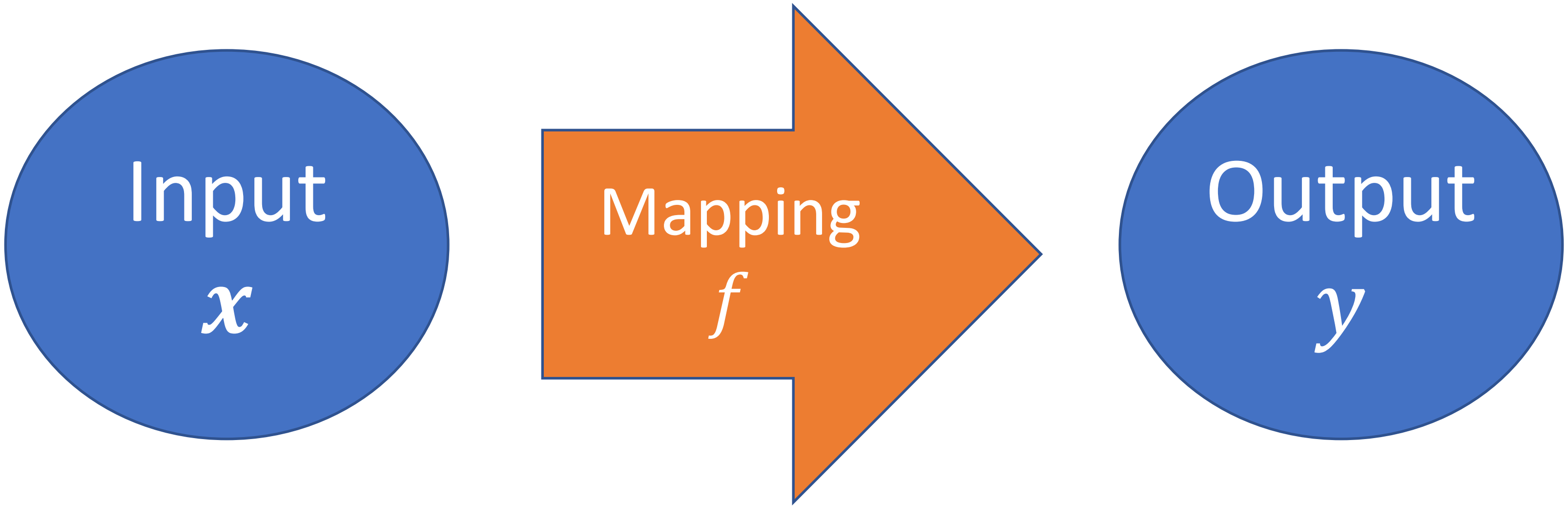
Introduction to Machine Learning (and Notation)

David I. Inouye

Outline

- Supervised learning
  - Regression
  - Classification
- Unsupervised learning
  - Dimensionality reduction (PCA)
  - Clustering
  - Generative models
- Other key concepts
  - Generalization
  - No free lunch theorem

The goal of supervised learning is to estimate a mapping (or function) between input and output



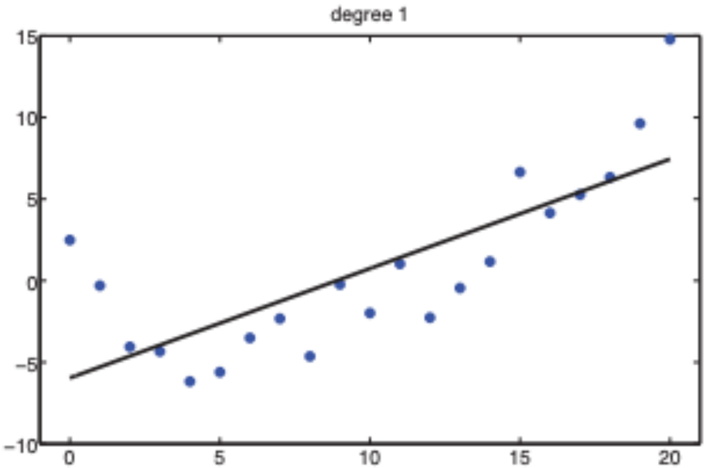
The goal of supervised learning is to estimate a mapping (or function) between input and output *given only input-output examples*

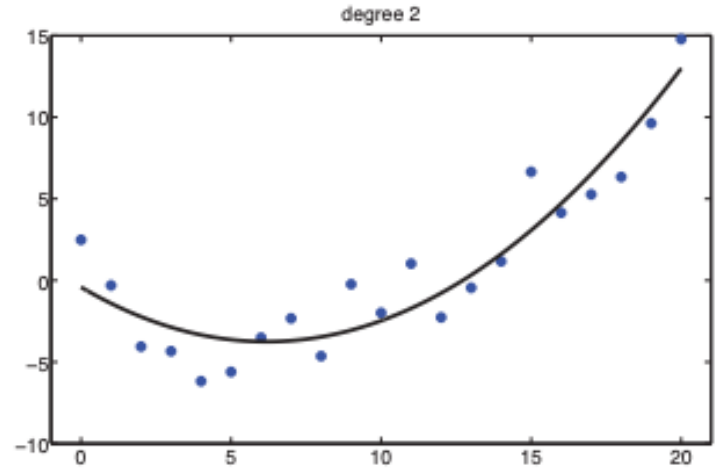
The set of input-output pairs is called a training set, denoted by $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$

- Input $\boldsymbol{x}_i$
  - Called features (ML), attributes, or covariates (Stats). Sometimes just variables.
  - Can be numeric, categorical, discrete, or nominal.
  - Examples
    - [height, weight, age, gender]
    - $[x_1, x_2, \cdots, x_d]$ – A *d*-dimensional vector of numbers
    - Image
    - Email message

- Output y_i
  - Called output, response, or target (or label)
  - Real-valued/numeric output: e.g., $y_i \in \mathcal{R}$
  - Categorical, discrete, or nominal output: y_i from *finite* set, i.e., $y_i \in \{1,2,\cdots,c\}$

If the output y_i is numeric, then the problem is known as regression





NOTE: Input x does not have to be numeric. Only the output y must be numeric.

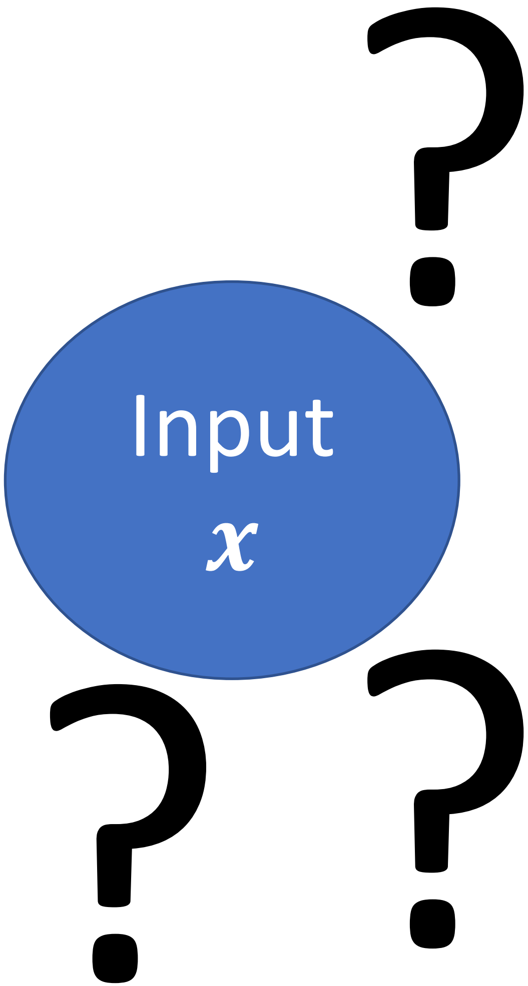
- Given height x_i, predict age y_i
- Predict GPA given SAT score
- Predict SAT score given GPA
- Predict GRE given SAT and GPA

If output is categorical, then the problem is known as classification

- Given height x, predict "male" ($y = 0$) or "female" ($y = 1$)

- Given salary x_1 and mortgage payment x_2, predict defaulting on loan ("yes" or "no")

The goal of <u>unsupervised learning</u> is to model or understand the input data directly

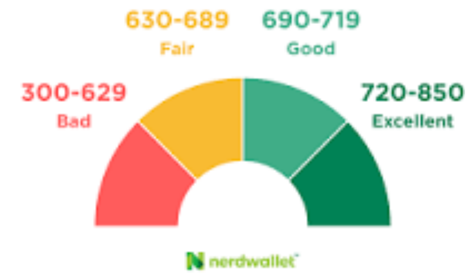


- Dimensionality reduction
- Clustering
- Generative models

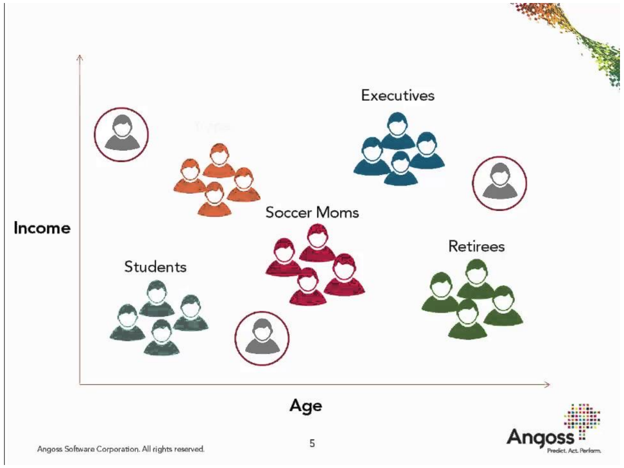
“What I cannot create I do not understand”
– Richard Feynman

In unsupervised learning, the training set is only a set of input values $\mathcal{D} = \{\boldsymbol{x}_i\}_{i=1}^{n}$

- [Dimensionality reduction] Estimate a single number that summarizes all variables of wealth (e.g. credit score)

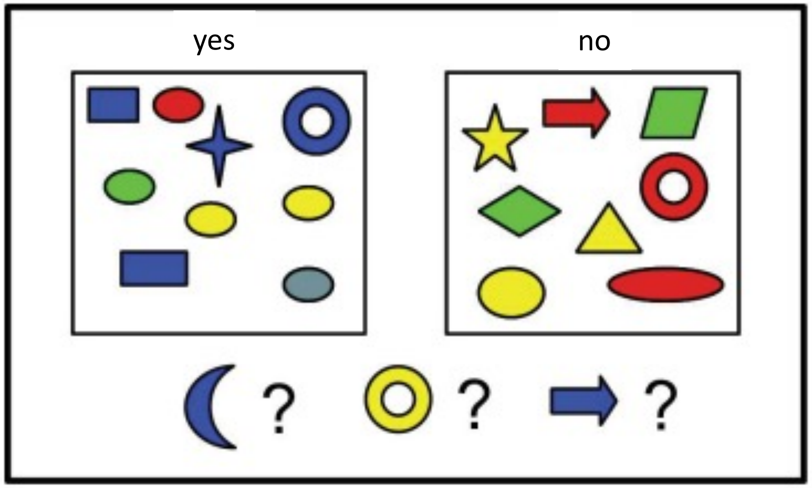


- [Clustering] Estimate natural groups of customers



- [Generative Models] Estimate the distribution of normal transactions to detect fraud (anomalies)

<u>Generalization</u> *beyond* the training set is the main goal of learning



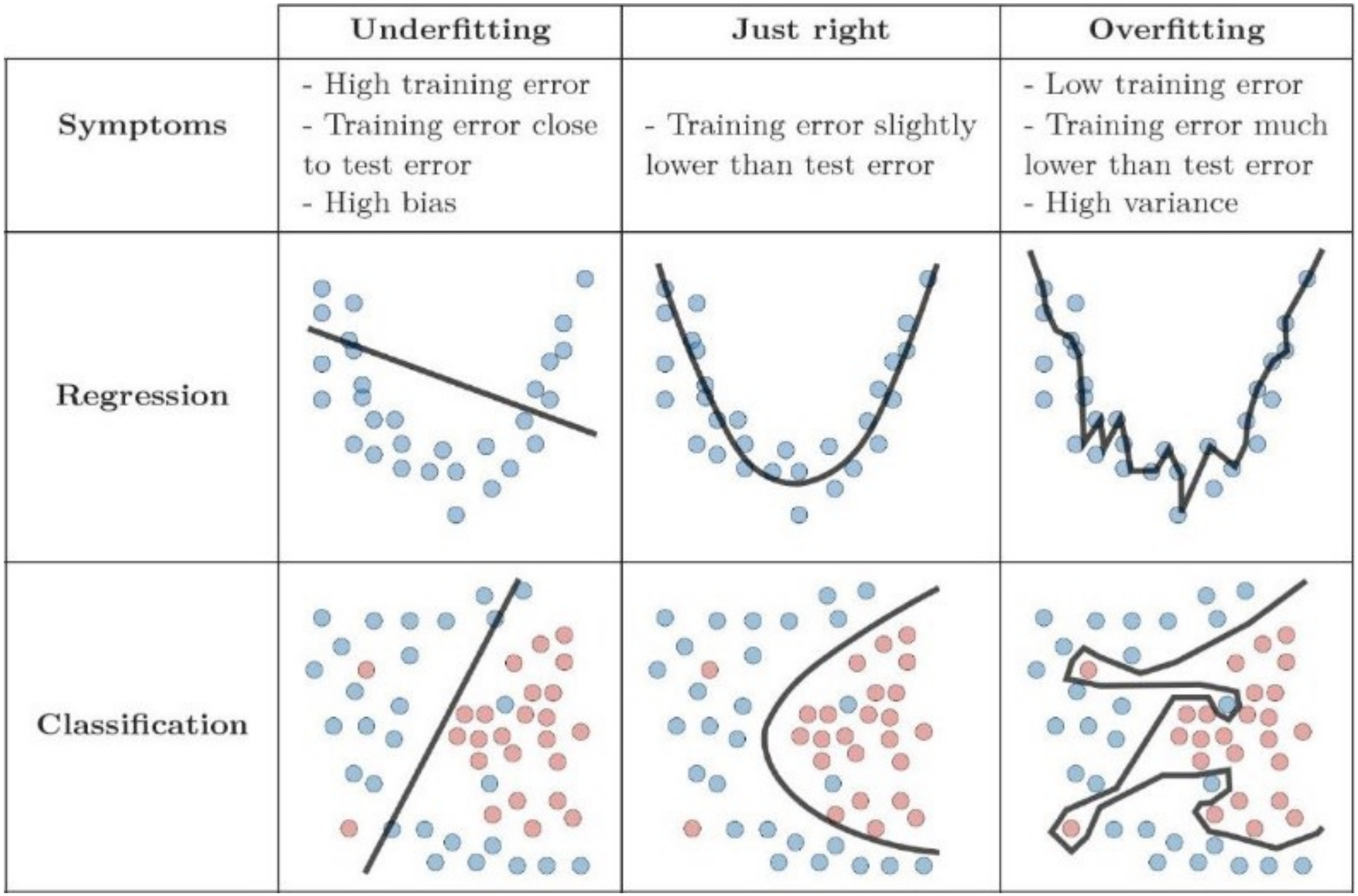
d features/attributes/covariates

n samples/ observations/ examples

	Color	Shape	Size (cm)	Is it good?	
x_1	Blue	Square	10	yes	y_1
x_2	Red	Ellipse	2.4	yes	y_2
	Red	Ellipse	20.7	no	

Example from Machine Learning: A Probabilistic Perspective, Ch. 1, Kevin P. Murphy, 2012.

<u>Generalization</u> *beyond* the training set is the main goal of learning

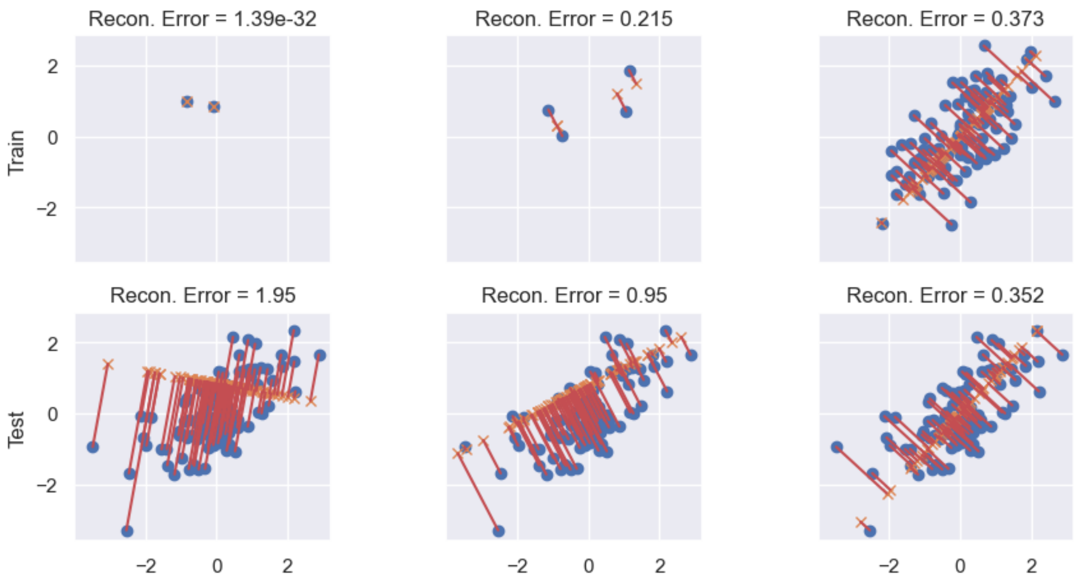


	Underfitting	Just right	Overfitting
Symptoms	- High training error - Training error close to test error - High bias	- Training error slightly lower than test error	- Low training error - Training error much lower than test error - High variance
Regression			
Classification			

Original source for figure unknown.

What does generalization look like for *unsupervised learning*?

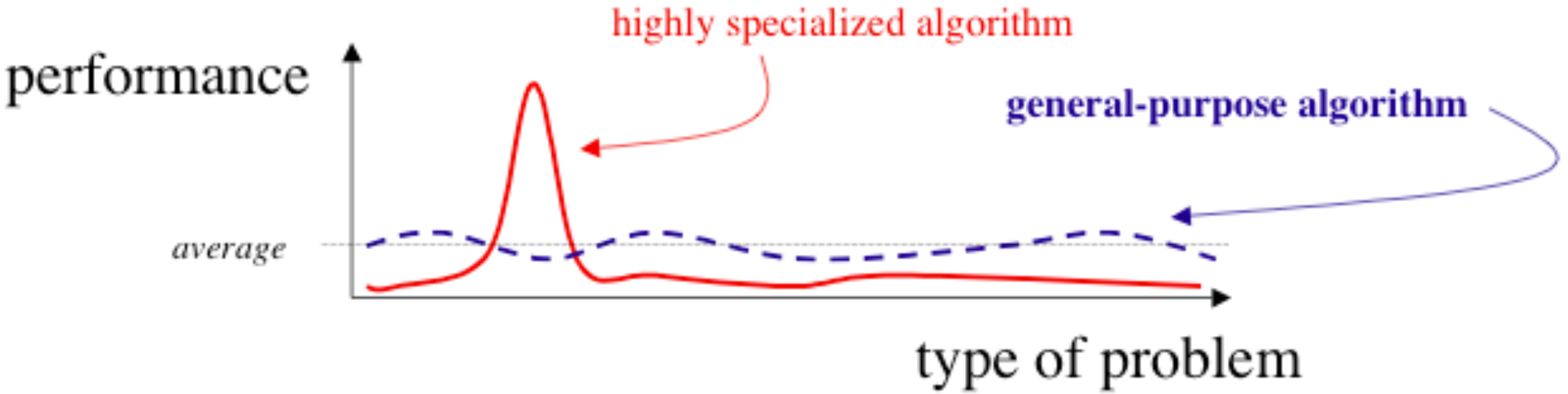
- Generalization in dimensionality reduction
  - Objective on train may be small but test may be large



- Generalization in generative models can be understood through the view of log likelihood.

“All models are wrong,
but some models are useful.”*

- All models are approximations
- All models make assumptions
- Assumptions are never perfect
- <u>No Free Lunch Theorem</u>



* George Box (Box and Draper 1987, page 424).