

Machine Learning for Newbies

[NUS Orbital 2016] – Mission Control 4A/4B



What is Machine Learning?

Please select your answers :

A) Memorizing something

B) Learning facts through observation and exploration

C) Generalizing a concept from experience

What type of learning to apply...

Depends on what type of feedback is available:

- **Supervised learning**: correct answers for each example
- **Unsupervised learning**: correct answers not given
- **Reinforcement learning**: occasional rewards



AlphaGo



Lee Sedol



Induction

Making predictions about the future based on the past. But is induction sound? Why believe that the future will look similar to the past?

If asked why we believe the sun will rise tomorrow, we shall naturally answer, “Because it has always risen every day.” We have a firm belief that it will rise in the future, because it has risen in the past.

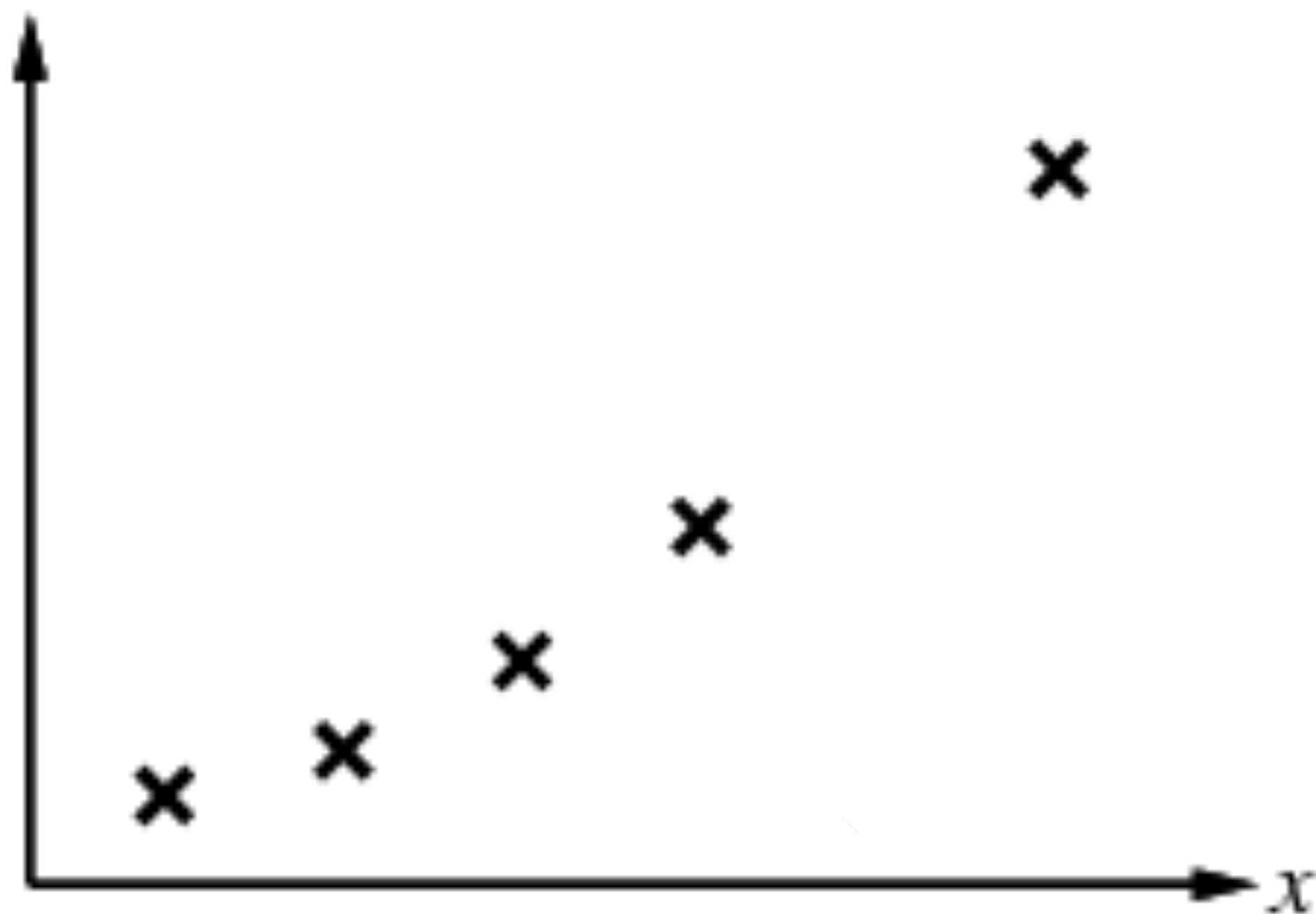
– [Bertrand Russell](#)

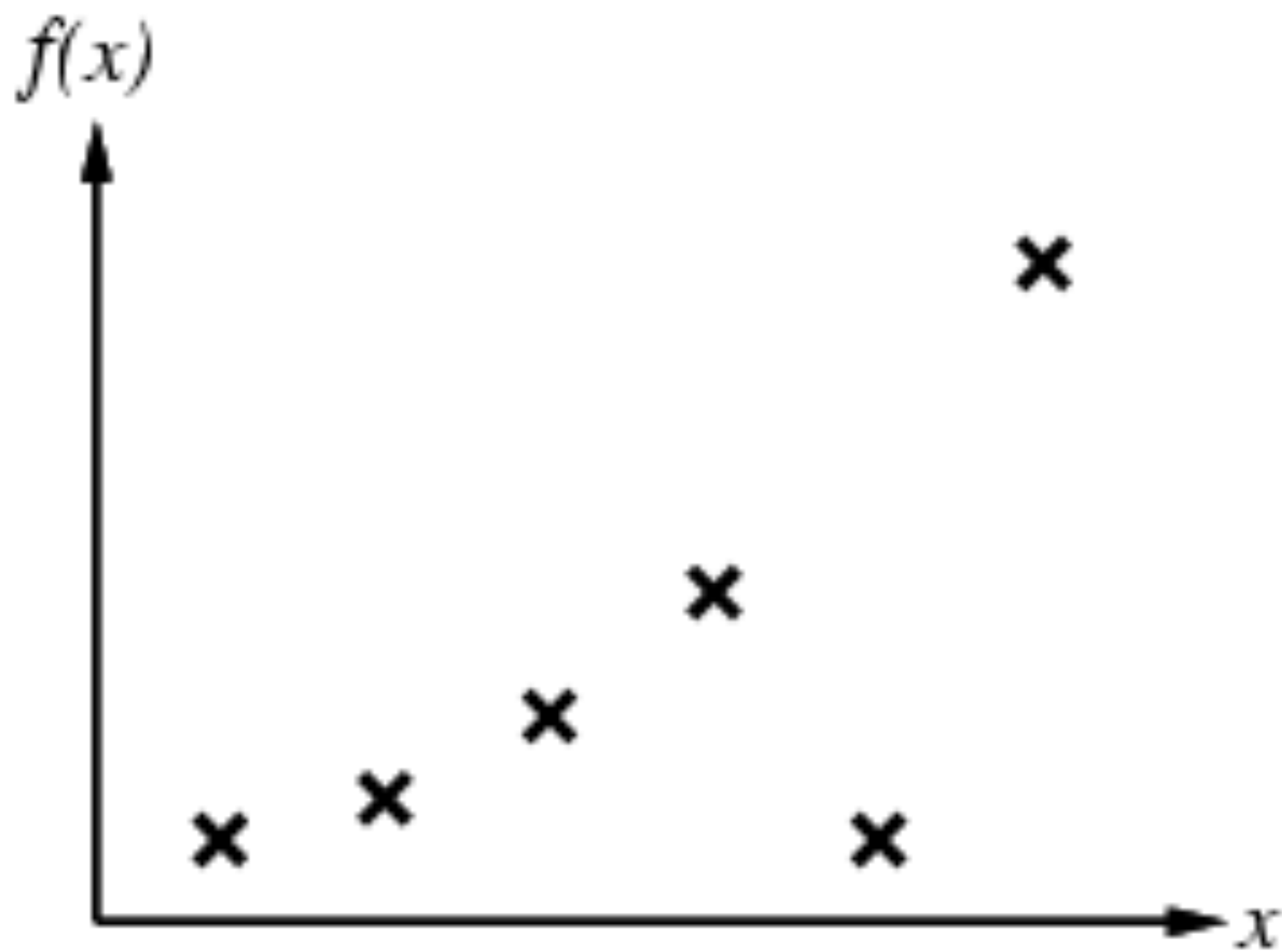
Inductive Learning

Learn a function from examples

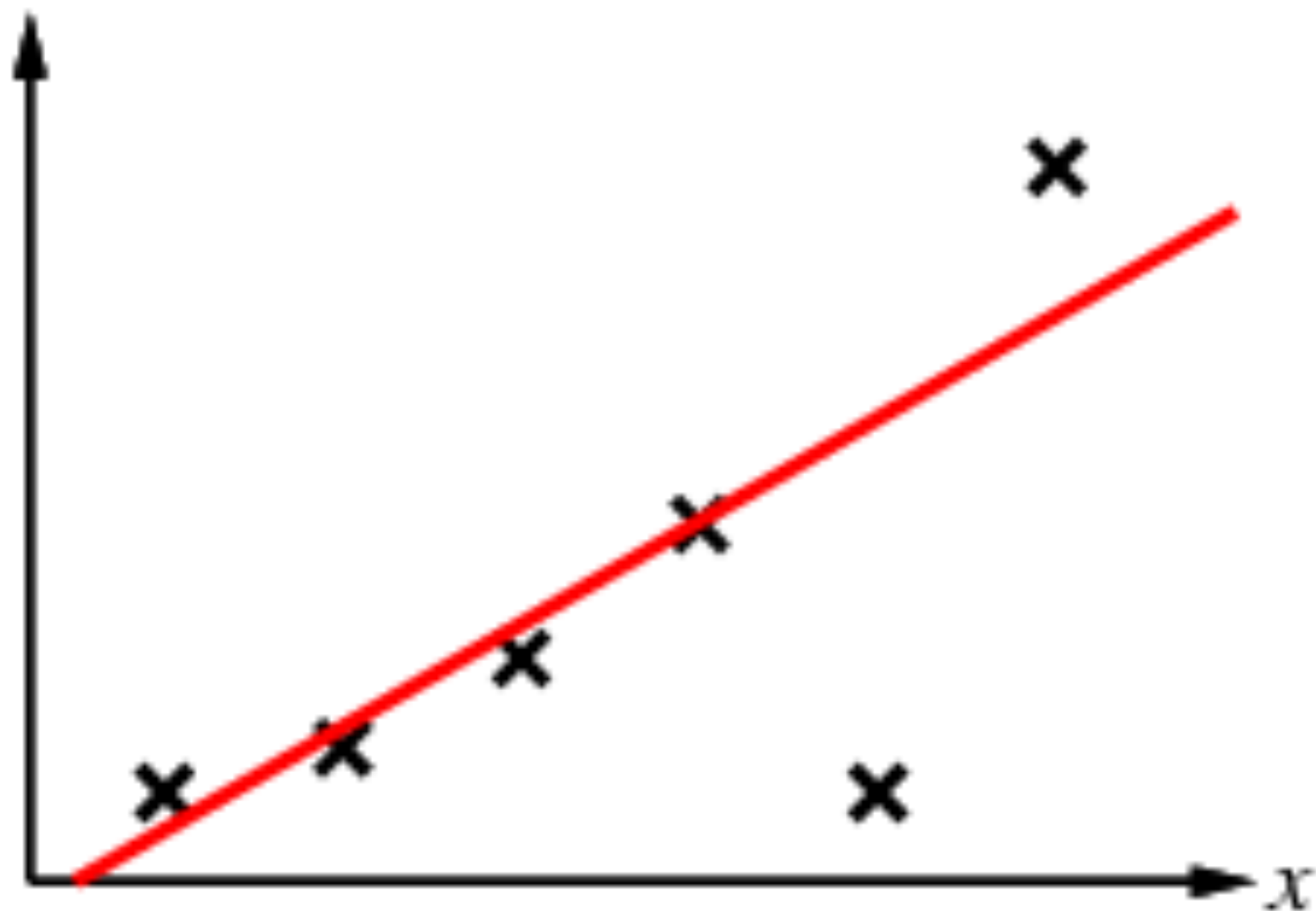
- f is the target function
- An example is a pair $\langle x, f(x) \rangle$
- Problem: find a hypothesis h
 - such that $h \approx f$, given a training set of examples

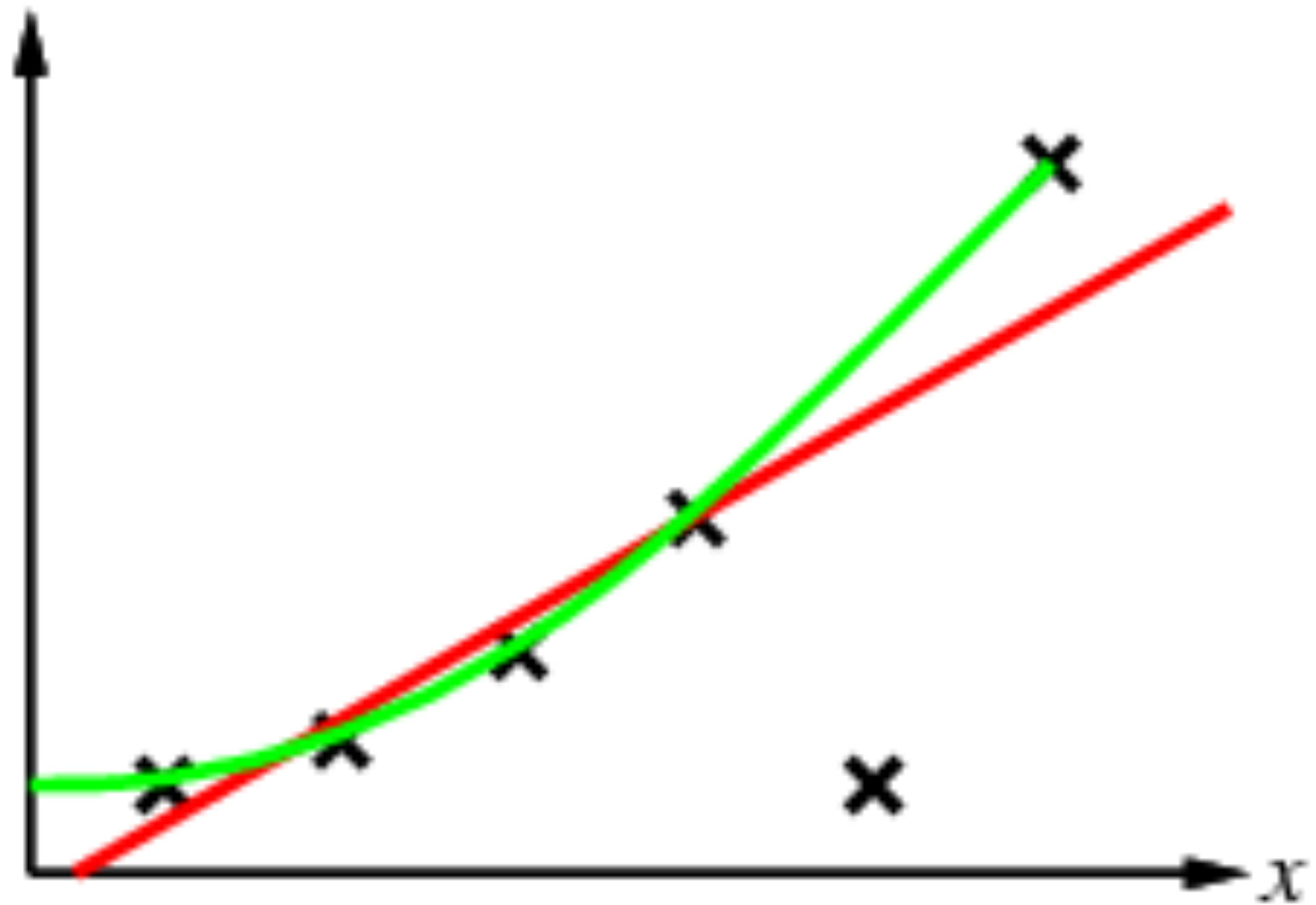
$f(x)$

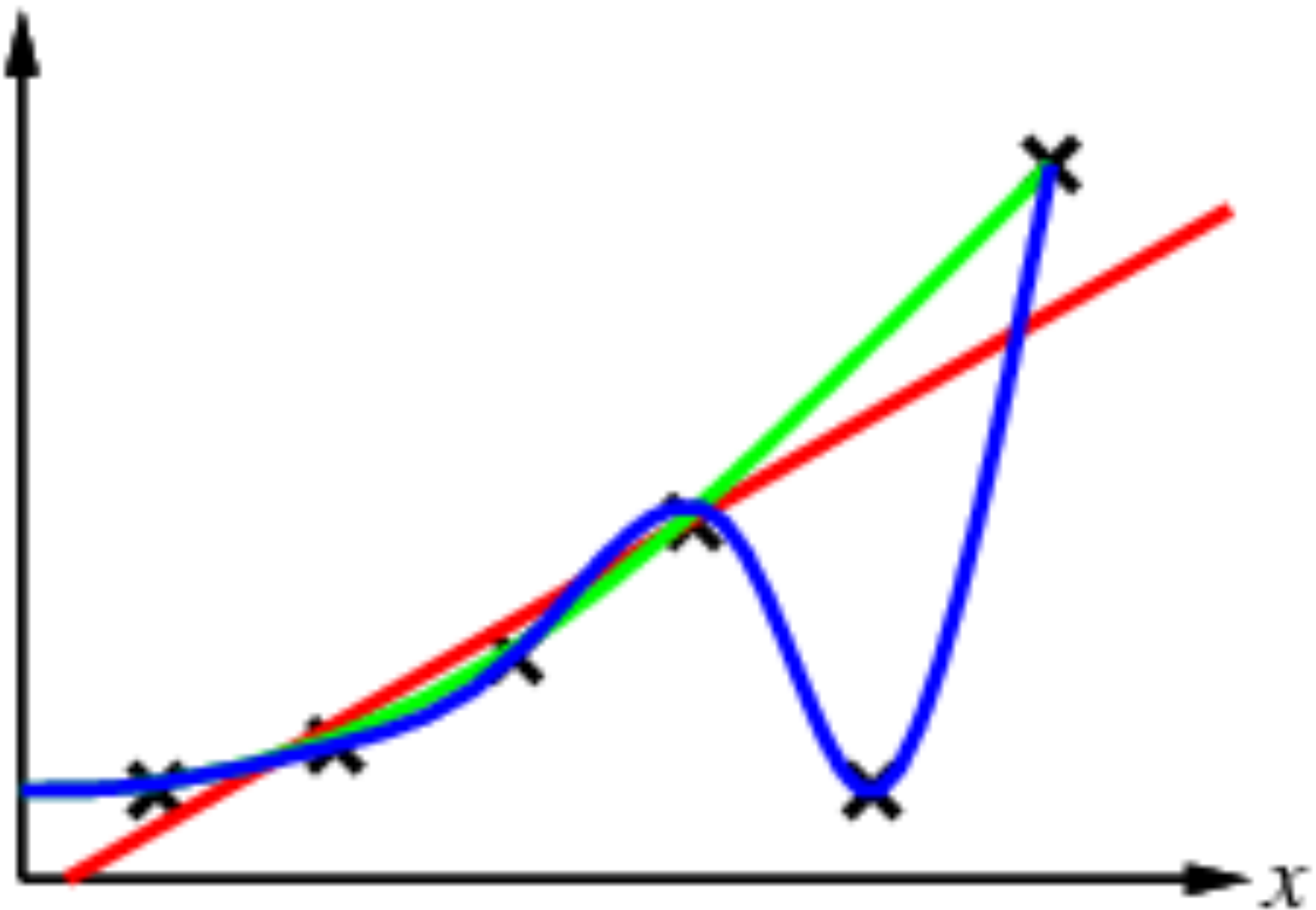


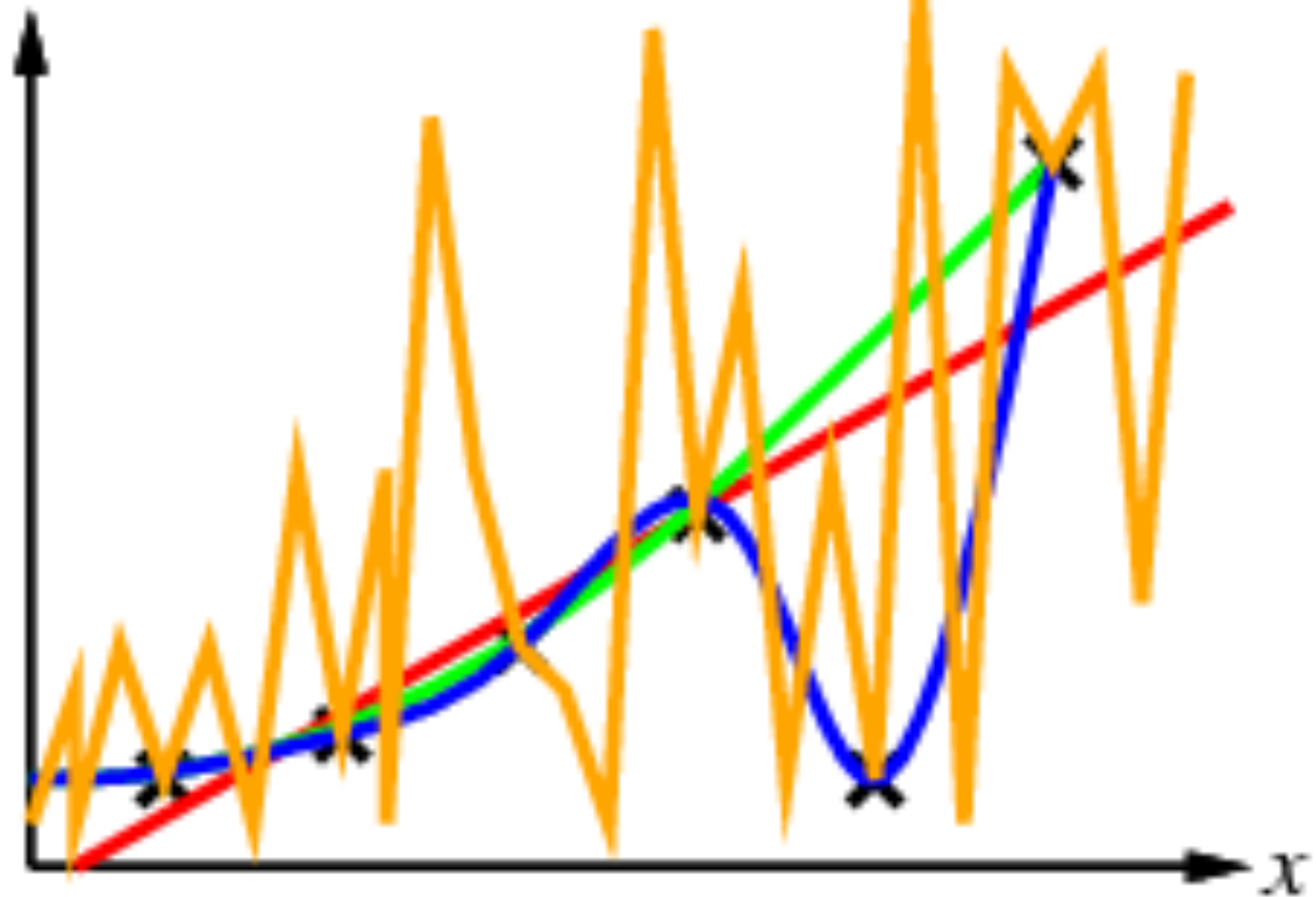


$f(x)$



$f(x)$ 

$f(x)$ 

$f(x)$ 

x is usually $\langle X \rangle$

That is, instead of learning from a single variable, we learn a function that is dependent on many inputs, or a vector.

Multiple, numeric inputs

Text, treated as words, then mapped to numbers

An application: Ad blocking

CNN.com International Edition | Netscape
MEMBER SERVICES MAKE CNN.com YOUR HOME PAGE

THE SOUTH BEACH DIET™ GET READY TO LOOK FABULOUS!
ONLINE
• 3 easy phases • 24h online support • Results, Results!

SEARCH The Web CNN.com Search Enhanced by: Google

Home Page Updated: 12:30 a.m. EST (05:30 GMT) March 15, 2004

Israeli helicopters fire at Gaza targets



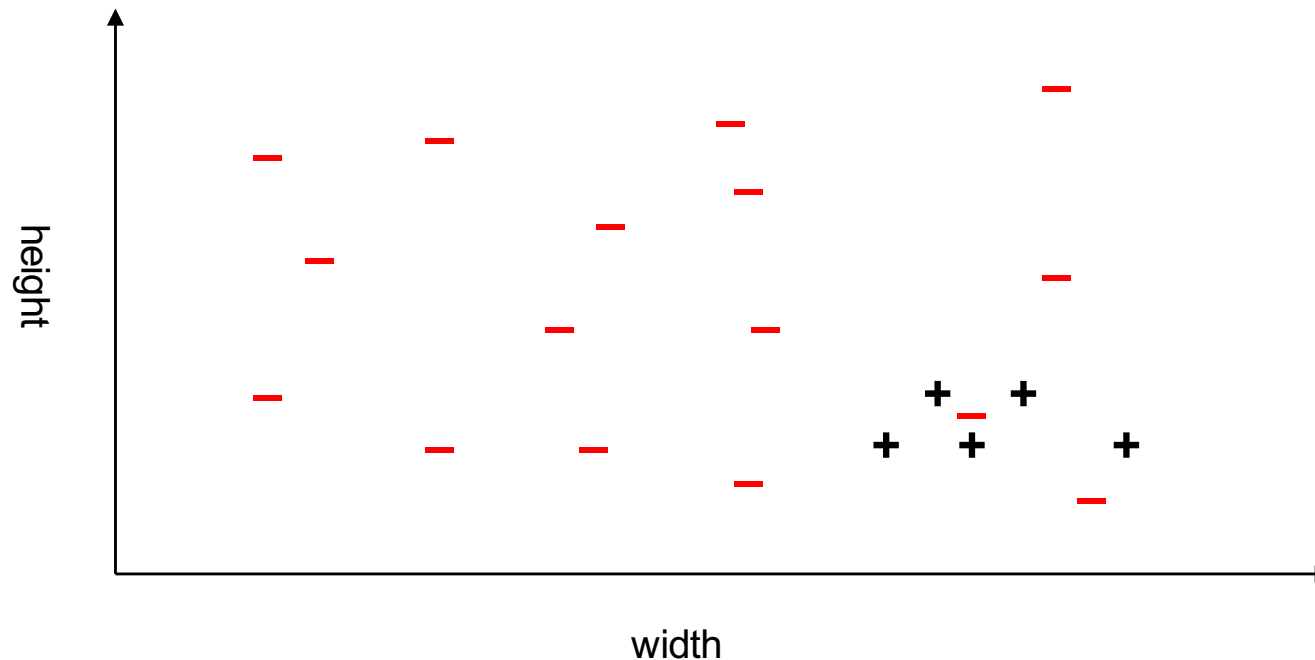
MORE TOP STORIES MOST POPULAR ▶

- [Terror-scarred Spain votes in Socialists](#) | [Video](#) | [Gallery](#)
- [Son: Murder suspect the 'best dad'](#) | [Video](#) | [Gallery](#)
- CNN/Money: [Microsoft facing European sanctions](#)
- [Ashcroft discharged from hospital](#)
- SI.com: [March Madness field unveiled](#) | [March Madness](#)
- [1794 silver dollar may be the country's first](#)
- ['Passion' stays on top at box office](#)
- **AMERICA VOTES 2004** ▶ [Complete election coverage](#)

CNN RADIO VIDEO MORE VIDEO ▶

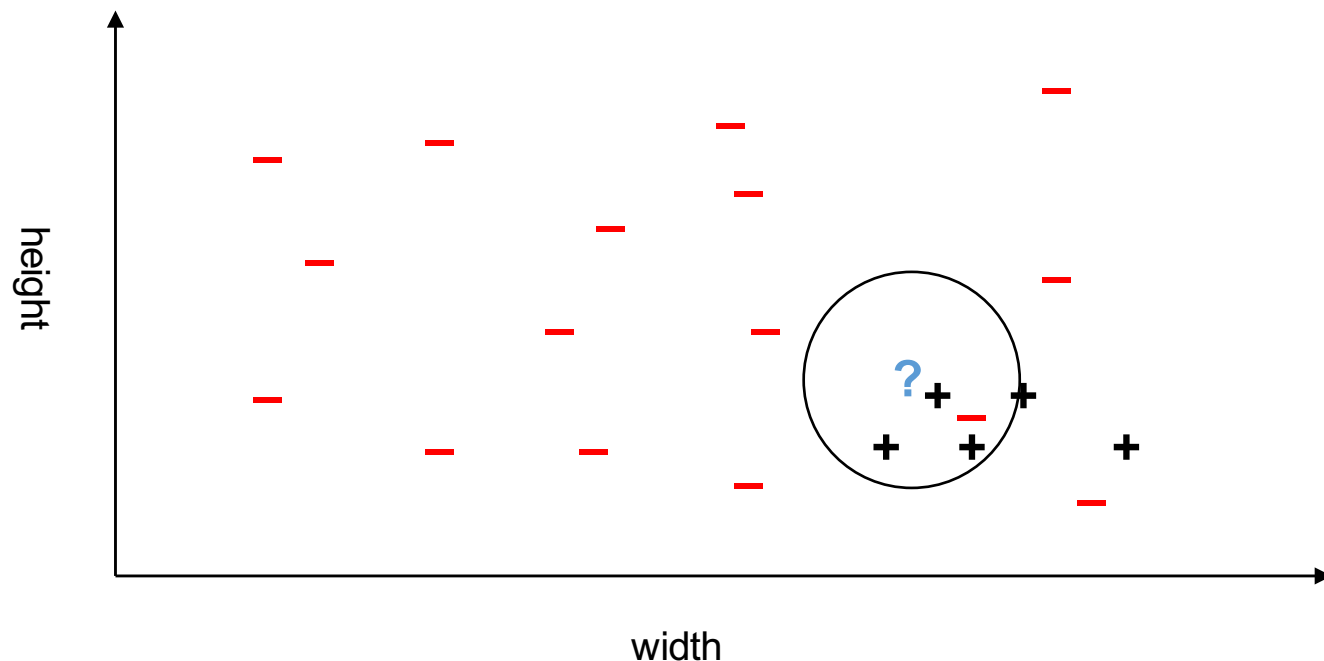
Learning Ad Blocking

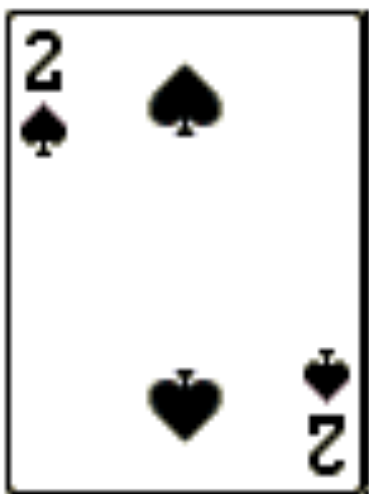
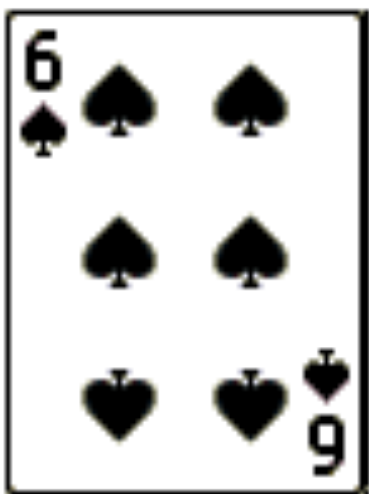
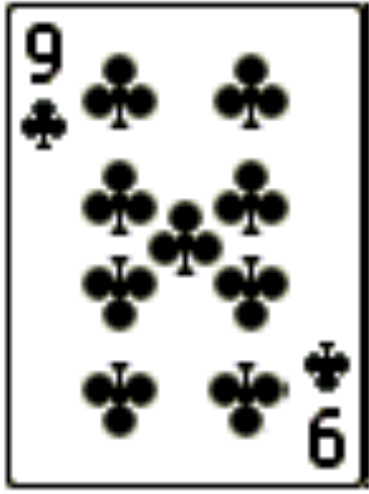
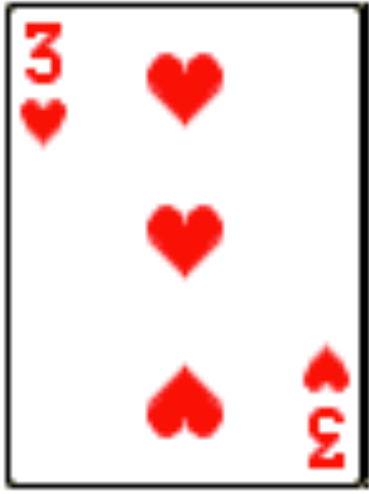
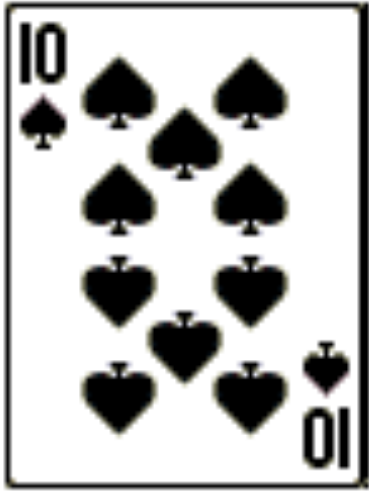
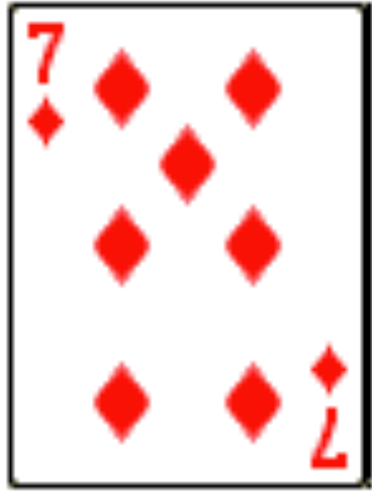
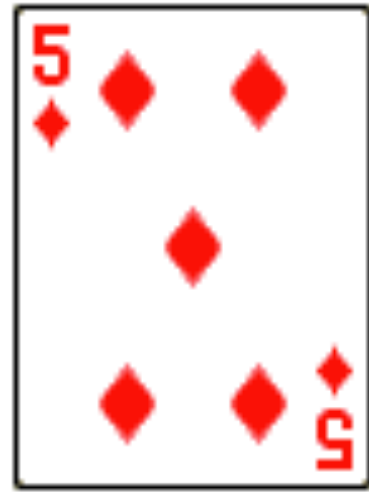
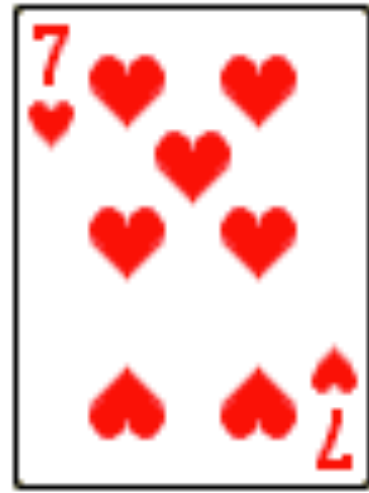
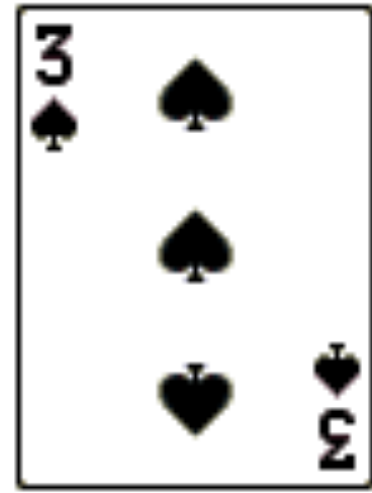
- Features $\langle X \rangle$: Width and height of image
- Binary Classification y : ad or not_ad?



(k) Nearest Neighbors

- Pick nearest neighbor, and associate the new example with its neighbors' class
- Generalized to k nearest neighbors





Twenty Questions

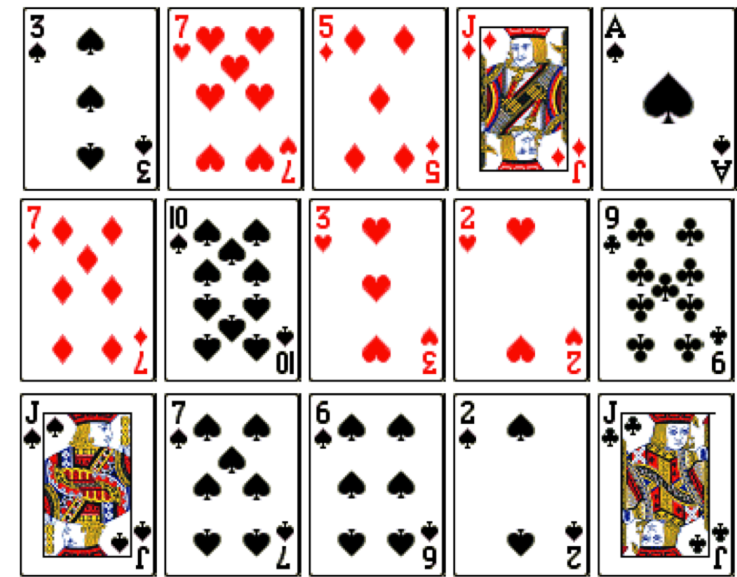
- Idea: pick a good attribute.

Good means it splits the training set into subsets that are (ideally) "all positive" or "all negative"

You try:

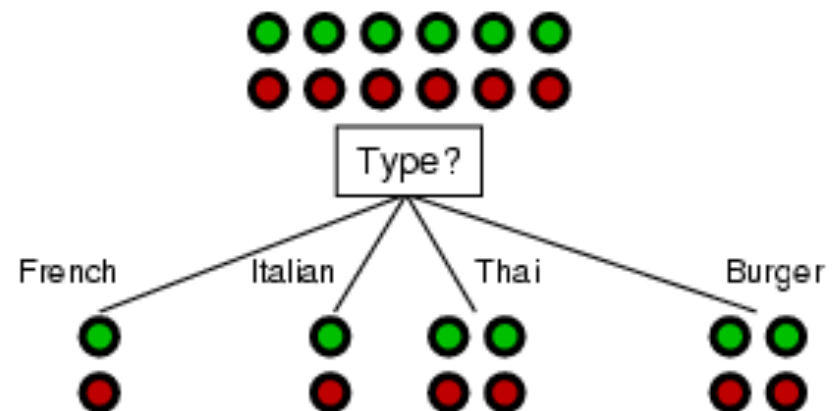
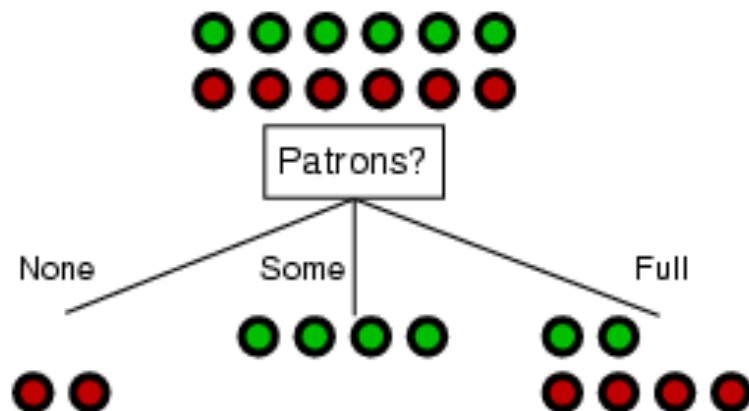
For the hidden function "Black Jacks", which question (attribute) is better?

- Are the correct cards red?
- Are the correct cards in the first row?



Decision Trees

- This is the basis for Decision Trees: one of the easiest to explain algorithms for learning
- Split a dataset by a variable and factor the task into a set of smaller problems
- The smaller problems stop being problems when the class is pure. E.g., for a “eating out” problem:



y is often Y , too

That is, instead of a binary output, many tasks may be better cast as a multiclass classification problem.

Some classifiers do this by making n binary classifiers, and choosing the class that has the highest confidence:

{CS, IS, InfoSec, CompBio, CompEng, BZA}

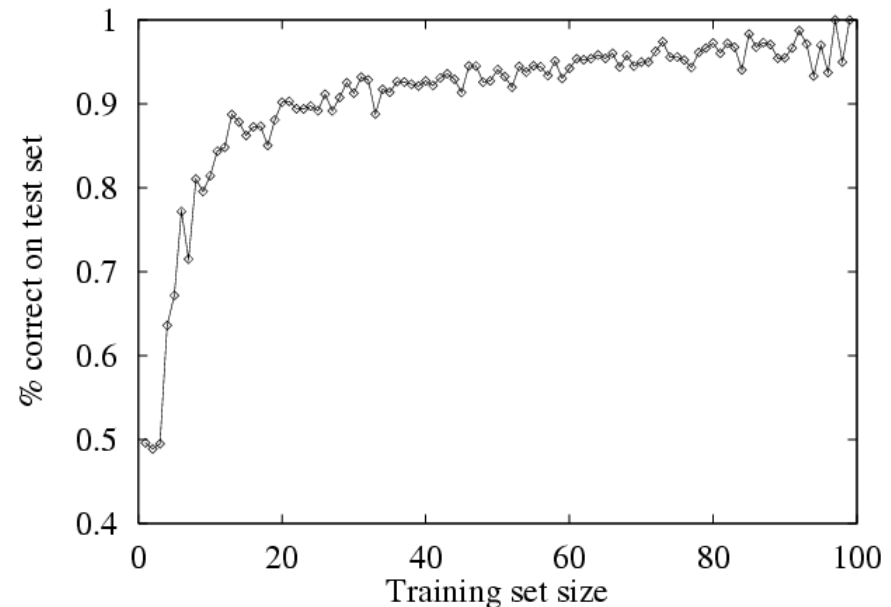
They can also be multi-label (“Select all that apply”):
{Monday, Tuesday, Wednesday, Thursday, Friday}

Performance measurement

How do we know that $h \approx f$?

Try h on a new **test set** of examples

Learning curve = % correct on test set as a function of training set size



Training and testing sets

Where does the test set come from?

1. Collect a large set of examples
2. Divide it into **training** and **testing data**
3. Train on training data, assess on testing
4. Repeat 1-3 for different splits of the set = k fold cross validation

