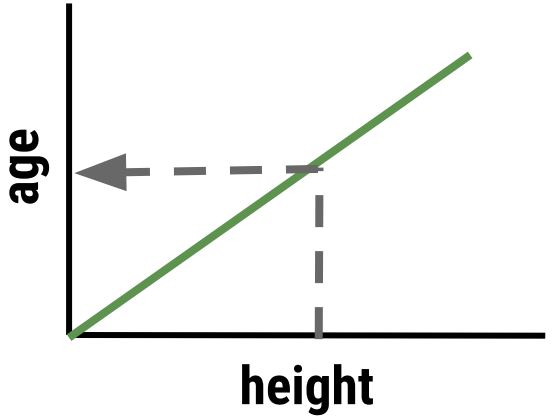
regression

continuous variable prediction

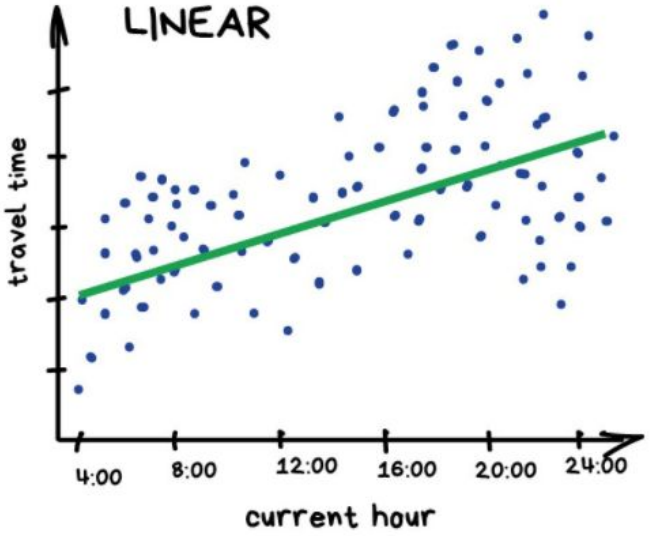


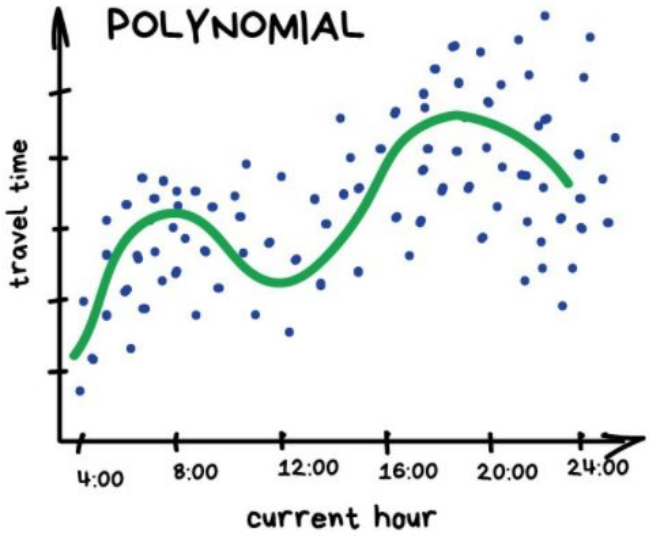






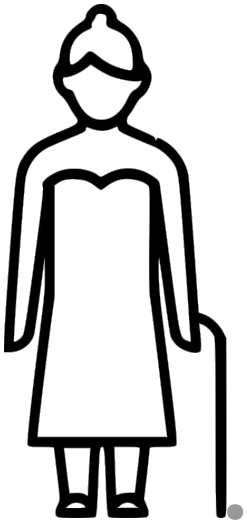
PREDICT TRAFFIC JAMS





REGRESSION

Image source: https://vas3k.com/blog/machine_learning/



Regression:

predicting continuous variables
(i.e. Age)

Classification:

predicting categorical variables
(i.e. education level)

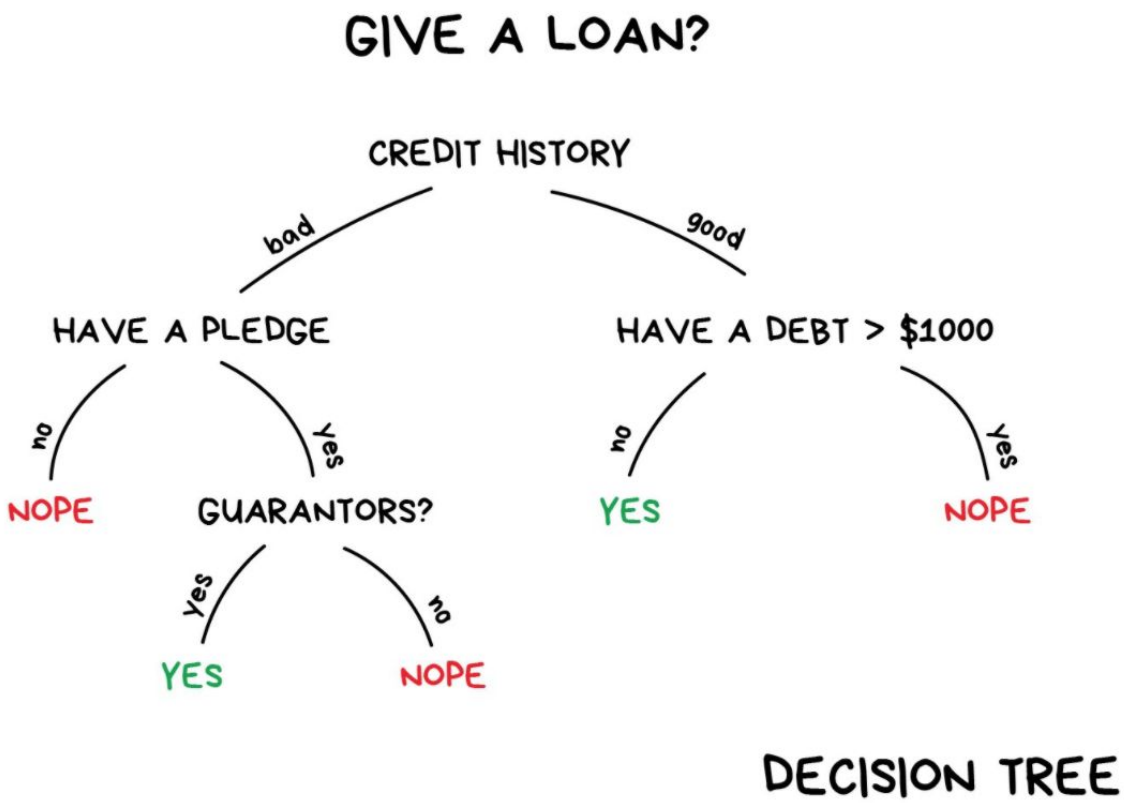
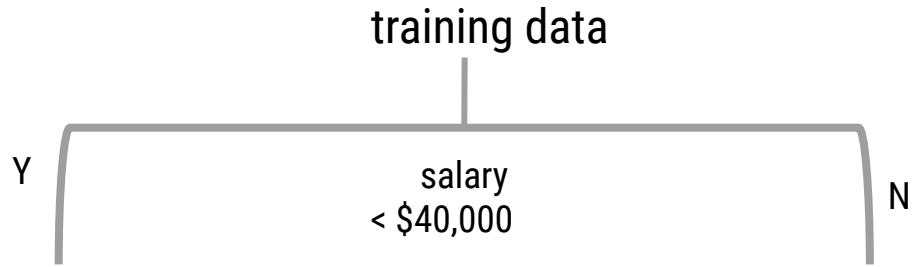


Image source: https://vas3k.com/blog/machine_learning/

Supervised Learning

classification

categorical variable prediction



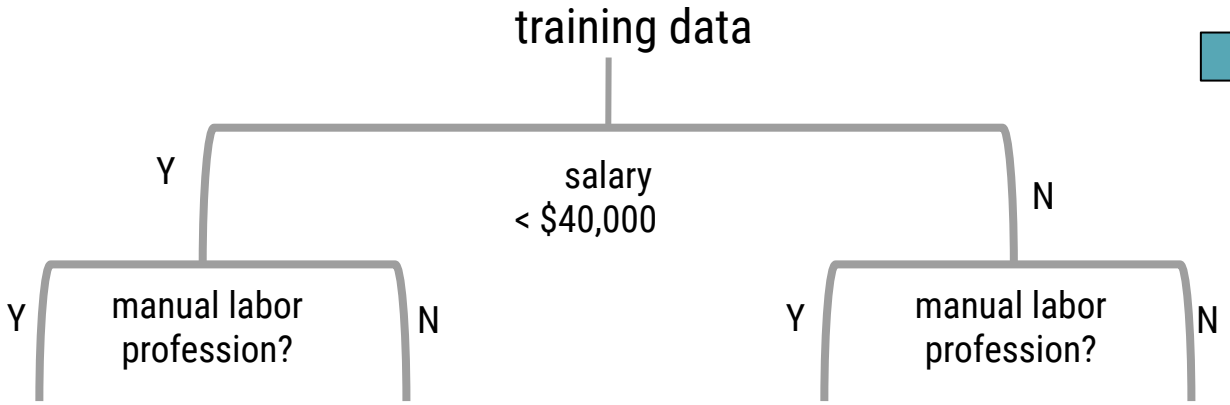
All the people who make *less* than 40K over here

All the people who make *more* than 40K over here

Supervised Learning

classification

categorical variable prediction

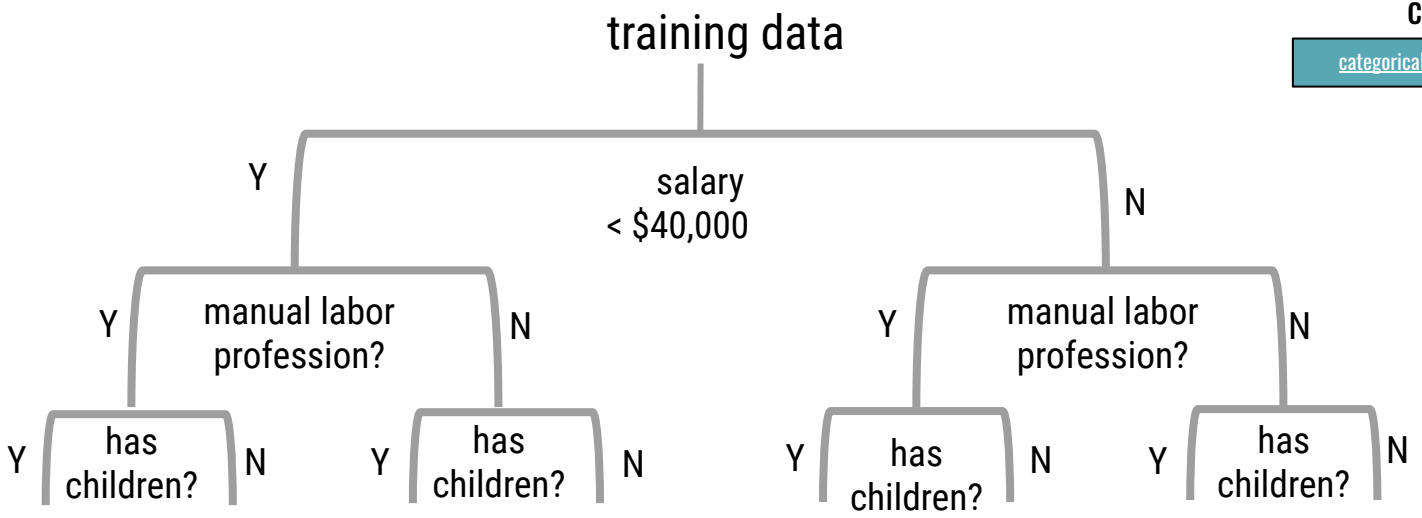


Continue adding *branches* to the **decision tree** where the variables and information in the training data decide which observations goes down which branch

Supervised Learning

classification

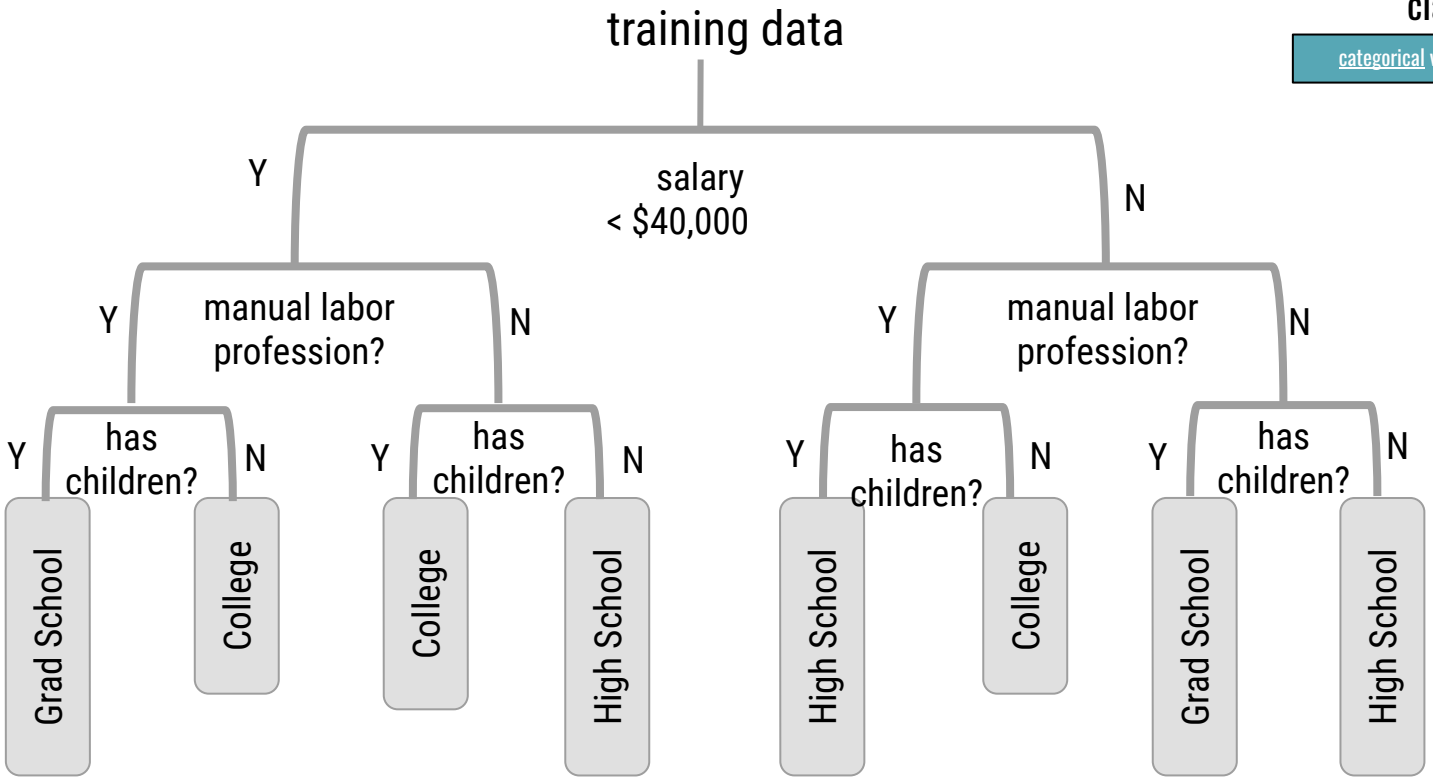
categorical variable prediction



Supervised Learning

classification

categorical variable prediction

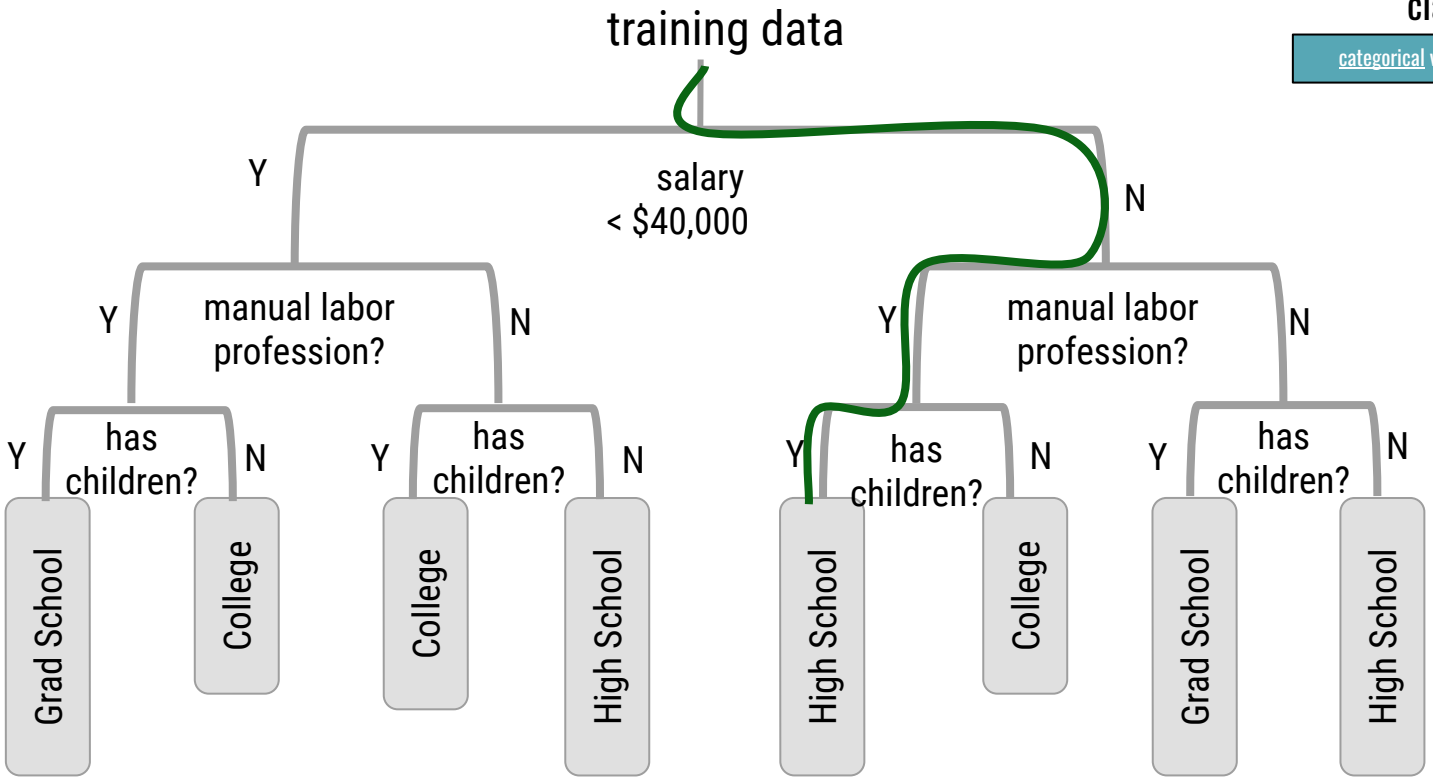


At the end of the tree, labels will be applied to each *leaf* of the tree

Supervised Learning

classification

categorical variable prediction

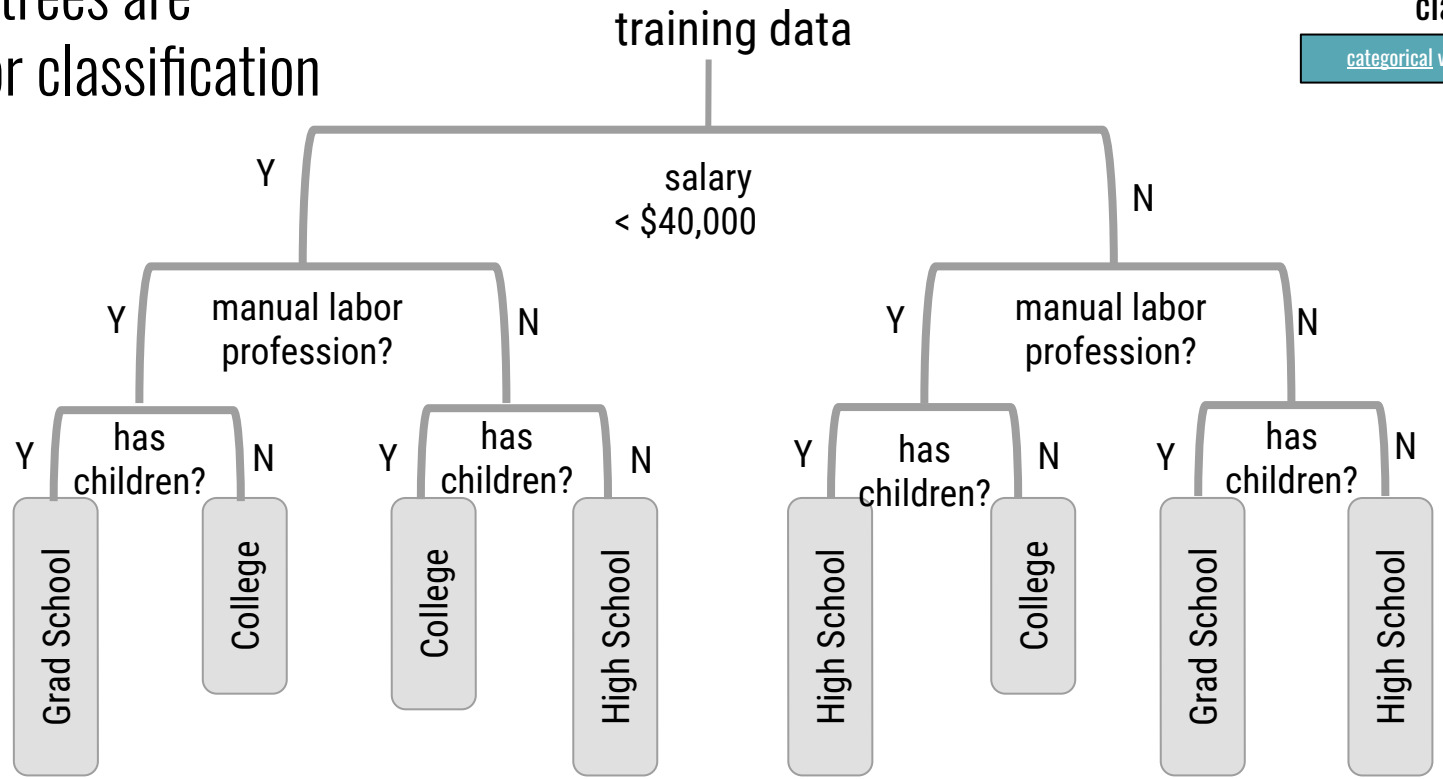


Decision trees are helpful for classification

Supervised Learning

classification

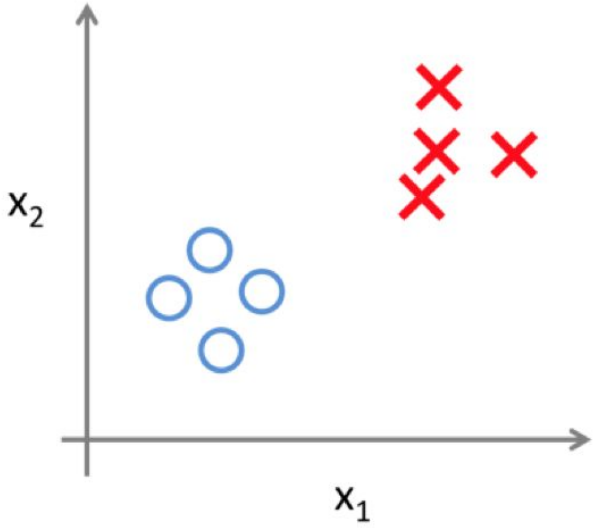
categorical variable prediction



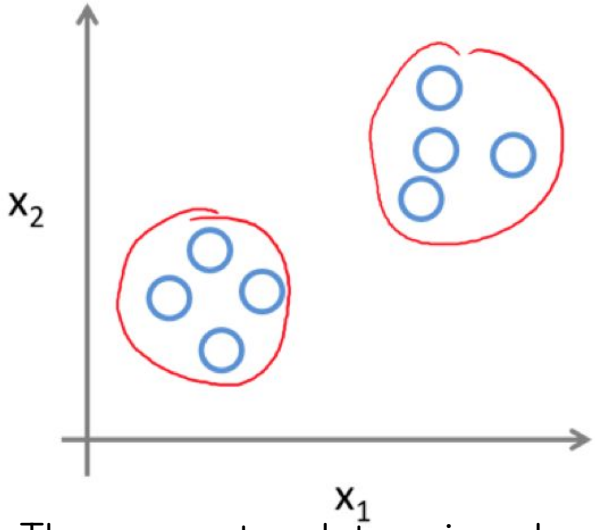
Unsupervised Learning

To modes of machine learning

Supervised Learning



Unsupervised Learning



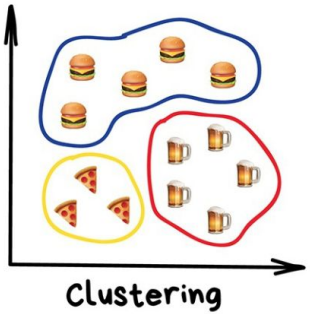
The computer determines how to classify based on properties within the data

Clustering

"Divides objects based on unknown features. Machine chooses the best way"

Nowadays used:

- For market segmentation (types of customers, loyalty)
- To merge close points on a map
- For image compression
- To analyze and label new data
- To detect abnormal behavior



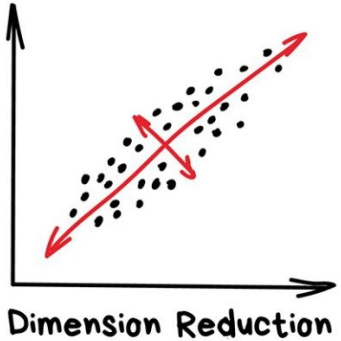
Popular algorithms: K-means_clustering, Mean-Shift, DBSCAN

Dimensionality Reduction (Generalization)

"Assembles specific features into more high-level ones"

Nowadays is used for:

- Recommender systems (★)
- Beautiful visualizations
- Topic modeling and similar document search
- Fake image analysis
- Risk management



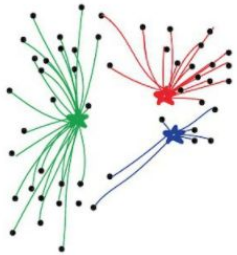
Popular algorithms: Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Latent Dirichlet allocation (LDA), Latent Semantic Analysis (LSA, pLSA, GLSA), t-SNE (for visualization)

Image source: https://vas3k.com/blog/machine_learning/

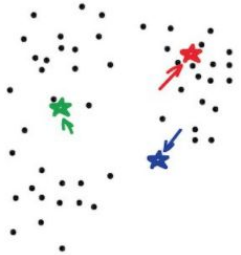
PUT KEBAB KIOSKS IN THE OPTIMAL WAY

(also illustrating the K-means method)

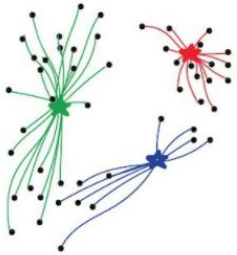
1. Put kebab kiosks in random places in city



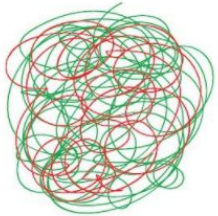
2. Watch how buyers choose the nearest one



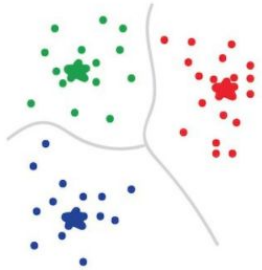
3. Move kiosks closer to the centers of their popularity



4. Watch and move again



5. Repeat a million times



6. Done!
You're god of kebabs!

Image source: https://vas3k.com/blog/machine_learning/

Prediction Approach

You want to predict someone's emotion based on an image.

How would you approach this with machine learning?

A
Supervised, Regression

B
Supervised, Classification

C
Unsupervised, dimensionality reduction

D
Unsupervised, clustering

E
Unsupervised, Neural Network

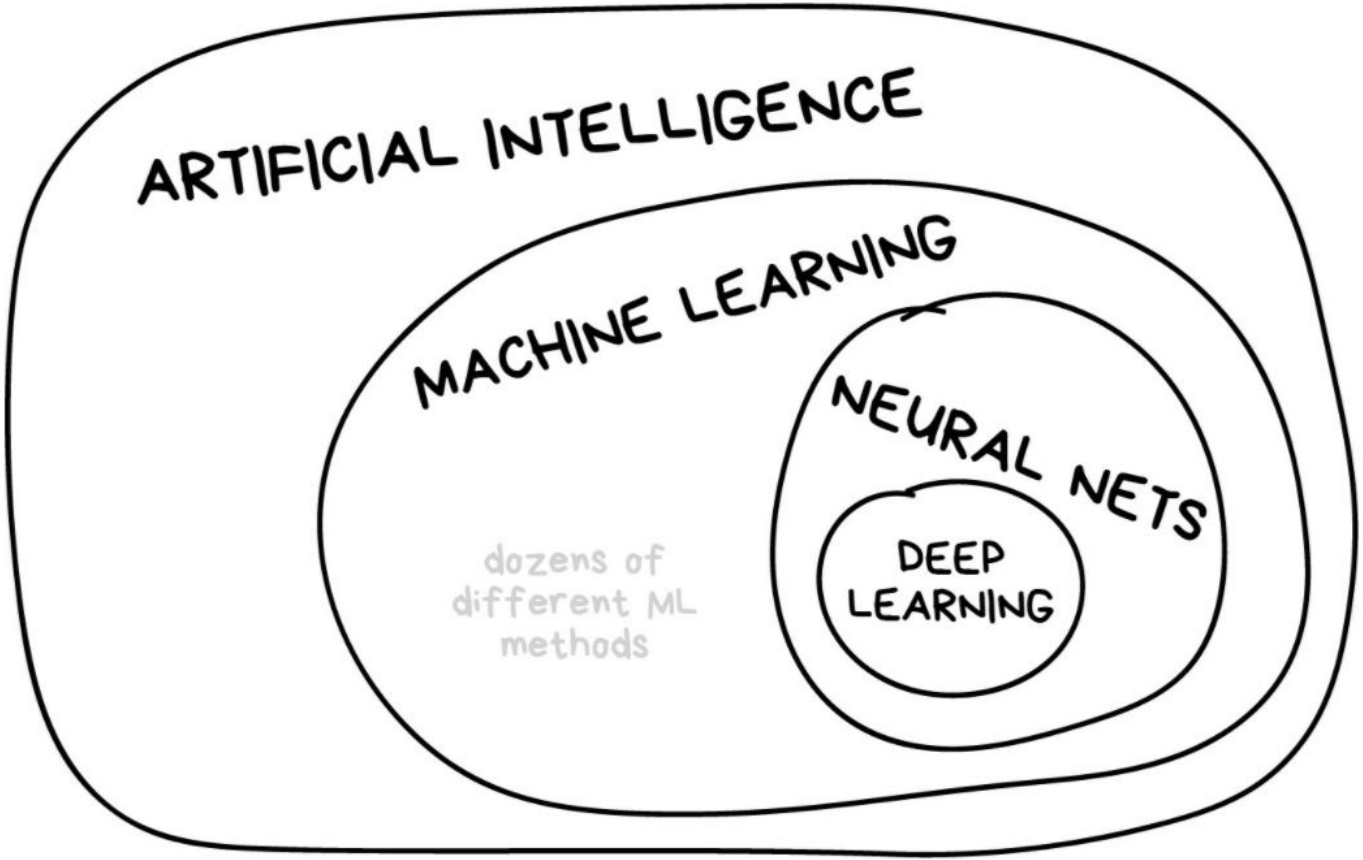


Image source: https://vas3k.com/blog/machine_learning/

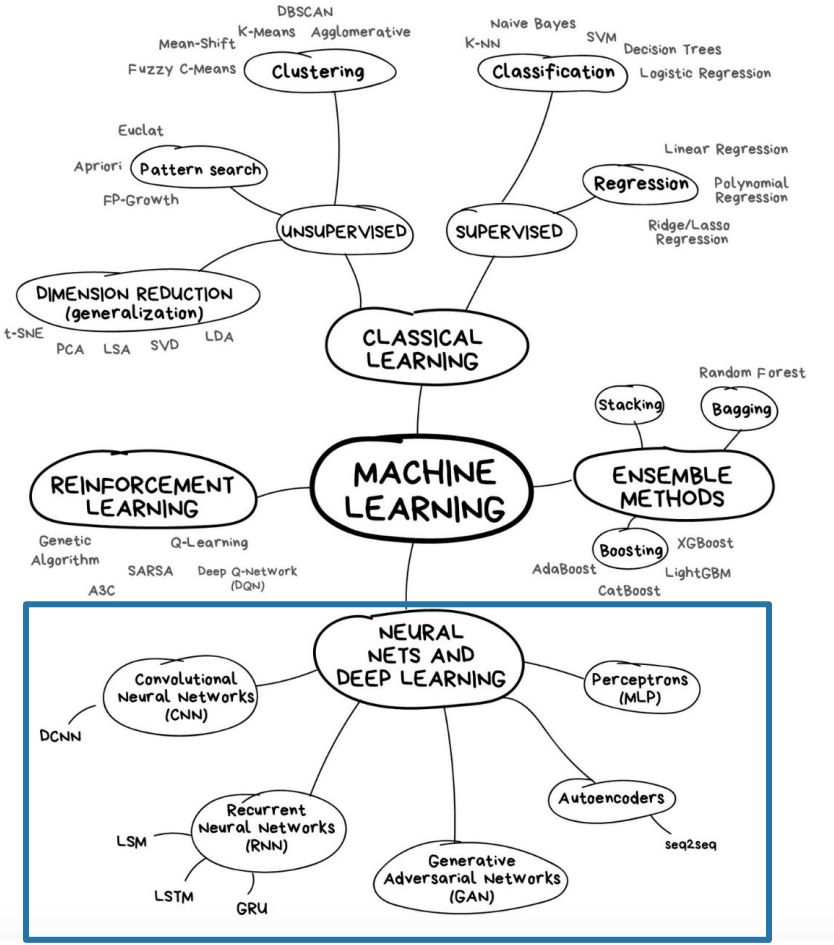
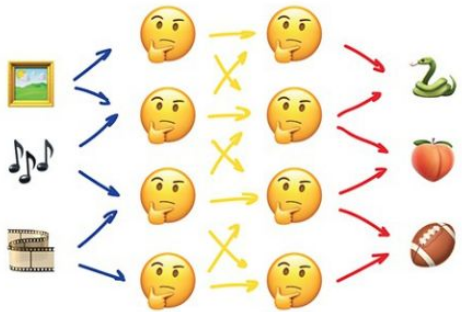


Image source: https://vas3k.com/blog/machine_learning/

"We have a thousand-layer network, dozens of video cards, but still no idea where to use it. Let's generate cat pics!"

Used today for:

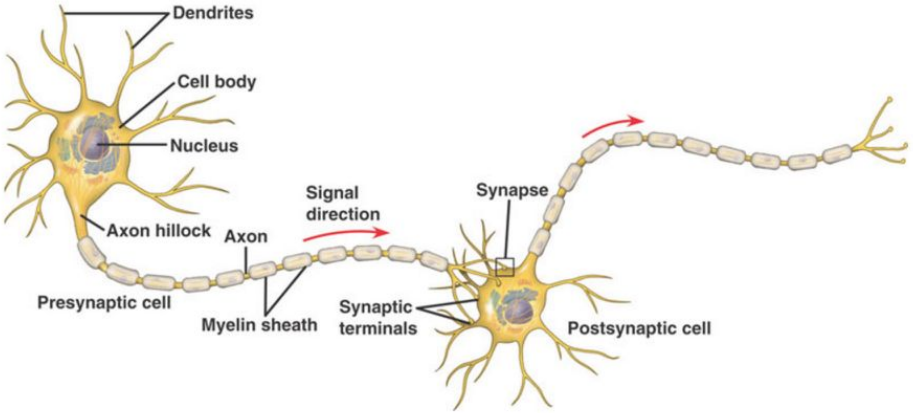
- Replacement of all algorithms above
- Object identification on photos and videos
- Speech recognition and synthesis
- Image processing, style transfer
- Machine translation



Popular architectures: Perceptron, Convolutional Network (CNN), Recurrent Networks (RNN), Autoencoders

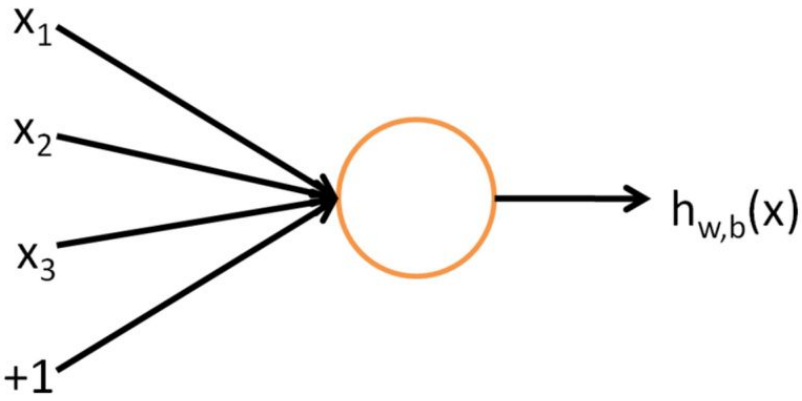
Image source: https://vas3k.com/blog/machine_learning/

WHAT IS A NEURON?



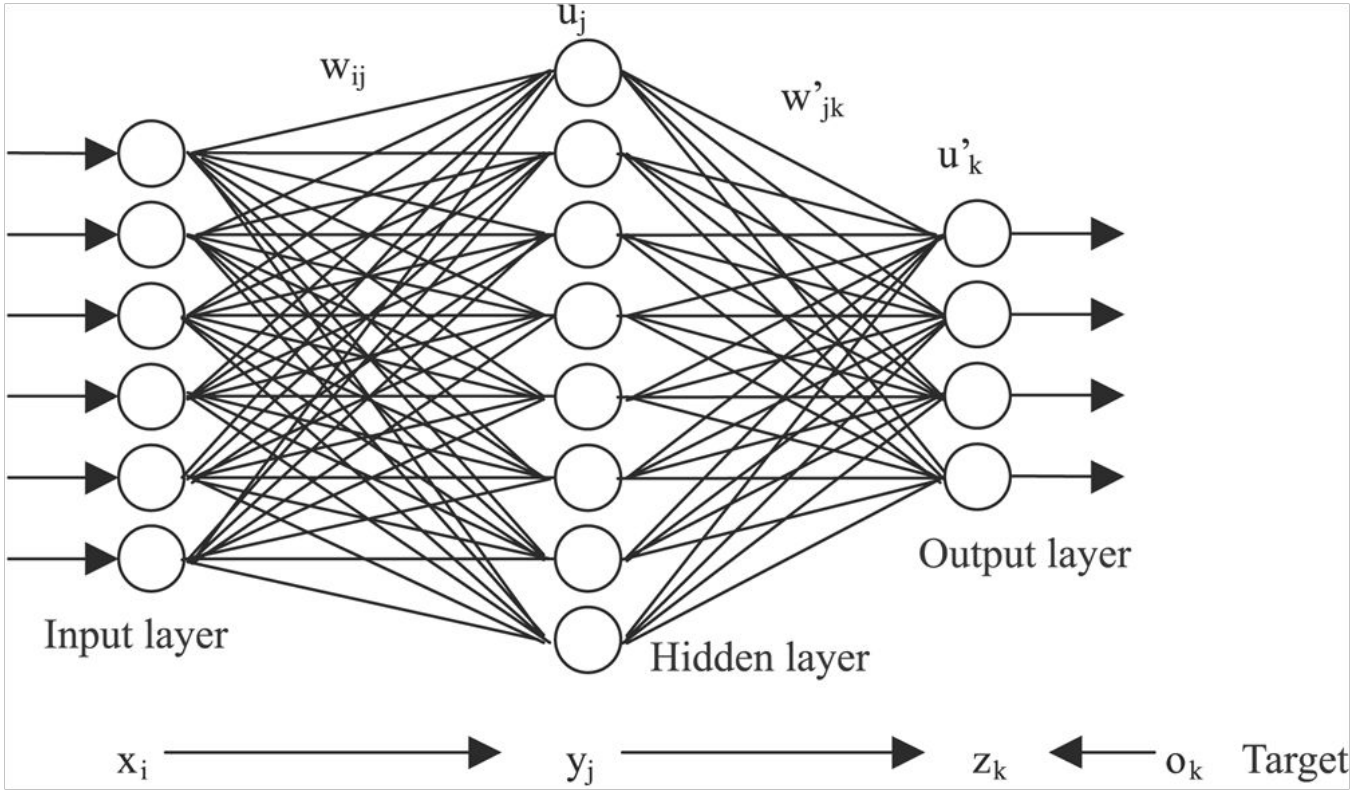
- Receives signal on synapse
- When trigger sends signal on axon

MATHEMATICAL NEURON



- Mathematical abstraction, inspired by biological neuron
- Either on or off based on sum of input

This will likely not be the last time you see this (mostly unhelpful) neural net image



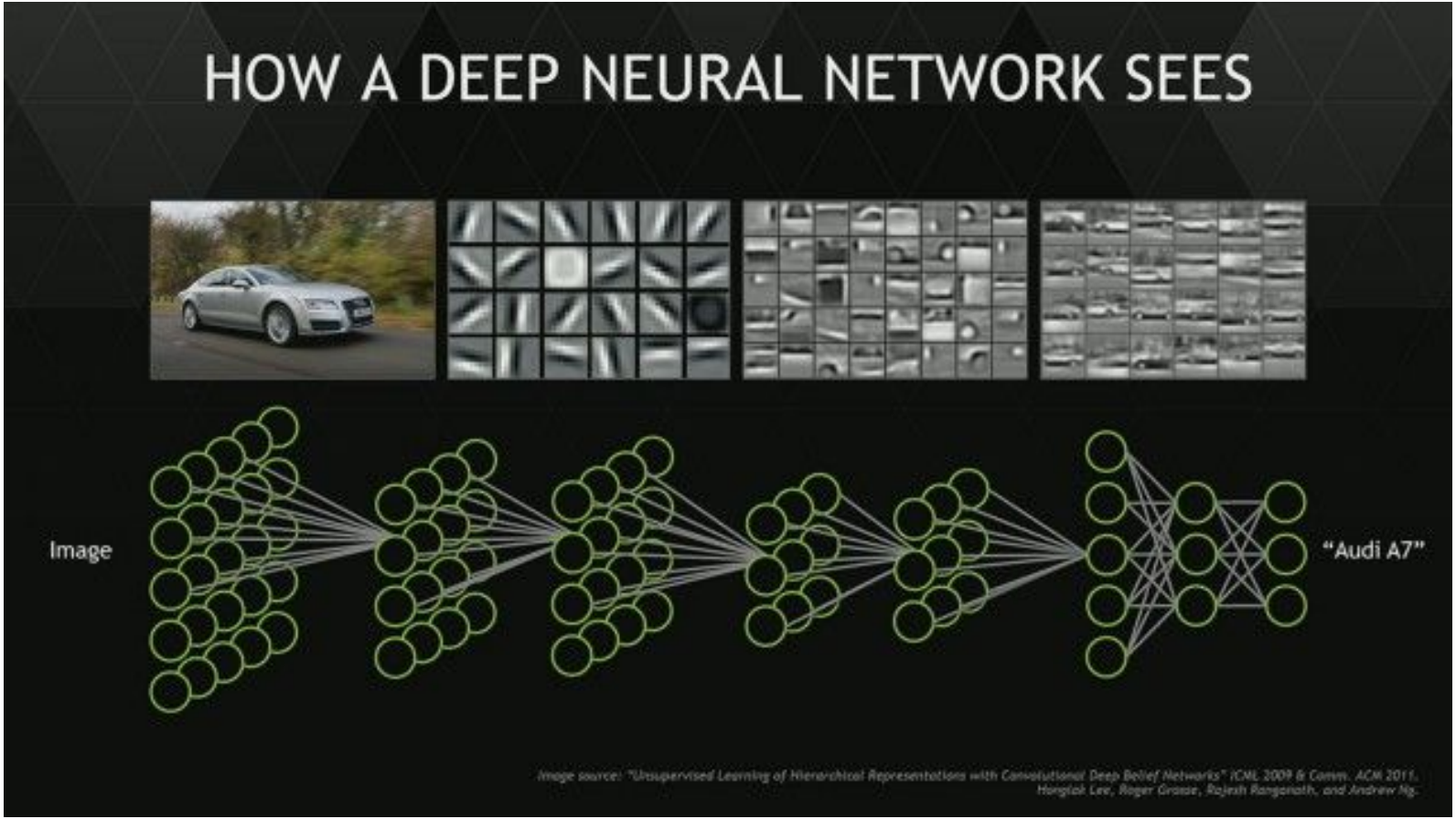


Image Source: https://towardsdatascience.com/understanding-residual-networks-9add4b664b03

These weights tell the neuron to respond more to one input and less to another. Weights are adjusted when training — that's how the network learns. Basically, that's all there is to it.

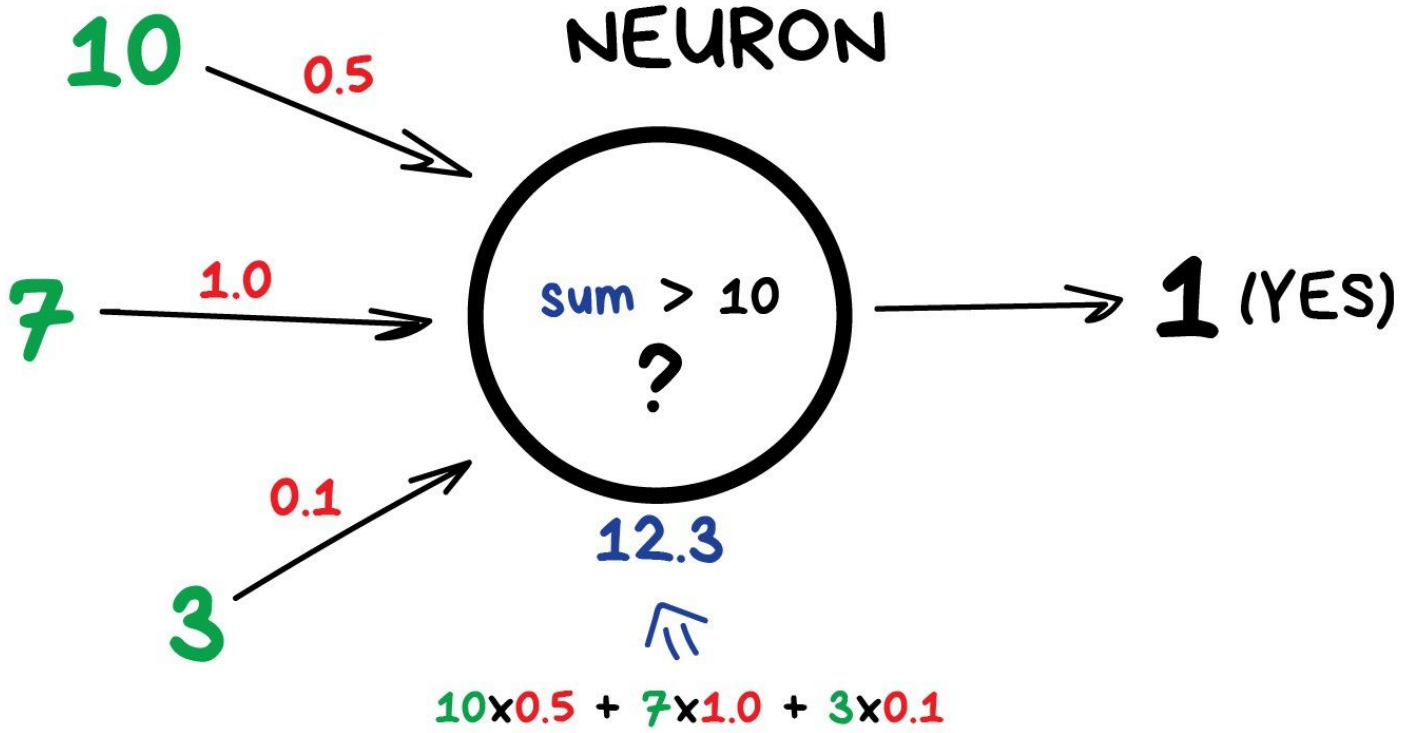


Image source: https://vas3k.com/blog/machine_learning/

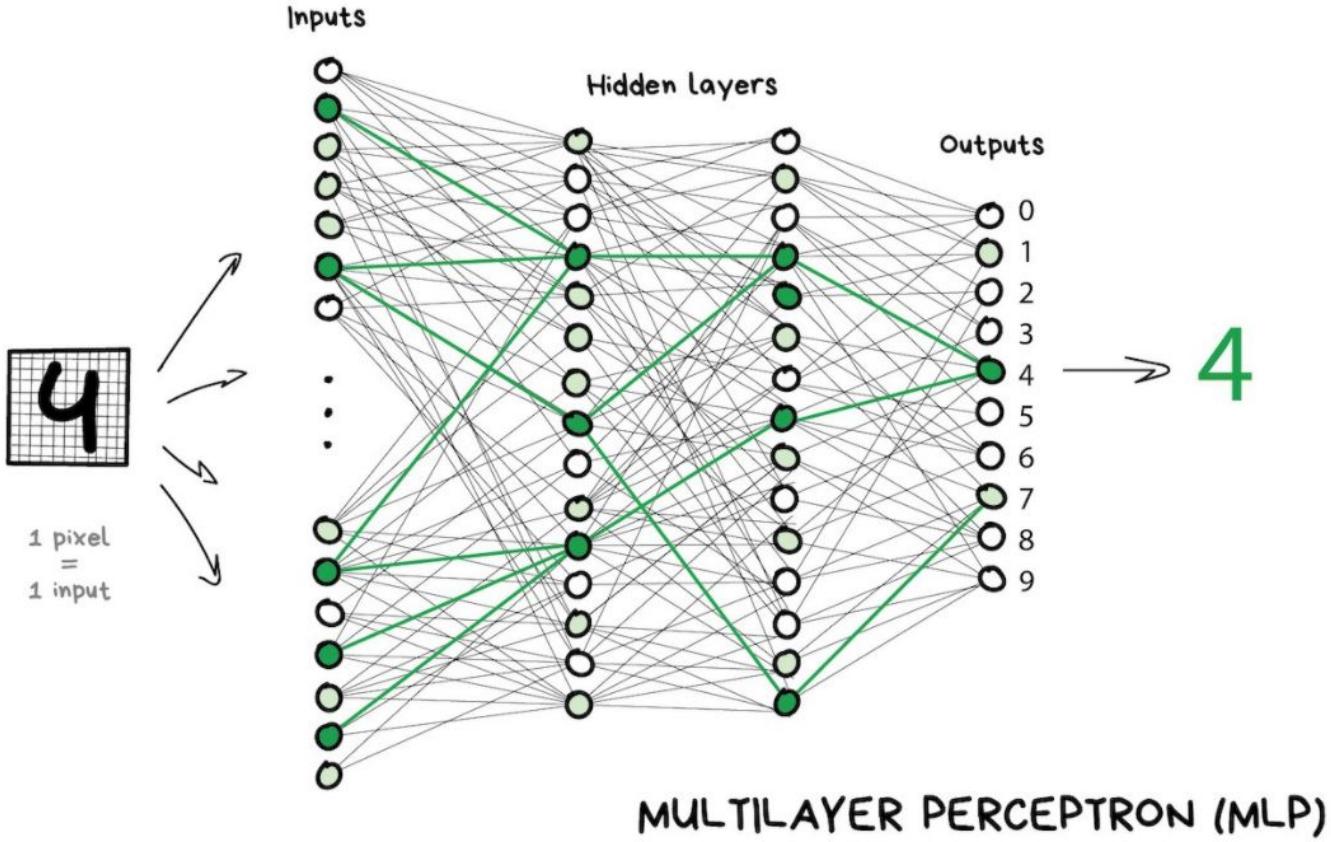


Image source: https://vas3k.com/blog/machine_learning/

Manually labeling used to be the way...

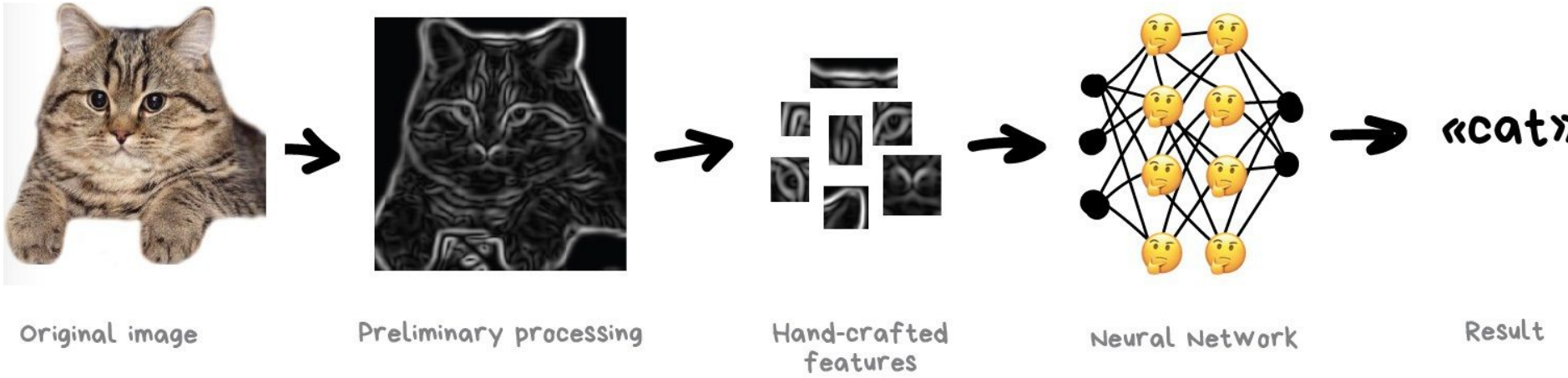
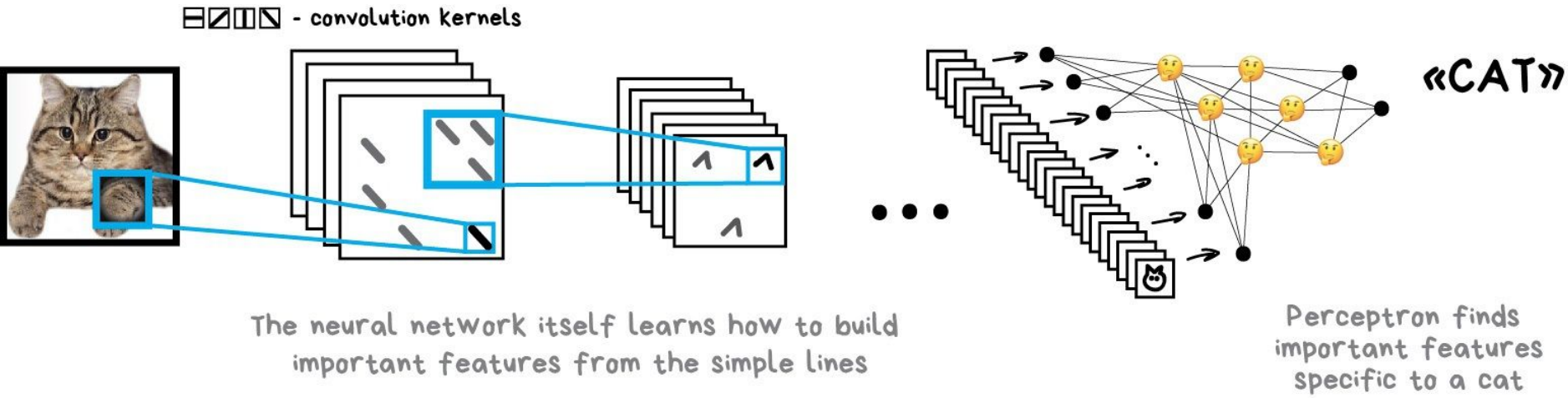


Image source: https://vas3k.com/blog/machine_learning/

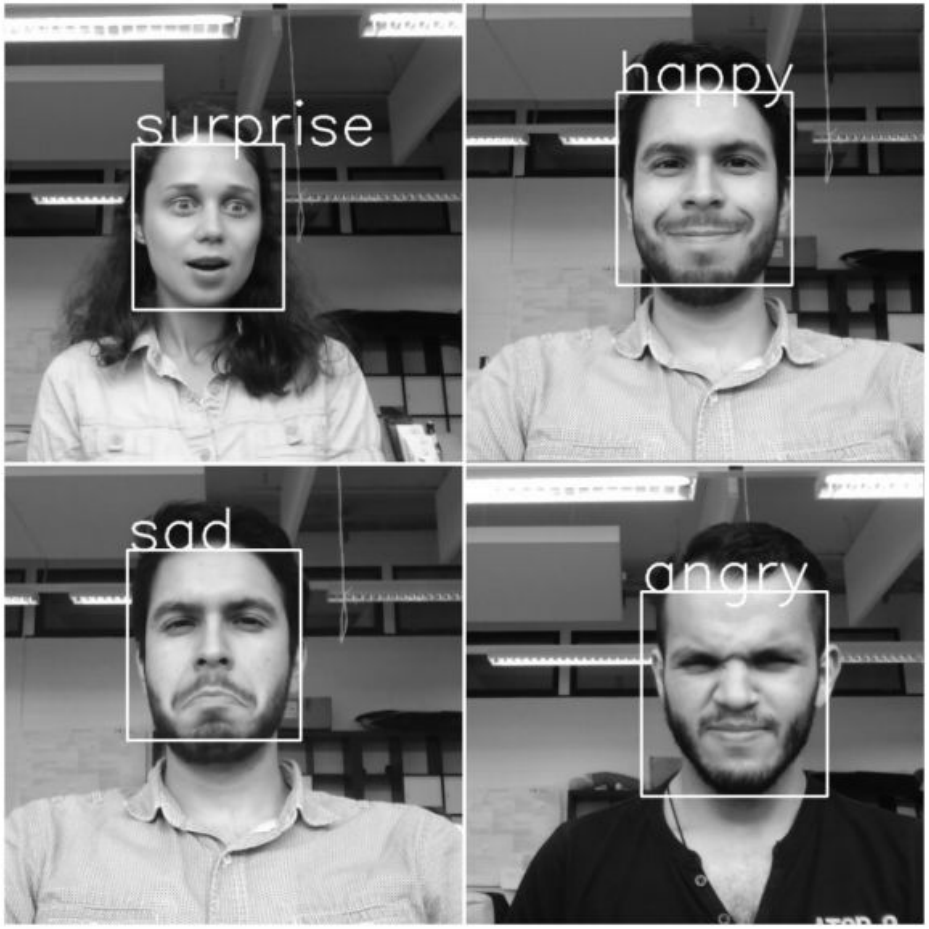
CNNs avoid manual labeling



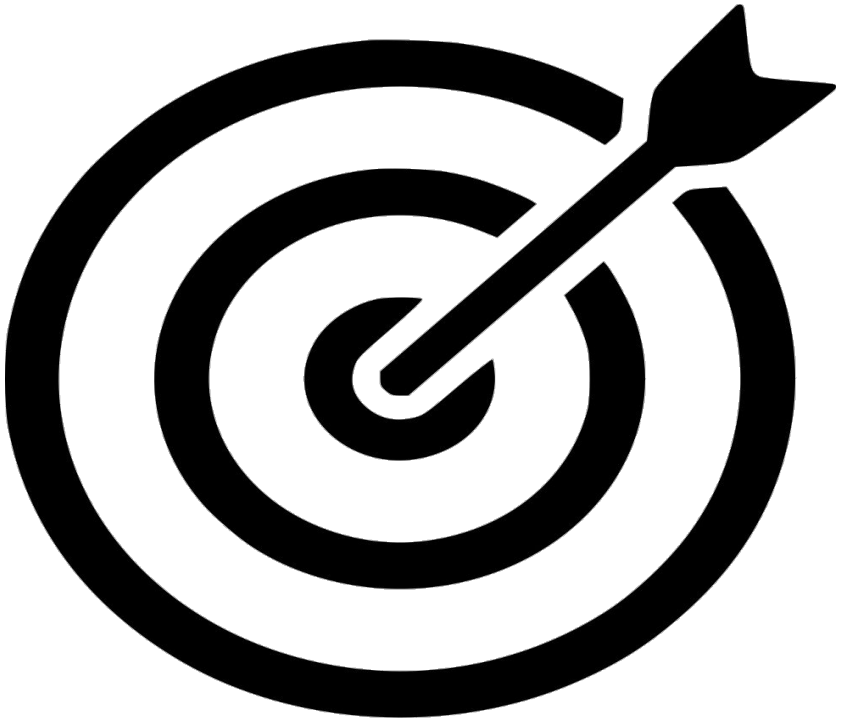
"CNNs are all the rage right now. They are used to search for objects on photos and in videos, face recognition, style transfer, generating and enhancing images, creating effects like slow-mo and improving image quality. Nowadays CNNs are used in all the cases that involve pictures and videos."

CONVOLUTIONAL NEURAL NETWORK (CNN)

Image source: https://vas3k.com/blog/machine_learning/



https://github.com/oarriaga/face_classification



model assessment

continuous variable prediction

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

A few outliers can lead to a big increase in RMSE, even if all the other predictions are pretty good

categorical variable prediction

continuous variable prediction

$$\text{Accuracy} = \frac{\#\text{ of samples predicted correctly}}{\#\text{ of samples predicted}} * 100$$

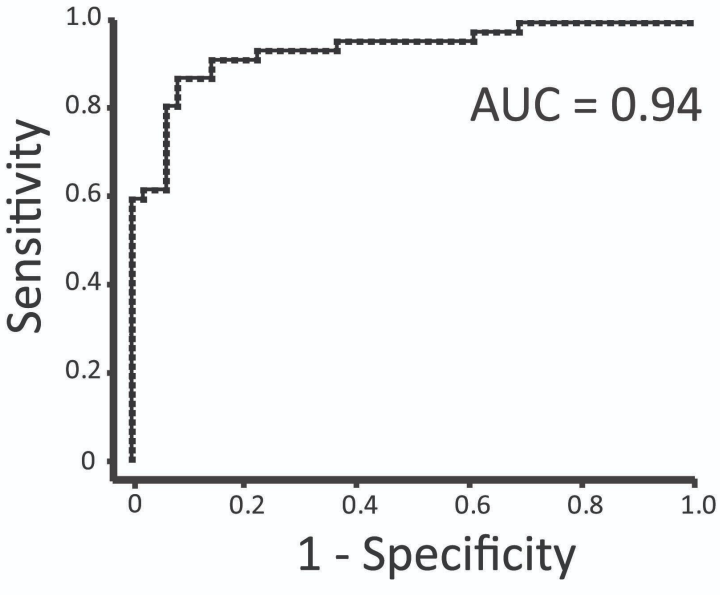
categorical variable prediction

		Actual	
		Positive	Negative
Predicted	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

A 2x2 table is a type of confusion matrix

Sensitivity

$$\frac{TP}{TP + FN}$$



Specificity

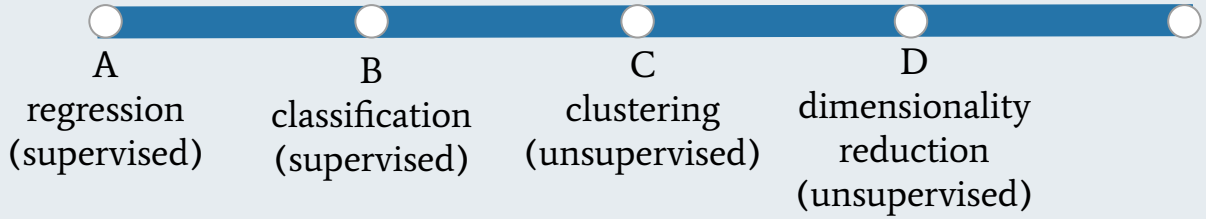
$$\frac{TN}{TN + FP}$$

Accuracy	What % were predicted correctly?
Sensitivity	Of those that *were* positives, what % were predicted to be positive?
Specificity	Of those that were *negatives*, what % were predicted to be negative?

Prediction Approach

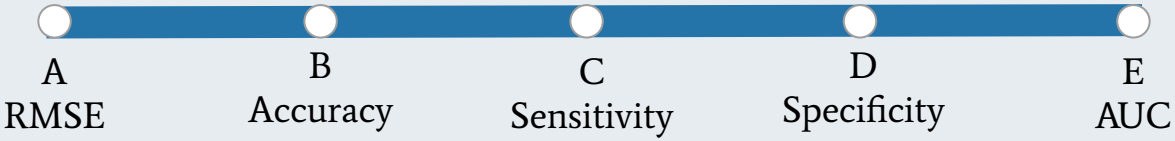
You've been given a dataset with a number of features and have been asked to predict each individual's age.

What prediction approach would you use?



Prediction Approach

After predicting each person's age, how would you assess your model?



Prediction Approach

Which would be the error value you'd want from your model?