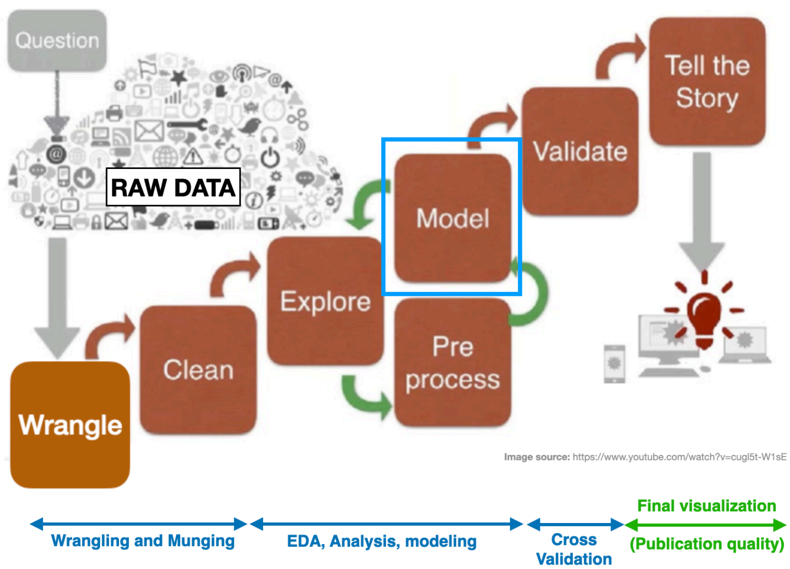
Source[1]

Machine Learning paradigm overview

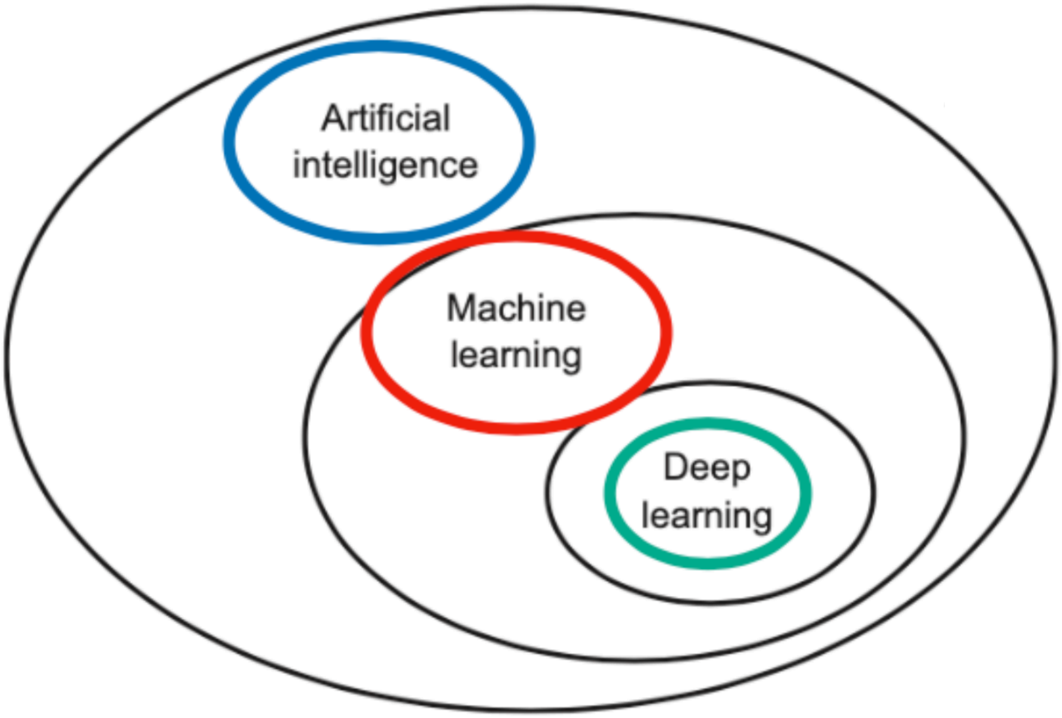
A brief introduction

Typically we want a predictive model the capture trends in the data



Artificial intelligence (AI)

- AI is any effort to automate intellectual tasks normally performed by humans
- This is a very broad category; It encompasses all forms of in-organic “intelligence”, e.g. an old calculator technically qualifies as a form of AI



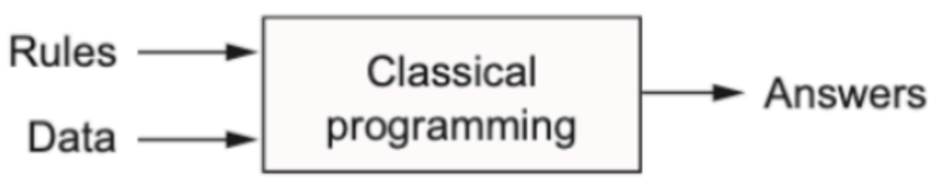
Traditional programming

- This is based on explicit instructions; the human tries to tell the machine what to do for EVERY possible situation
- Not scalable to large complex problems

Algorithm only knows what you explicitly tell it what to do, doesn't adapt

explicit rules: for example hard coded "if statements"

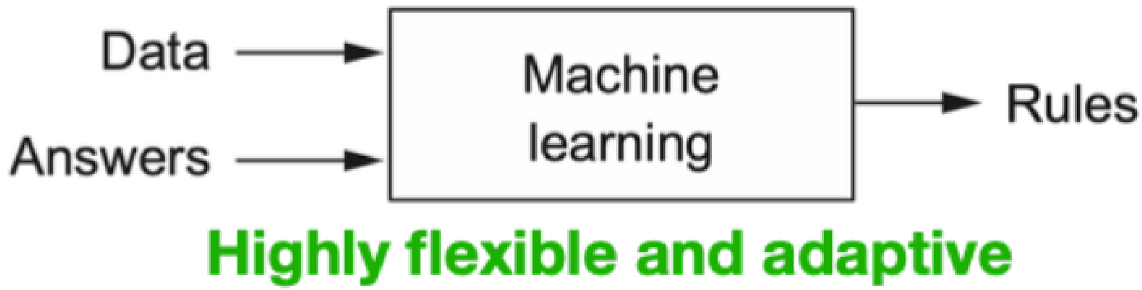
"symbolic AI" (1950-1980's) **Inflexible (i.e. rigid)**



- Can the algorithm instead *learn* from data, instead of being told exactly what to do?

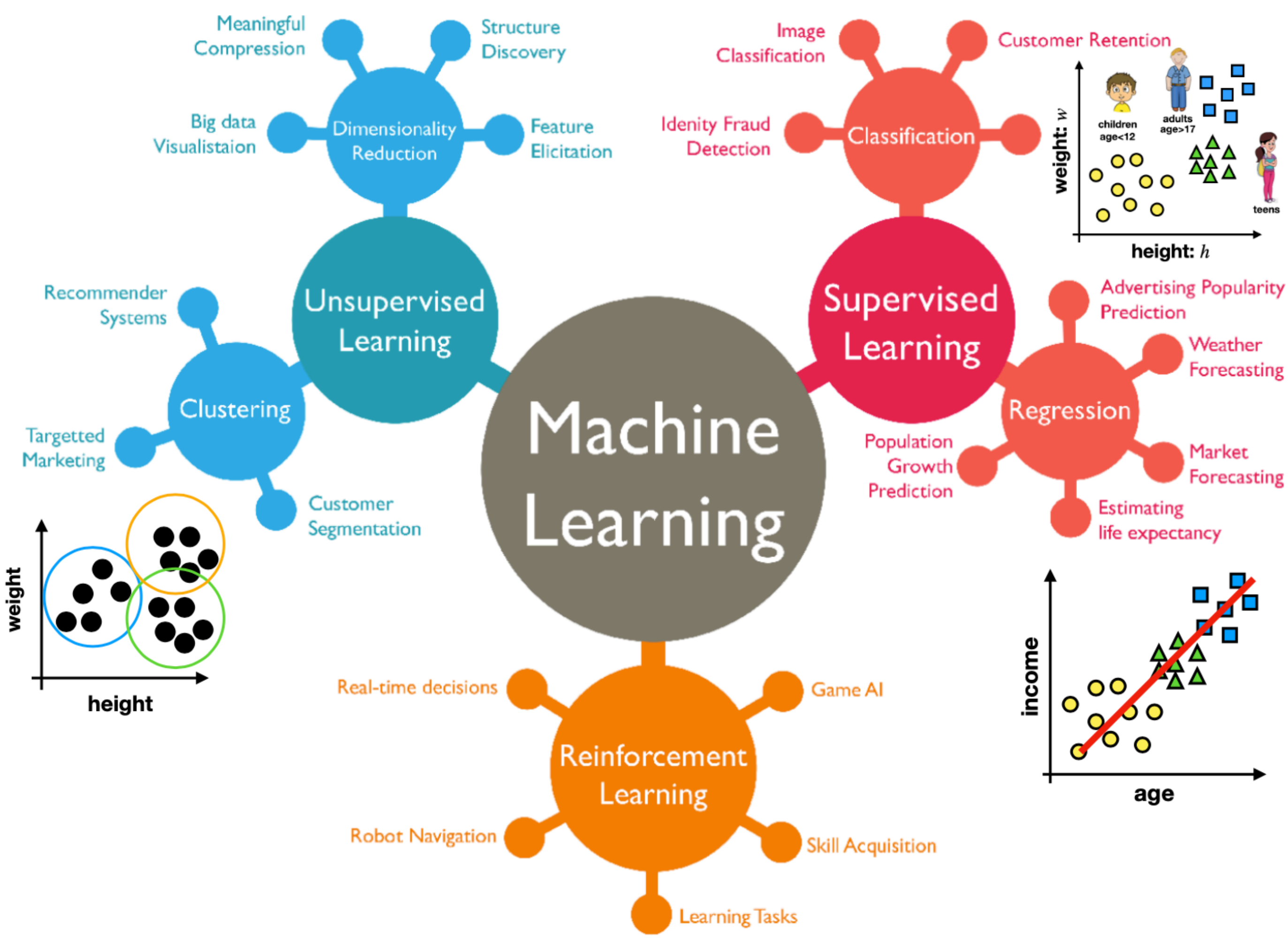
Machine learning

- Machine learning (ML): Machine learns rules from data given examples of what's expected (1980 to present)



- **Definition-1:** Machine learning is searching for useful representations of some input data, within a predefined space of possibilities, using guidance from a feedback signal.
- **Definition-2:** A field of inquiry devoted to understanding and building methods that ‘learn’, that is, methods that leverage data to improve performance. [1]
- **Training**: Learning describes an automatic search process to find better representations of the data.

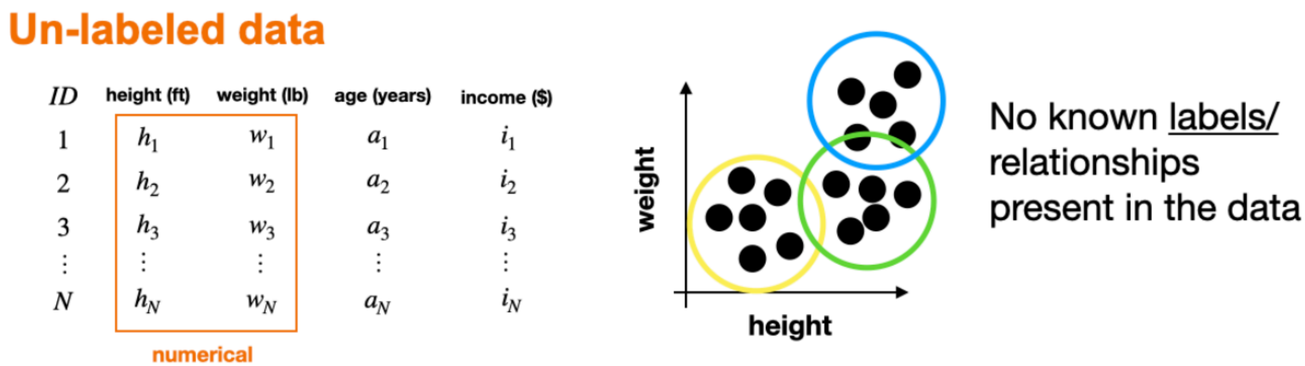
Machine learning paradigms



SOURCE: https://www.wordstream.com/blog/ws/2017/07/28/machine-learning-applications

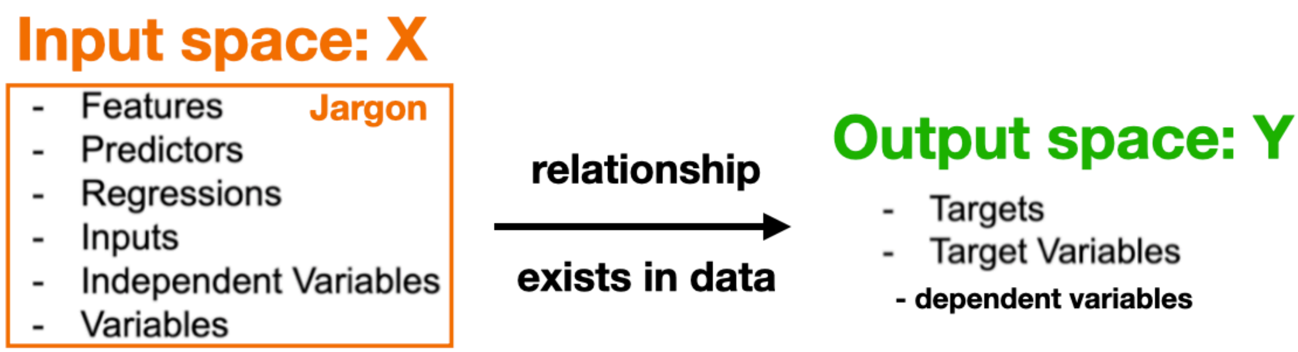
Unsupervised learning

- The nature of the `data dictates` what type of `modeling options` are available
- **Un-labeled data**: No a-priori known relationships exist in the data
- No outcome variable, just a set of predictors (features) measured on a set of samples.
- The objective is more fuzzy - find groups of samples that behave similarly, find features that behave similarly, find linear combinations of features with the most variation.
- `Very useful pre-processing and data exploration tool`

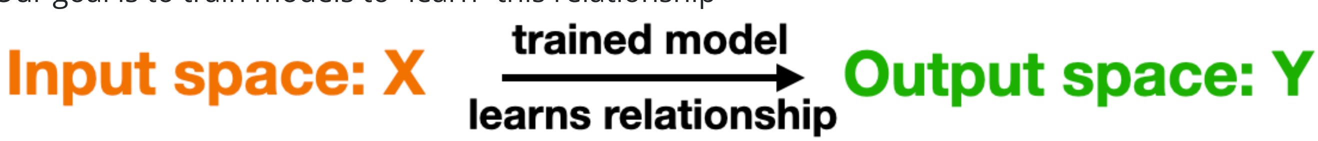


Supervised learning: Big picture

- Models leverage structure and relationships in the data



- Our goal is to train models to "learn" this relationship



- Once trained, we use the model to make predictions Y, given an `un-seen` input X
- If no relationship in the data exists, then the model with be useless, the stronger the relationship, the better a well trained model will likely be

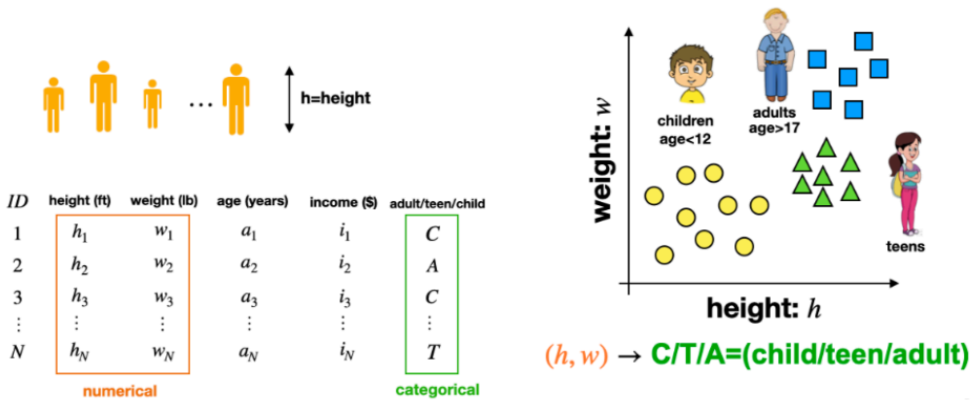
Supervised learning: Formalism

- **Given the following**
  - Outcome measurement Y (also called dependent variable, response, target).
  - Vector of p predictor measurements X (inputs, features, independent variables).
  - **Regression problem**, Y is quantitative (e.g price, blood pressure).
  - **Classification problem** Y takes values in a finite, unordered set (survived/died, digit $0-9$, cancer class of tissue sample).

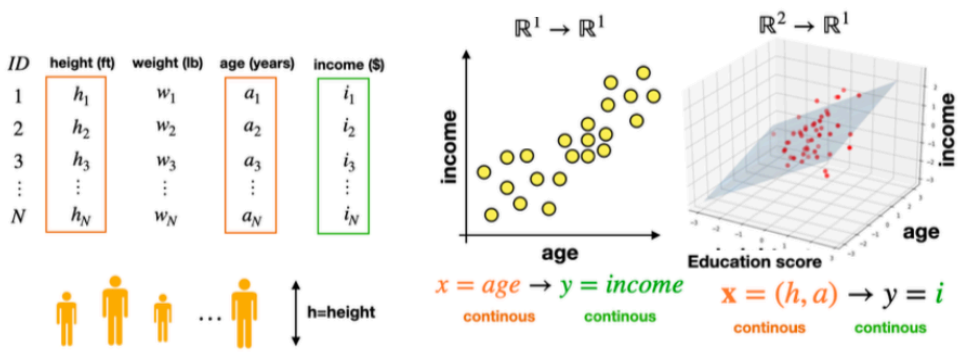
- We have training data $(x_1, y_1), \ldots, (x_N, y_N)$. These are observations (examples, instances) of these measurements.
- **Train**: Develop a model to MIMIC the data (optimization)
- **Given a trained model**
  - Accurately predict unseen test cases.
  - Understand which inputs affect the outcome, and how.
  - Assess the quality of our predictions and inferences.

Supervised learning: Example

- The structure and type of data dictate what type of modeling options are available
- In many cases, "learning" just means fitting a parameterization of a function
- **Classification problem**: Known Numerical $\longrightarrow$ Categorical relationships exist.



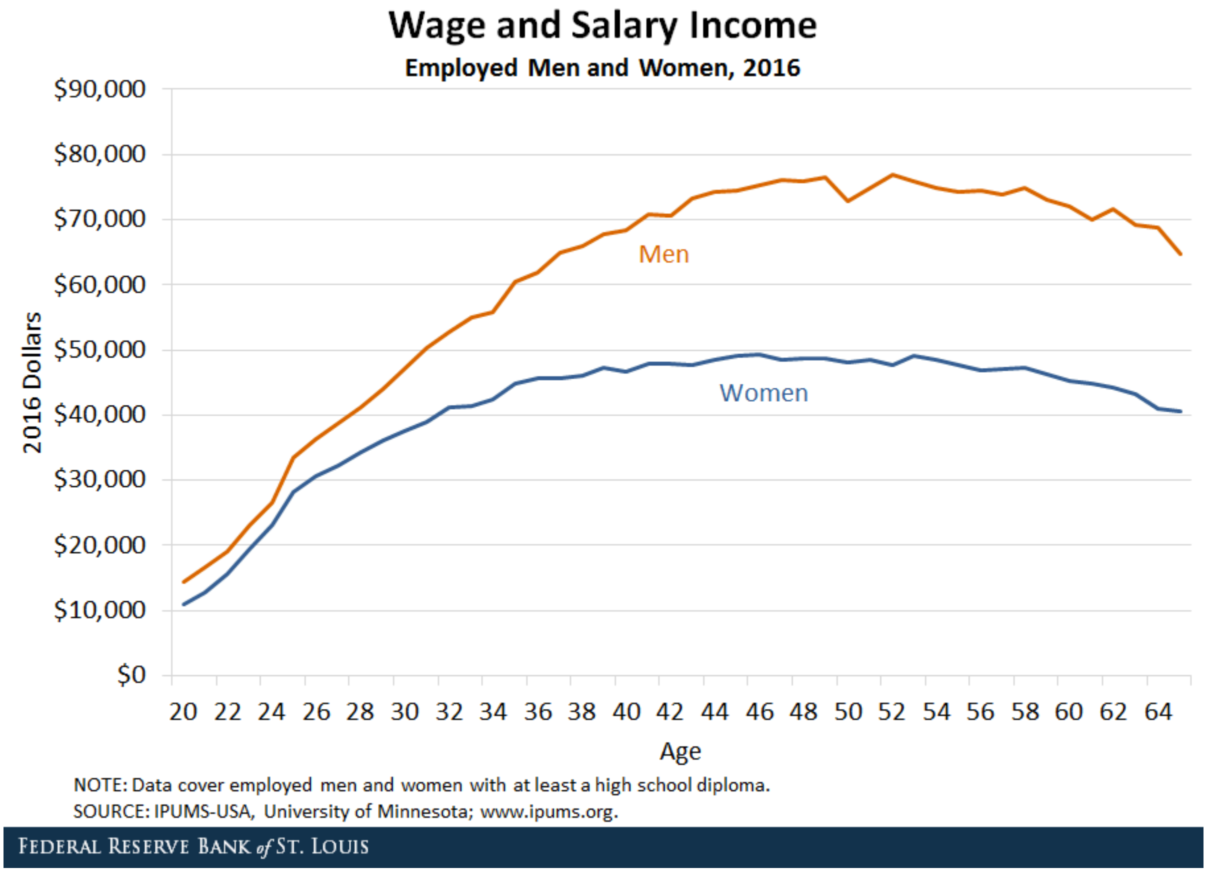
- **Regression problem**: Known Numerical $\longrightarrow$ Numerical relationships exist.



Age and income

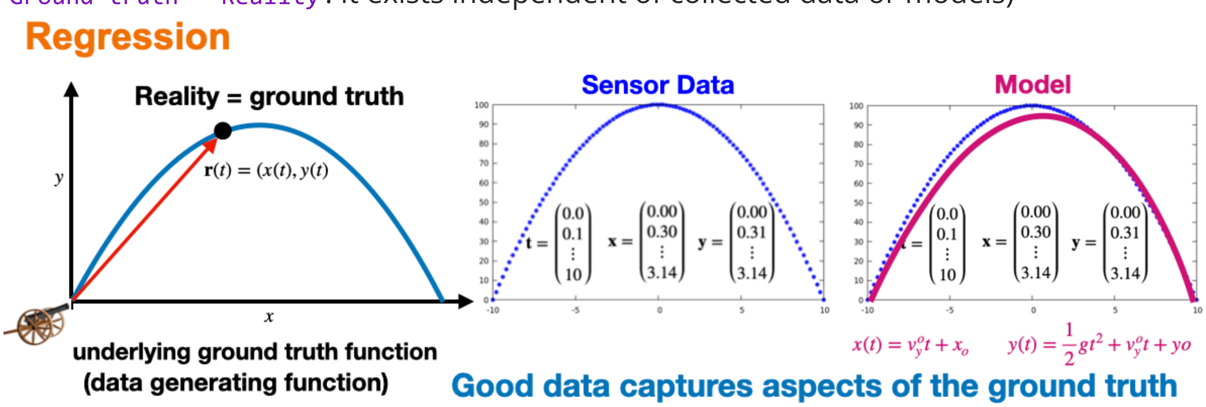
- **Note**: on the previous slide the relationship between age and income appeared `linear`
  - That was just a `cartoon` to demonstrate a point.
- In reality the relationship is `non-linear` with a plateau as time progresses.

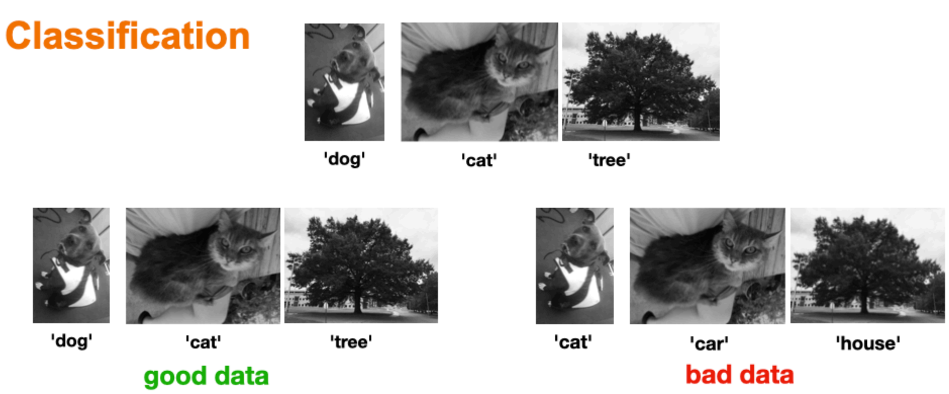
- The phenomenon would be better modeled with either a logistic function or a parabola.



The ground truth

- `Ground truth` = `Reality`: It exists independent of collected data or models)



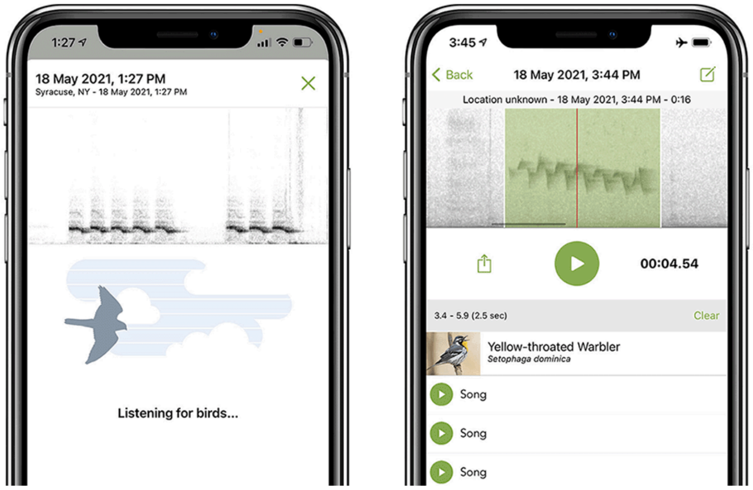


- Good data captures the ground truth, good models learn the ground truth

- `Data Provenance`: Tracing data origin, transformations, and changes. Ensures accountability, reliability, and quality in data-driven processes and decisions.

Real world example:

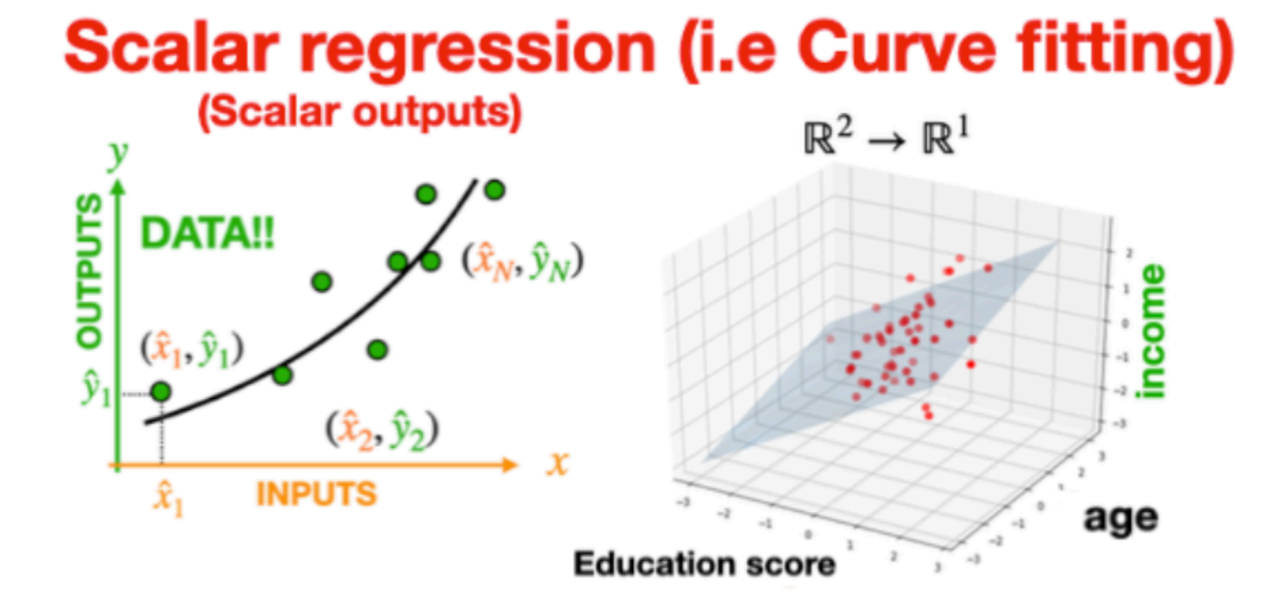
- Identify the birds you see or hear with Merlin Bird ID
    - https://merlin.allaboutbirds.org/
- Trained ML models on supervised sound data to map between;
    - `X=sound fingerprint` $\longrightarrow$ `Y=bird species`
    - This is an example of supervised multi-class classification

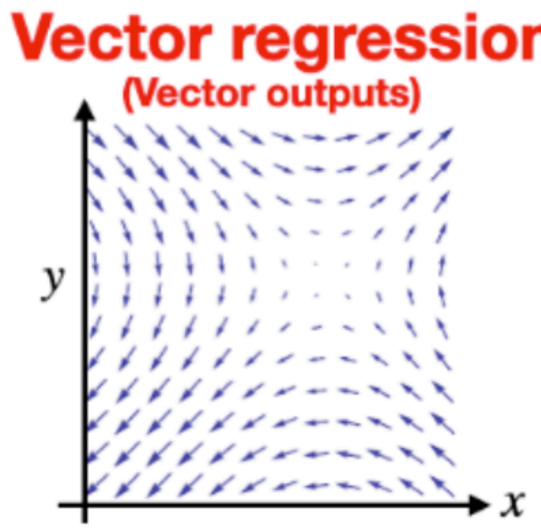


- `Prediction:` Now that the model is trained, anyone in the world figure out which birds are around based on a simple phone recording.

Regression tasks

- Scalar Regression: Targeting a single continuous scalar value, like predicting prices.
- Vector Regression: Targeting a set (i.e. vector) of continuous values, like bounding box coordinates in an image.

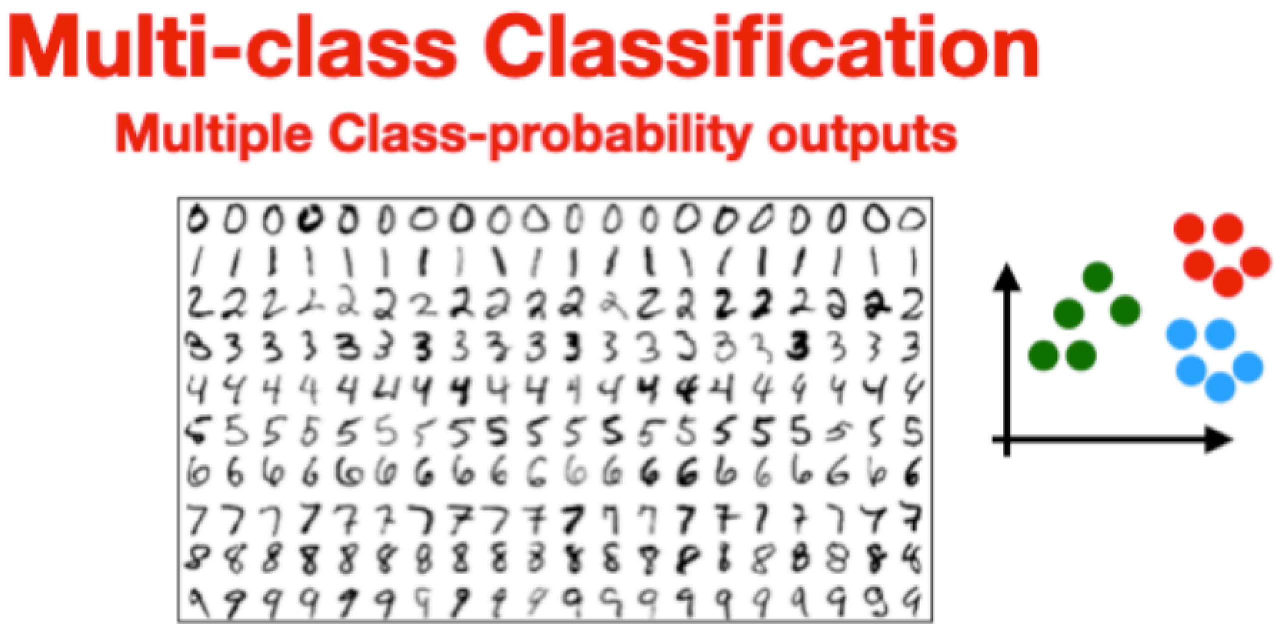




ID	height (ft)	weight (lb)	age (years)	income ($)
1	h_1	w_1	a_1	i_1
2	h_2	w_2	a_2	i_2
3	h_3	w_3	a_3	i_3
⋮	⋮	⋮	⋮	⋮
N	h_N	w_N	a_N	i_N
	numerical		numerical	

Classification tasks

- Binary Classification: Categorizing inputs into two categories.
- Multiclass Classification: Categorizing inputs into more than two categories.



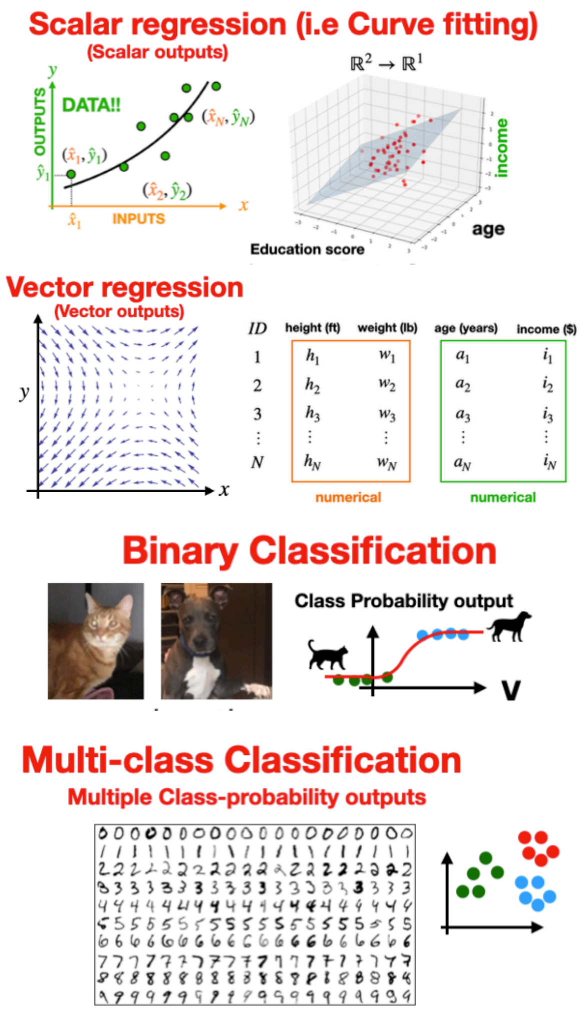
Supervised learning: Jargon

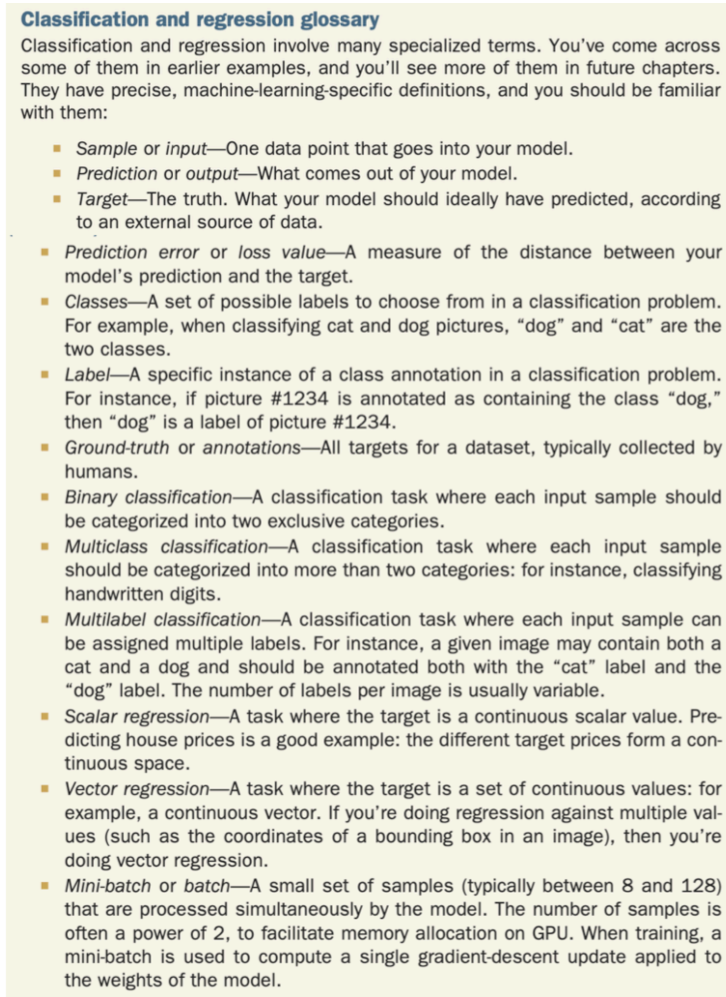
Classification and regression involve specialized terms with precise, machine-learning-specific definitions.

- Sample/Input: A data point fed into the model.
- Prediction/Output: The model's output.
- Target: The correct output from external data.
- Prediction Error/Loss: The difference between prediction and target.
- Classes: Possible labels in a classification task, like "cat" or "dog."
- Label: A specific annotation for a sample, such as "dog" for a given picture.
- Ground-Truth/Annotations: Dataset targets, typically human-collected.
- Binary Classification: Categorizing inputs into two categories.
- Multiclass Classification: Categorizing inputs into more than two categories.
- Multilabel Classification: Assigning multiple labels to a single input.
- Scalar Regression: Targeting a continuous scalar value, like predicting prices.
- Vector Regression: Targeting a set of continuous values, like bounding box coordinates.

- Mini-batch/Batch: A small set of samples processed together for gradient-descent updates.

Source:[1]





Classification and regression glossary

Classification and regression involve many specialized terms. You've come across some of them in earlier examples, and you'll see more of them in future chapters. They have precise, machine-learning-specific definitions, and you should be familiar with them:

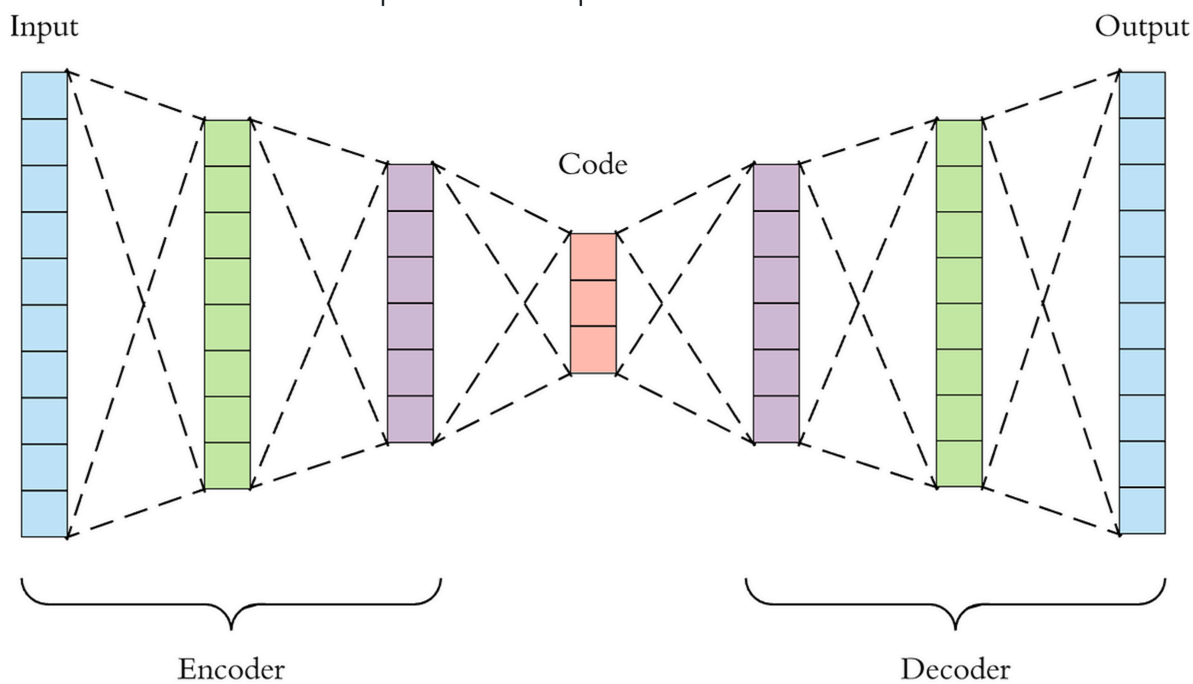
- *Sample* or *input*—One data point that goes into your model.
- *Prediction* or *output*—What comes out of your model.
- *Target*—The truth. What your model should ideally have predicted, according to an external source of data.
- *Prediction error* or *loss value*—A measure of the distance between your model's prediction and the target.
- *Classes*—A set of possible labels to choose from in a classification problem. For example, when classifying cat and dog pictures, "dog" and "cat" are the two classes.
- *Label*—A specific instance of a class annotation in a classification problem. For instance, if picture #1234 is annotated as containing the class "dog," then "dog" is a label of picture #1234.
- *Ground-truth* or *annotations*—All targets for a dataset, typically collected by humans.
- *Binary classification*—A classification task where each input sample should be categorized into two exclusive categories.
- *Multiclass classification*—A classification task where each input sample should be categorized into more than two categories: for instance, classifying handwritten digits.
- *Multilabel classification*—A classification task where each input sample can be assigned multiple labels. For instance, a given image may contain both a cat and a dog and should be annotated both with the "cat" label and the "dog" label. The number of labels per image is usually variable.
- *Scalar regression*—A task where the target is a continuous scalar value. Predicting house prices is a good example: the different target prices form a continuous space.
- *Vector regression*—A task where the target is a set of continuous values: for example, a continuous vector. If you're doing regression against multiple values (such as the coordinates of a bounding box in an image), then you're doing vector regression.
- *Mini-batch* or *batch*—A small set of samples (typically between 8 and 128) that are processed simultaneously by the model. The number of samples is often a power of 2, to facilitate memory allocation on GPU. When training, a mini-batch is used to compute a single gradient-descent update applied to the weights of the model.

Source: Deep Learning with Python: FRANCOIS CHOLLET

Self supervised

- Training a model to learn from unlabeled data by creating surrogate supervisory tasks
- Enables it to acquire representations without manual labeling.
- Basically same as supervised learning, but we trick the data to learn from itself.

- Auto-encoders are an example of a Self supervised neural network

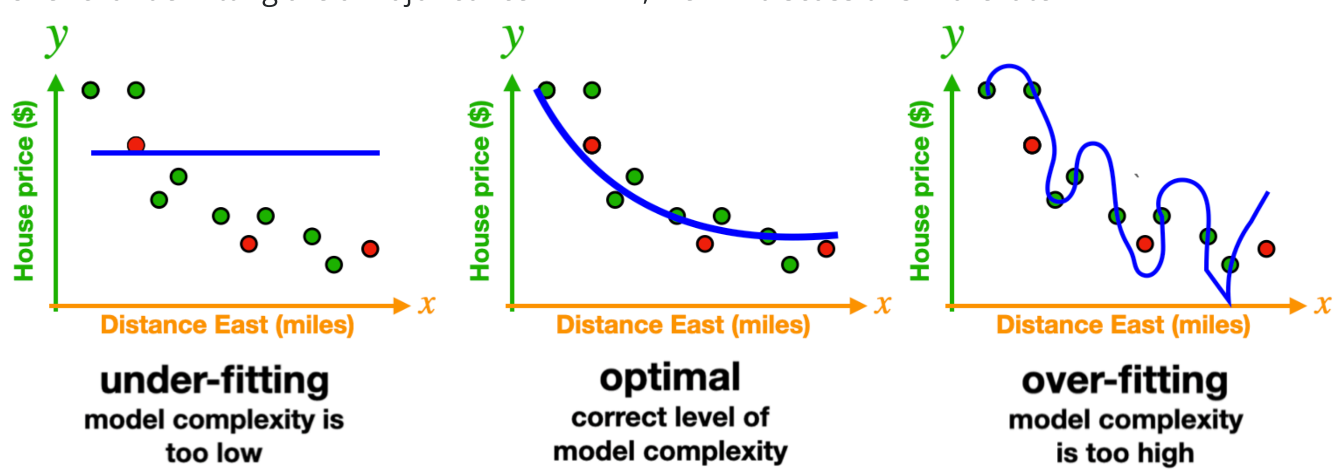


- Sometimes self-supervised & un-supervised are used inter-changeably, this is incorrect

Image source: [click here]

Monitoring model over-fitting

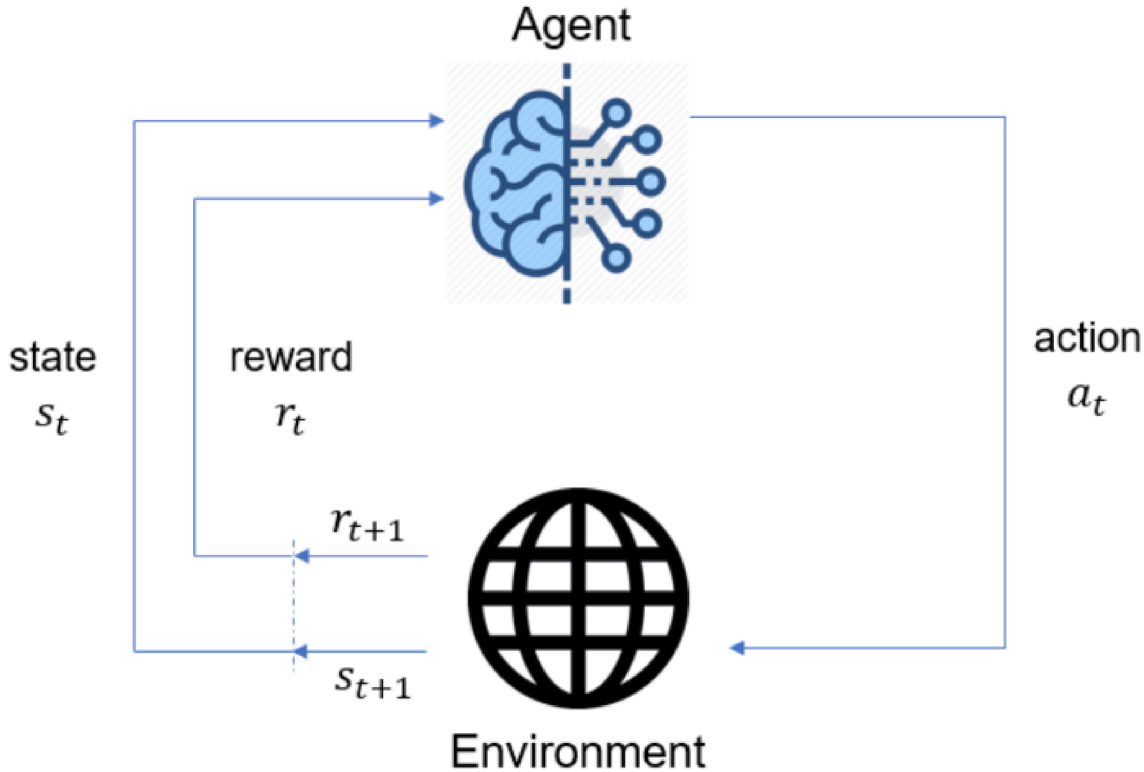
- Over & under fitting are a major concern in ML, we will discuss this more later



Reinforcement Learning (RL)

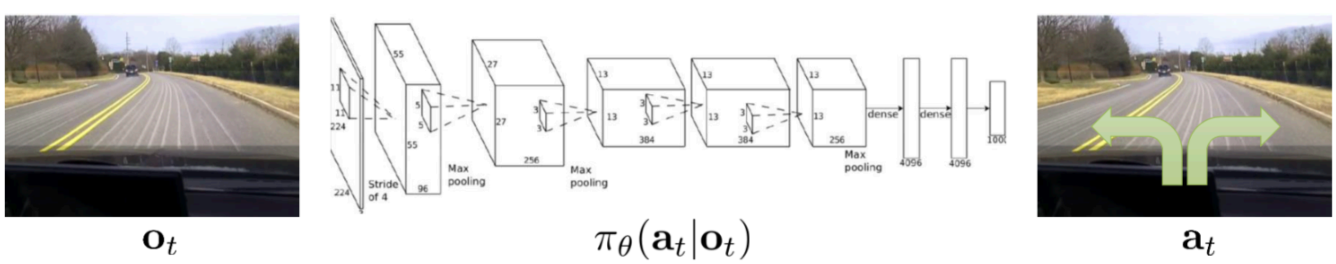
- RL creates AI agents for optimal decisions in complex, uncertain environments, maximizing long-term rewards through defined methodologies.
- The agent learns by trial and error, forming a 'policy' for diverse situations, guided by rewards. Over time, decision-making improves, paralleling human learning in RL.

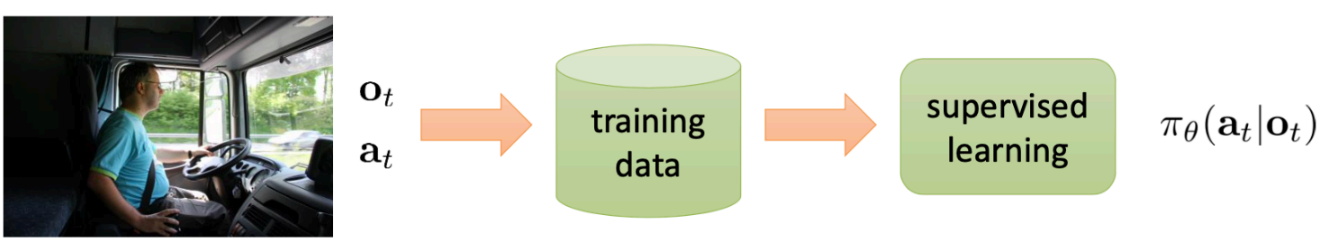
- In short, `RL = temporal decision making` in complex environments[2]



Aside: Imitation Learning

- Given a large amount of data from a trustworthy agent (i.e. the truck driver). We can actually train an RL in a supervised fashion (Image: Bojarski et al. 2016)

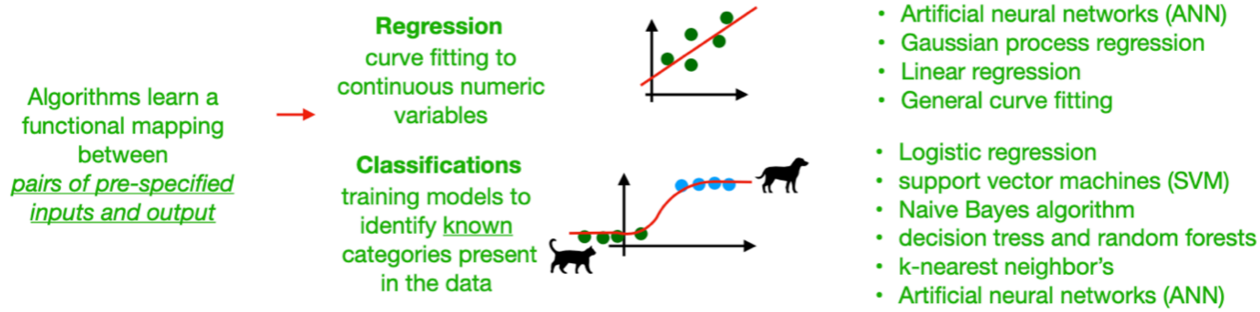




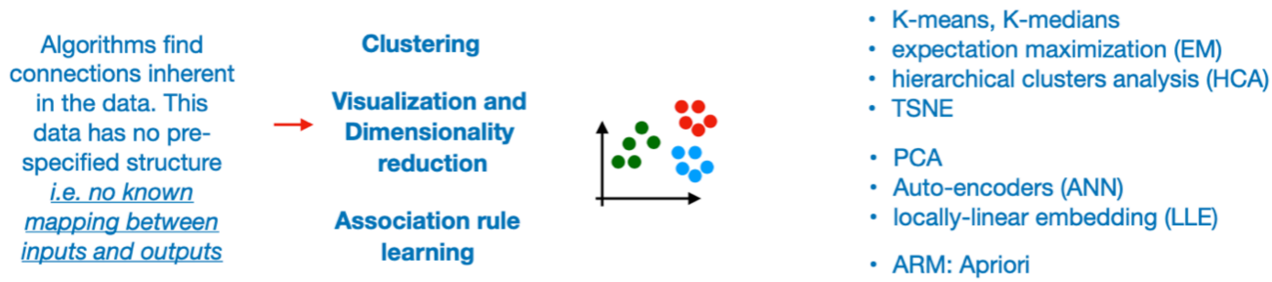
behavioral cloning

Algorithm summary

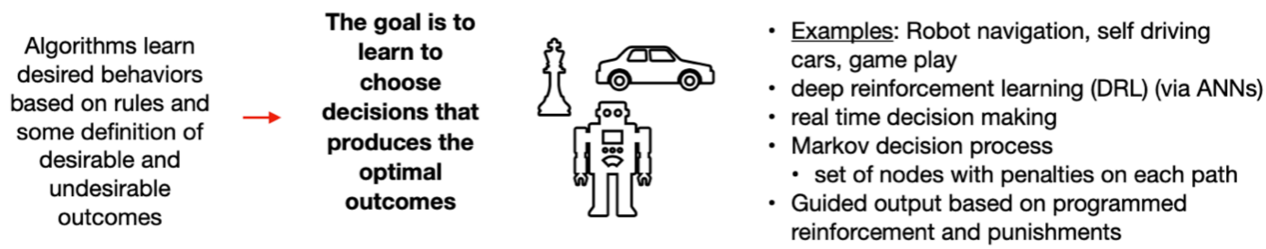
- **Supervised learning**.



- **Un-supervised learning**.



- **Reinforcement learning**.



References

1. 2018: Deep learning with python, francois chollet (1st edition).
2. 2020: Mastering reinforcement learning with python, enes bilgin (1st edition).

Footnotes

1. https://en.wikipedia.org/wiki/Machine_learning ↩