Introduction

Mathematics for Machine Learning

by Marina Barsky

Objectives: discuss

- What do we mean by Machine Learning (ML)
- What can we do with learned models?
- Demo the type of Math for ML
  - Linear Algebra → Data as vectors
  - Probability → Probabilistic classifiers
  - Statistics → Extracting models from data
  - Calculus → Adjusting model parameters

What do we mean by learning?

- Examples of learning:
  - Based on the previous experiences assign a label to a new object

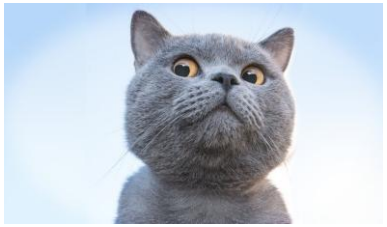
Label: **house**

?

What do we mean by learning?

- Examples of learning:
  - Based on the previous experiences assign a label to a new object

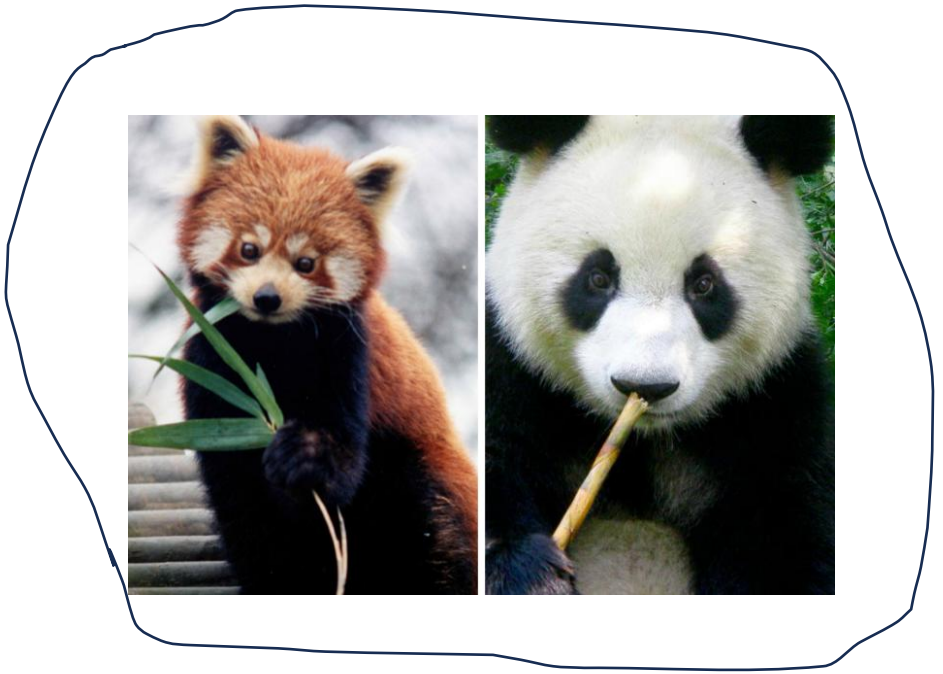




What do we mean by learning?

- Examples of learning:
  - Based on the previous experiences assign a label to a new object
  - Group similar things together into a single category

Cats?



Bears?

What do we mean by learning?

- Examples of learning:
  - Based on the previous experiences assign a label to a new object
  - Group similar things together into a single category
  - Identify repeated patterns

What is repeating in all these dogs?

What is Machine Learning?

- Machine learning occurs when machines **learn by themselves**, without explicit instructions

- How can a machine learn something new, if all the instructions are written by a human programmer?

ML algorithms extract models from data!

- ML algorithms learn from data – a previously recorded collection of examples of some phenomenon
- If data is non-random, it contains *patterns*
- Based on these patterns, ML algorithms discover a *generalized model of data*
- That allows to make *assumptions* about other data that it might see in the future

We learn from the past – to predict the future

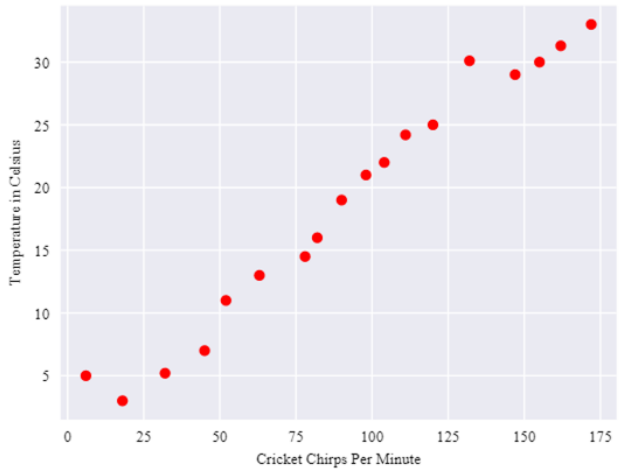
Descriptive analytics

Predictive analytics
Machine Learning
Data mining

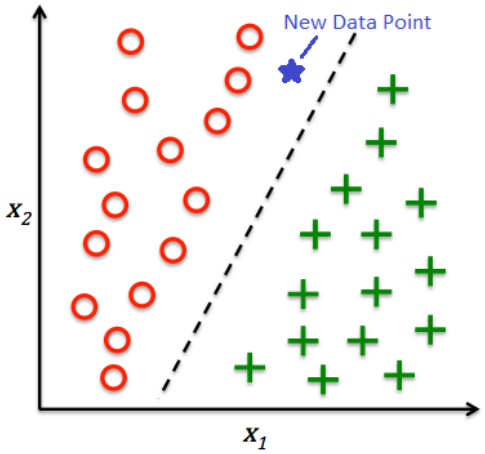
now

What can we do with ML

Supervised learning

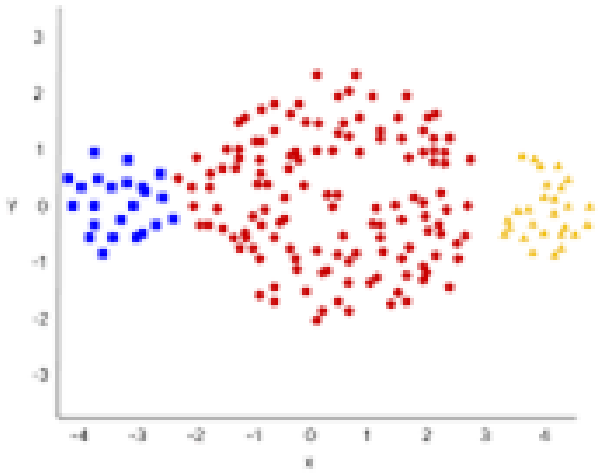


Prediction



Classification

Unsupervised learning



Clustering

TransactionId	Items
1	{A,C,D}
2	{B,C,D}
3	{A,B,C,D}
4	{B,D}
5	{A,B,C,D}

Associations

Example: Learning to classify

My neighbor dataset

class label

Temp	Precip	Day	Shop	Clothes	
25	None	Sat	No	Casual	**Walk**
-5	Snow	Mon	Yes	Casual	**Drive**
15	Snow	Mon	Yes	Casual	**Walk**

(Adopted from Leslie Kaelbling's example in the MIT courseware)

Classify:

class label

Temp	Precip	Day	Shop	Clothes	
25	None	Sat	No	Casual	**Walk**
-5	Snow	Mon	Yes	Casual	**Drive**
15	Snow	Mon	Yes	Casual	**Walk**
-5	Snow	Mon	Yes	Casual	**?**

Classify: memory

class label

Temp	Precip	Day	Shop	Clothes	
25	None	Sat	No	Casual	**Walk**
-5	Snow	Mon	Yes	Casual	**Drive**
15	Snow	Mon	Yes	Casual	**Walk**
-5	Snow	Mon	Yes	Casual	**Drive**

Classification problem: noise

Temp	**Precip**	**Day**	**Clothes**	
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Drive**
25	None	Sat	Casual	**Drive**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**?**

Classification: averaging

Temp	**Precip**	**Day**	**Clothes**	
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Drive**
25	None	Sat	Casual	**Drive**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Walk**

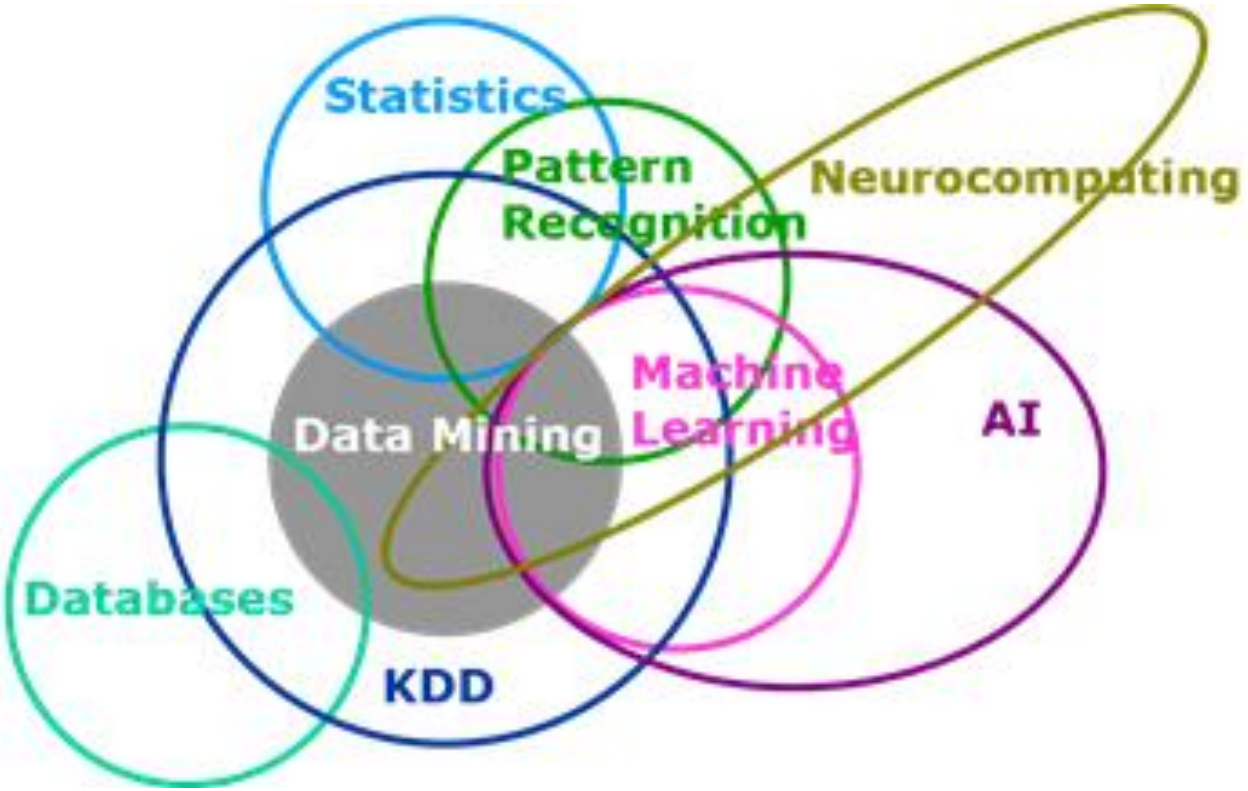
Classification: generalization

Temp	**Precip**	**Day**	**Clothes**	
22	None	Fri	Casual	**Walk**
3	None	Sun	Casual	**Walk**
10	Rain	Wed	Casual	**Walk**
30	None	Mon	Casual	**Drive**
20	None	Sat	Formal	**Drive**
25	None	Sat	Casual	**Drive**
-5	Snow	Mon	Casual	**Drive**
27	None	Tue	Casual	**Drive**
24	Rain	Mon	Casual	**?**

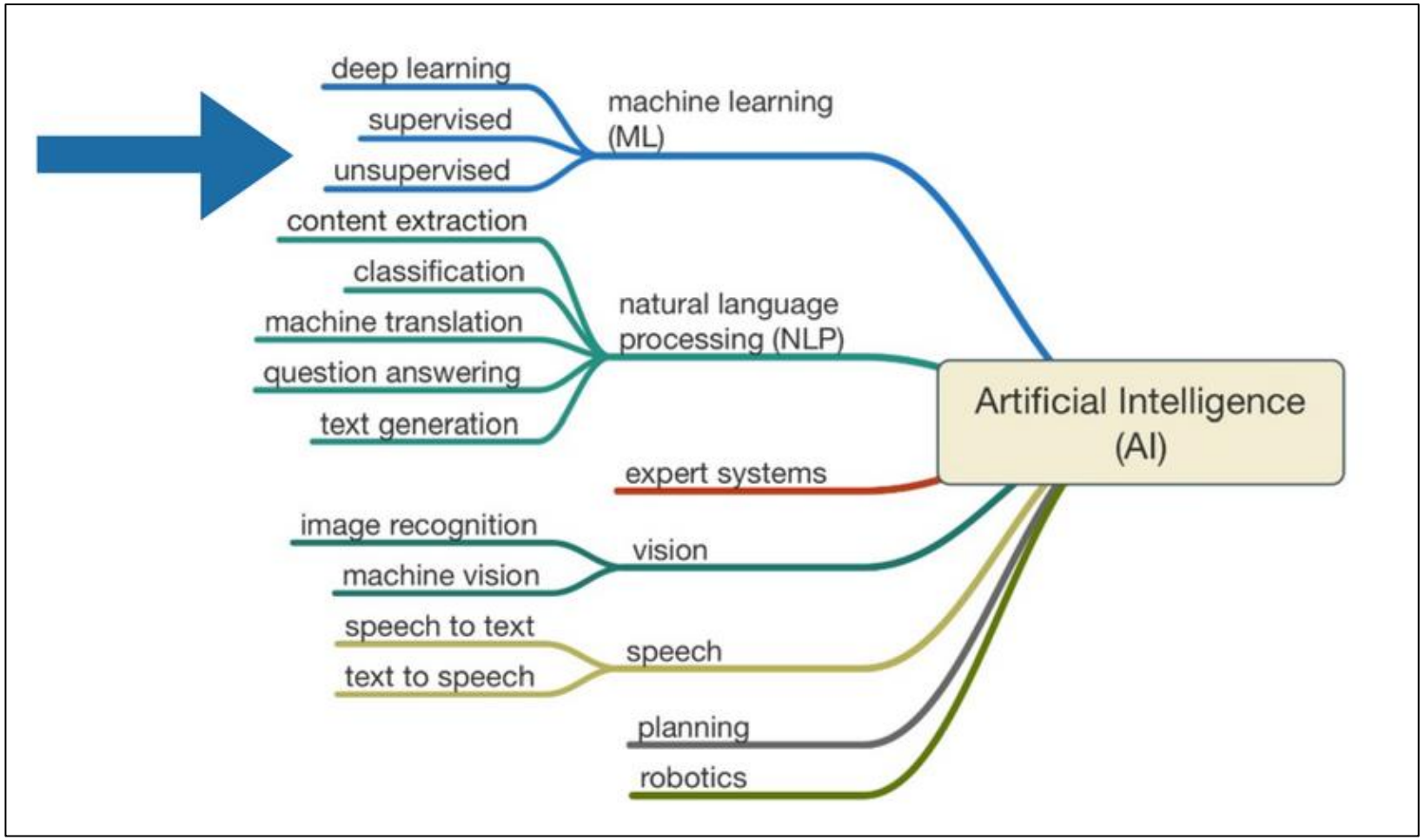
Classification model built on:

- memory
- averaging
- generalization

Statistics, Data Mining, ML, AI: What is the difference?



Machine Learning is a subfield of AI



Different subfields overlap

- **Statistics**: statistically sound **sampling** techniques
- **Machine Learning**: incorporates statistics plus linear algebra plus optimizations plus reinforcement learning plus …
- **Data Mining**: incorporates all the all the Machine Learning algorithms plus their efficient implementation for very large datasets (Big Data, parallel processing)

There is also:

- **Data Science?**

a new way of learning about the world – from data

Evolution of Science

- Empirical Science – collect and systematize facts
- Theoretical Science – formulate theories and empirically test them
- Computational Science –run automatic proofs, simulations
- **e-Science** (**Data Science**) – collect data without clear goal - and test theories, find patterns **in the data itself**

Data science

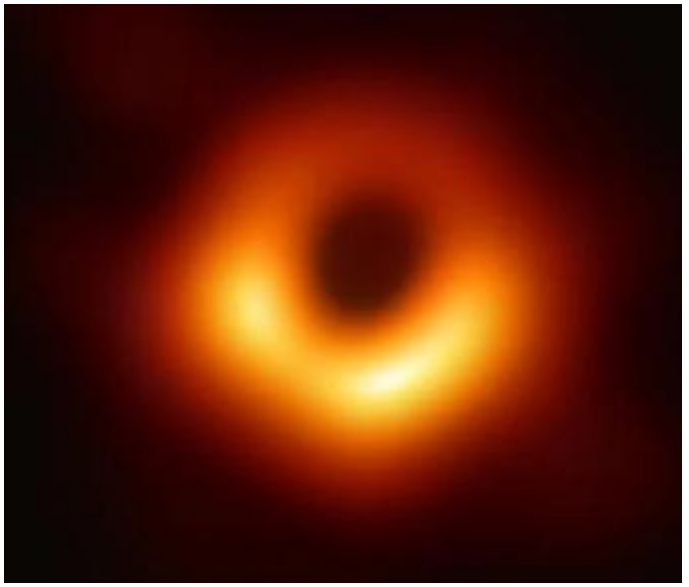
- Traditionally: “Query the world”

Data acquisition for a specific hypotheses

- Data science: “Download the world”

Data acquired en masse in support of future hypotheses

Query the world through data



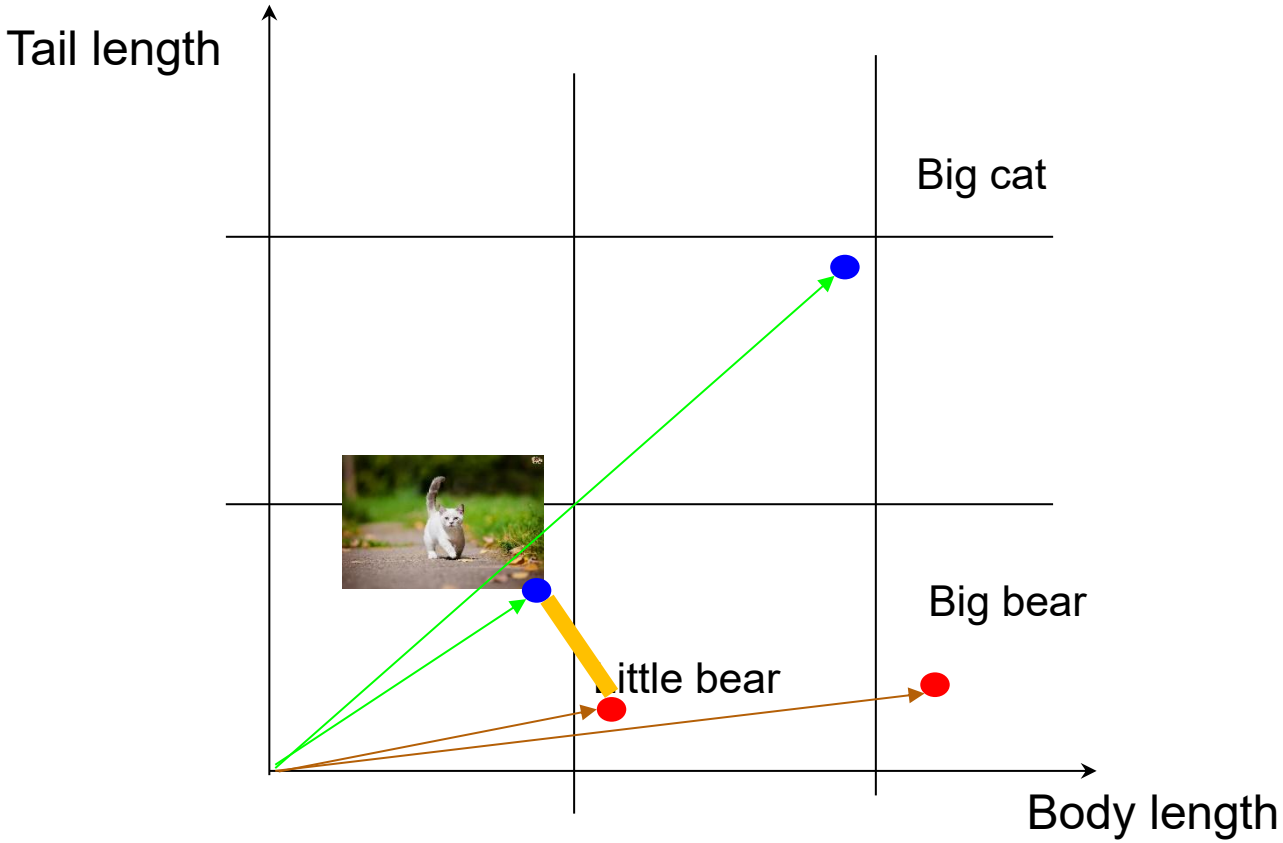
ML techniques are used to process the enormous amounts of data from the global network of radio telescopes to create the first image of black holes, including the one at the center of the Milky Way.

ML algorithms: what kind of Math?

- Linear algebra
  - We think of data as points in multi-dimensional space: need to understand *vectors* and *matrices*
- Probability
  - All ML predictions are based on *probabilistic reasoning*
- Statistics
  - We generate *statistical models* of data
- Calculus
  - We use derivatives to learn model parameters

Example: vectors

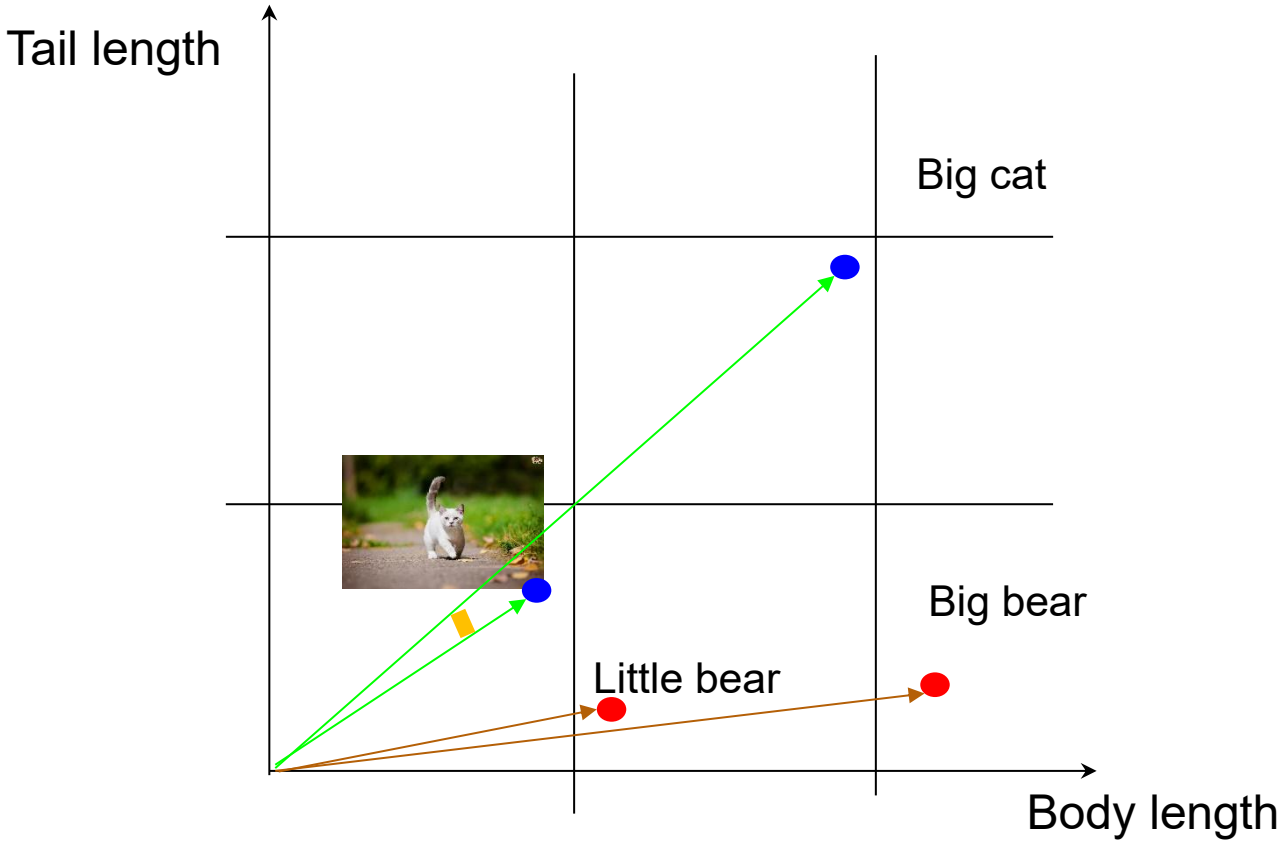
Cat or Bear?



Consider Euclidean distance

Example: vectors

Cat or Bear?



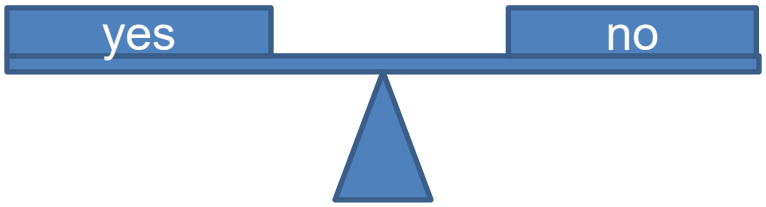
Consider Angle between vectors

Example: Probabilistic reasoning

I believe that John will be at the party

In the absence of facts

John will be at the party



What are the odds?

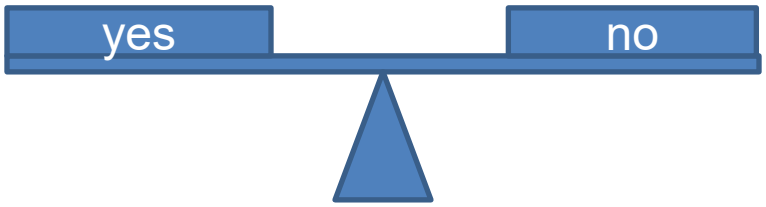
Probabilistic reasoning

I believe that John will be at the party

Invalid (illogical) reasoning

I do not like John

John will be at the party



What are the odds?

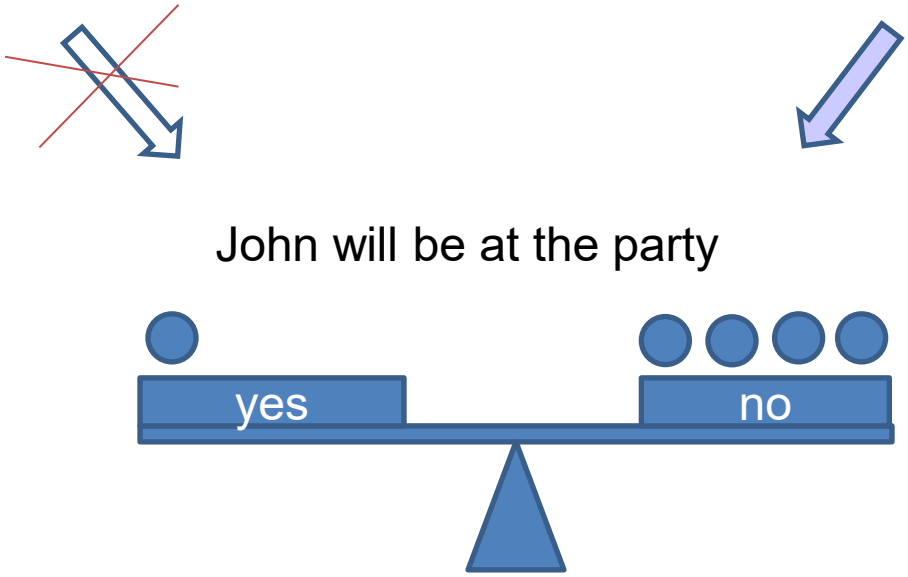
Probabilistic reasoning

I believe that John will be at the party

Probabilistic reasoning: valid fact (evidence)

I do not like John

John is very shy



What are the odds?

Probabilistic reasoning

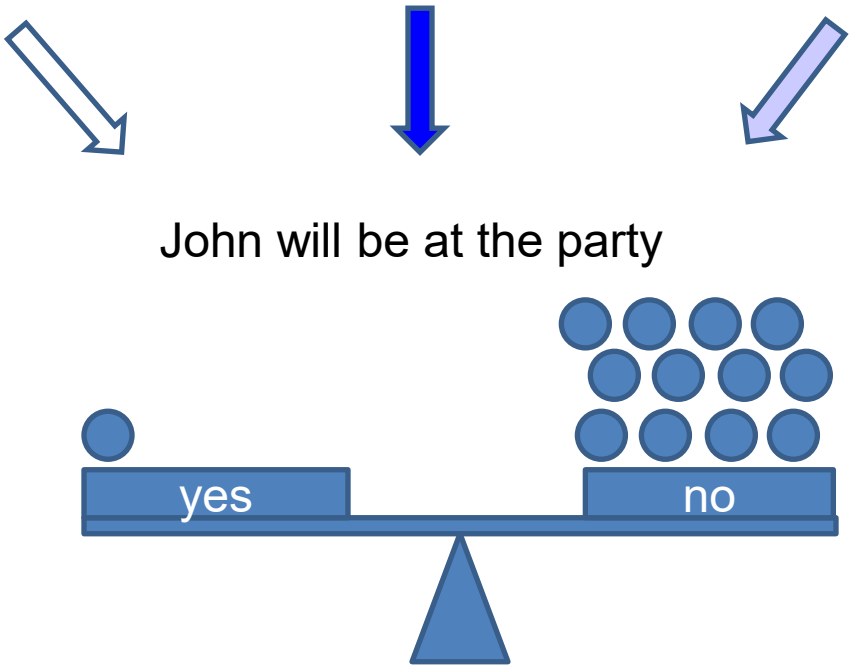
I believe that John will be at the party

More facts – update your beliefs

I do not like John

John is in Beijing

John is very shy



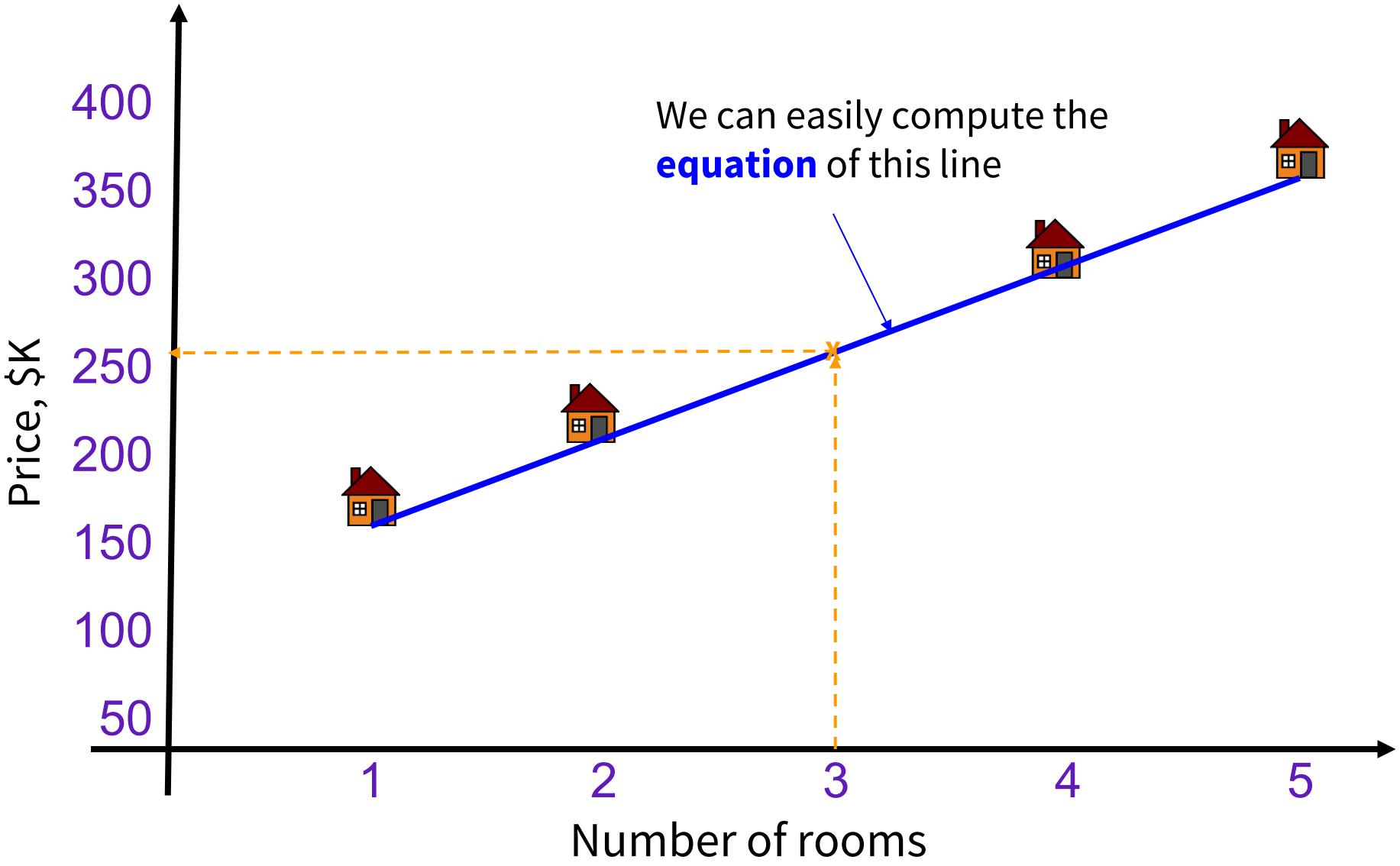
What are the odds?

Never tell me the odds!

- The Scene: In The Empire Strikes Back, Han Solo flies through an asteroid field while being pursued by TIE fighters.
- C-3PO attempts to tell Han the "3,720 : 1" odds against successfully navigating the field.

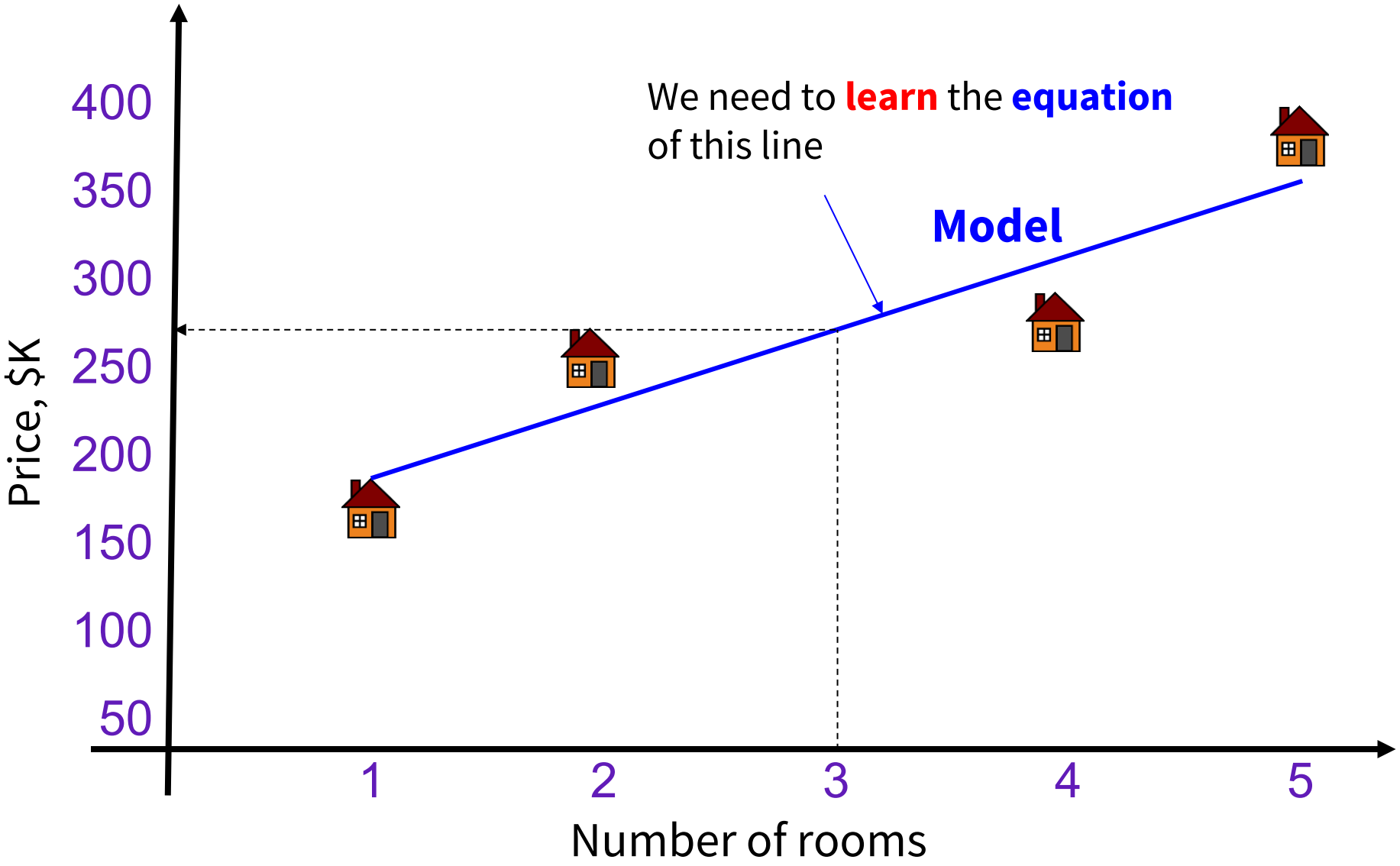
Model of home prices

Price of 3-bedroom home: ideal world



Model of home prices

Real world: line that fits the data points best

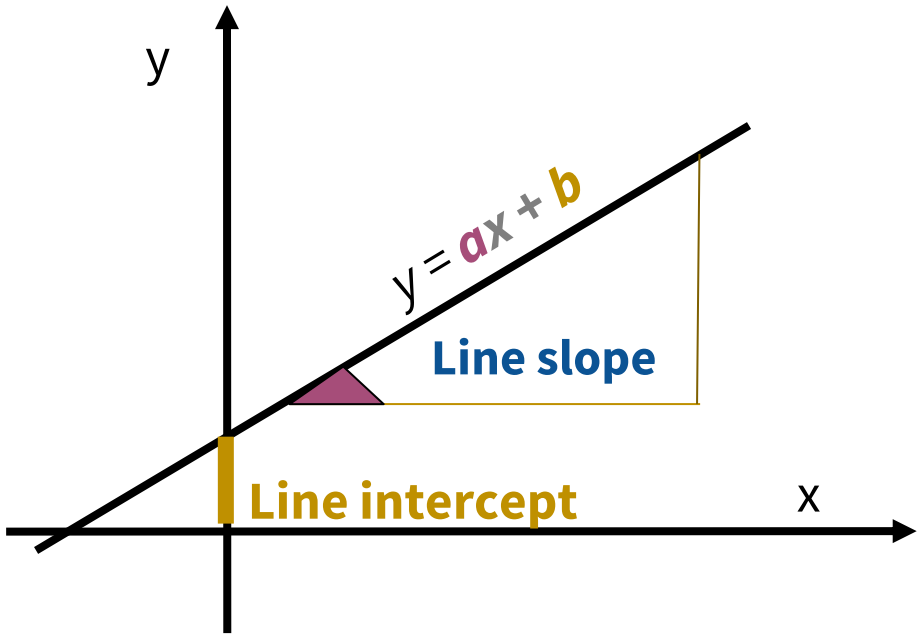


Model of home prices
Simple Linear Regression

We have a set of data points

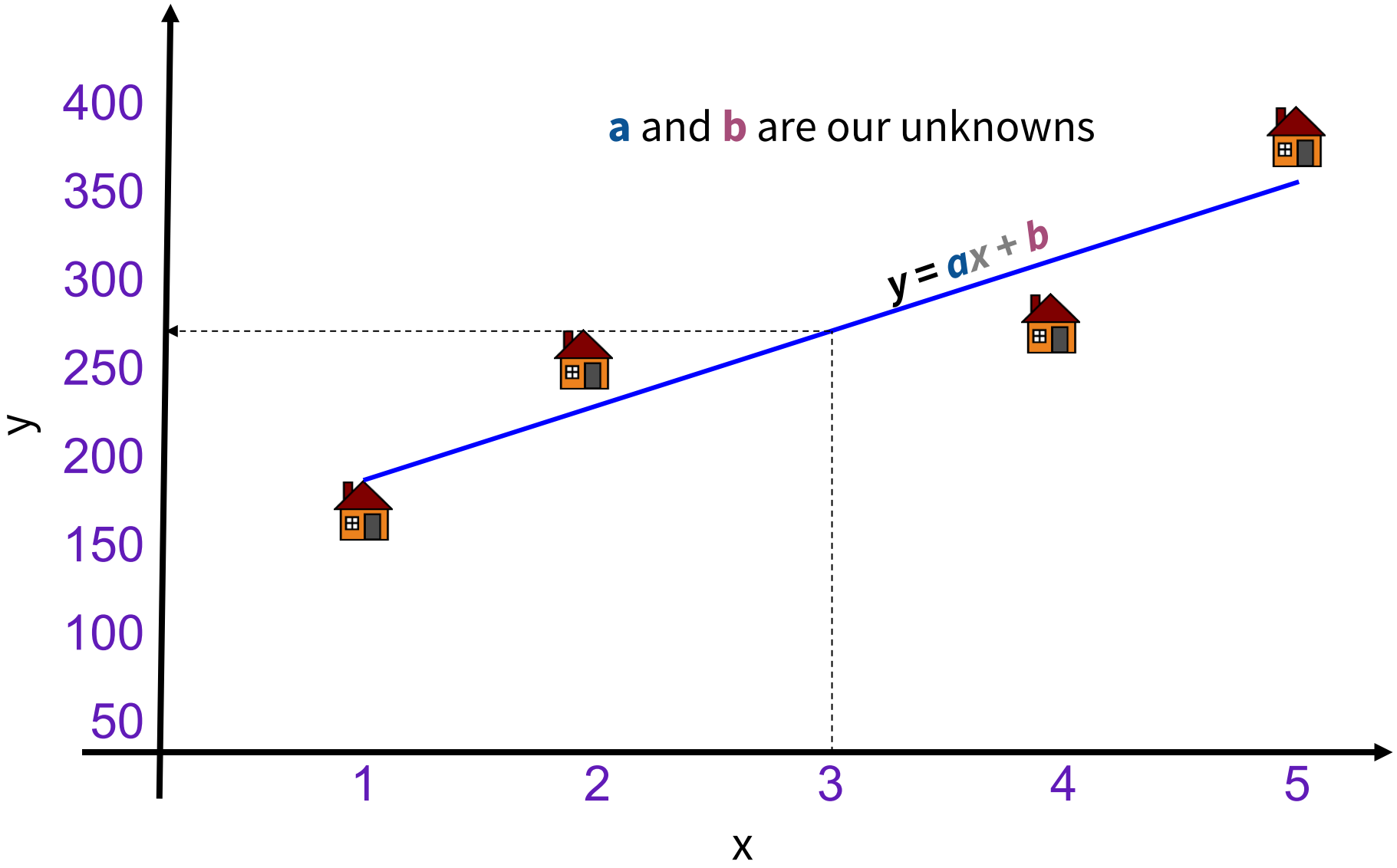
$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

Learning task:
Learn the linear function f(x) =$\boldsymbol{a}x$ + $\boldsymbol{b}$
which best describes linear relationship
between x and y



Model parameters

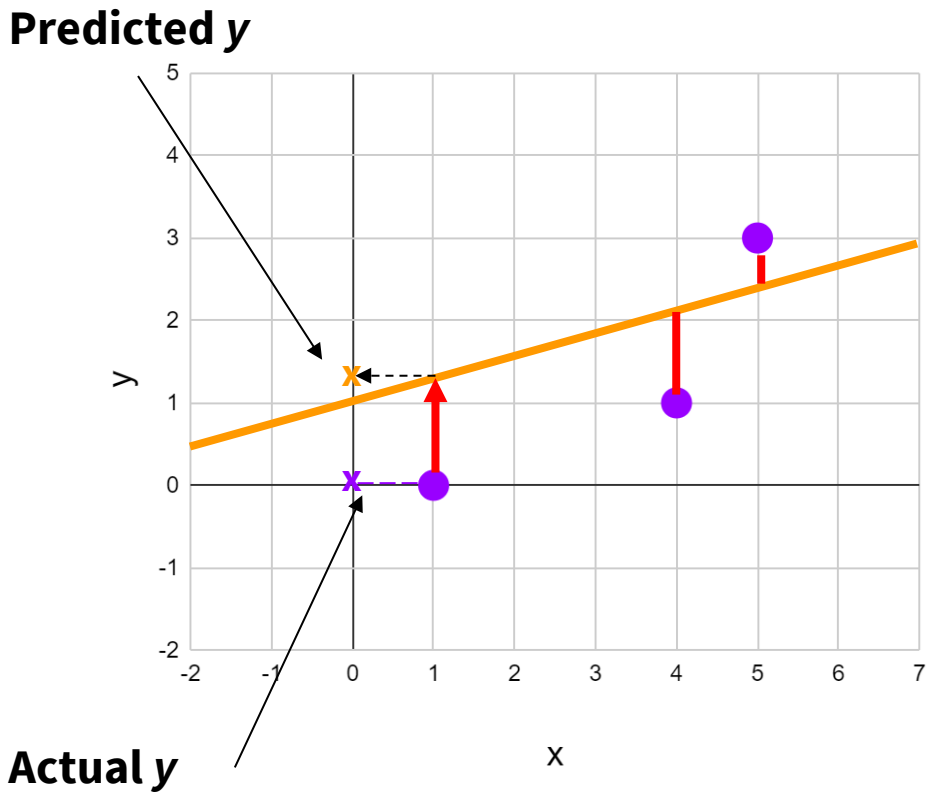
Line closest to a set of points



Model parameters
Fitting the line to data

We start with an arbitrary line.

The difference between the predicted and actual value is an error of the prediction



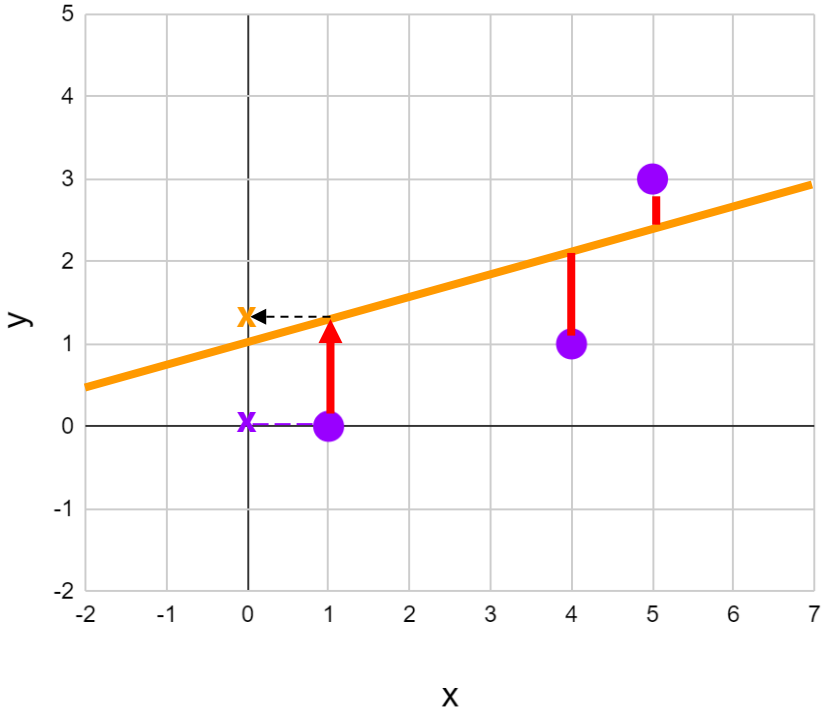
Data set: {[1,0], [4,1], [5,3]}

Model parameters
Fitting the line to data

We start with an arbitrary line.

The difference between the predicted and actual value is an error of the prediction

We formulate the error function as a function of two variables *a* and *b* and we try to minimize this error function



Model parameters

Using calculus

The error is a function of 2 variables: *a* and *b*

The goal is to find the combination of *a* and *b* that minimize the overall error.

This requires us to find the extrema of function E using partial derivatives of the error function:

$$\frac{\partial E}{\partial a} = 0, \frac{\partial E}{\partial b} = 0$$