A Short Tour of Machine Learning

Michael Mathioudakis

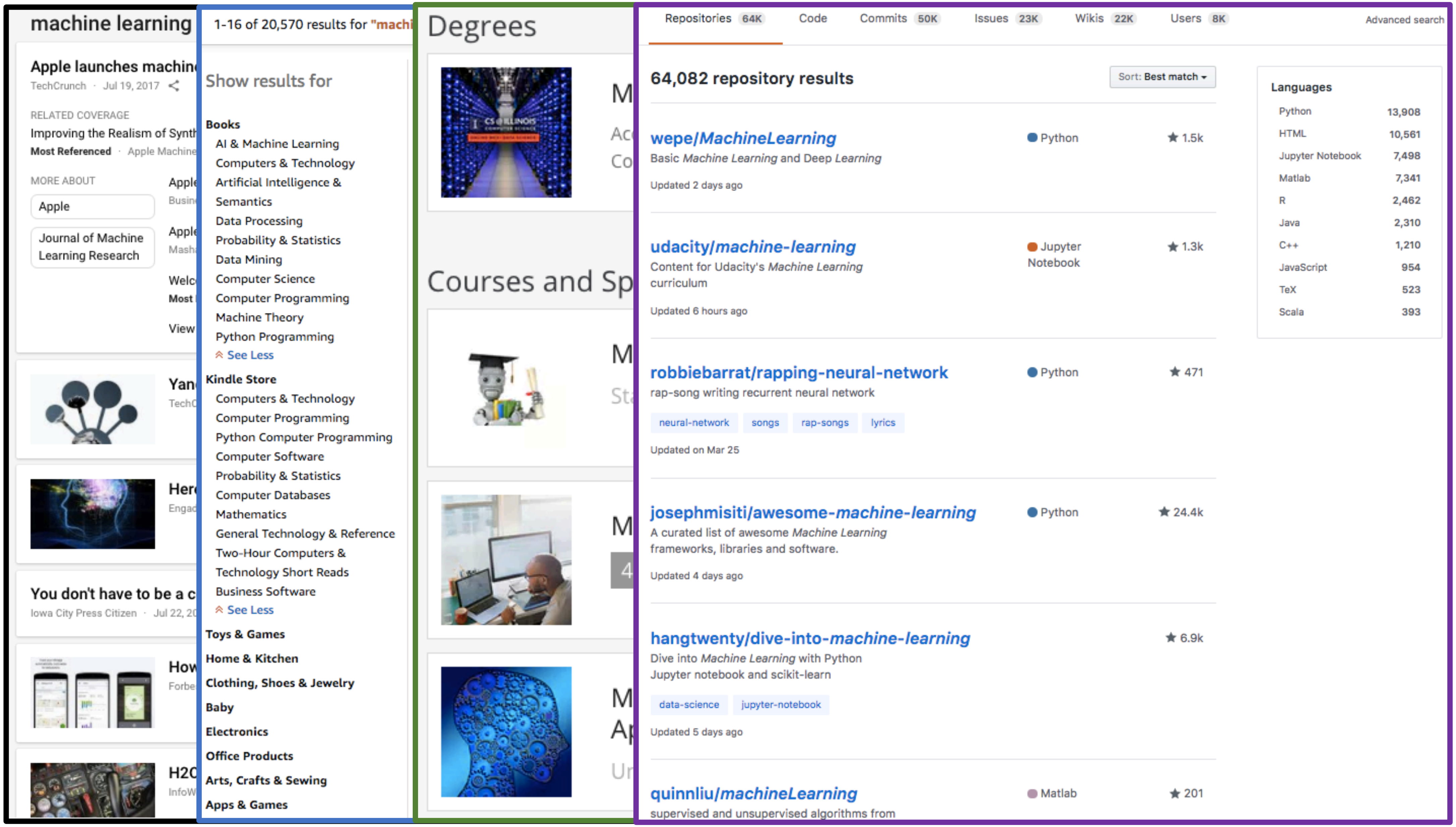
Covve, Athens, 2017-08-02

:HELVIA

today

- part 1: machine-learning algorithms
    a. basic concepts and the ML pipeline
    b. algorithms
- part 2: platforms and software
- part 3: hands-on session

part 1: machine learning algorithms


machine learning
Apple launches machin
TechCrunch · Jul 19, 2017
RELATED COVERAGE
Improving the Realism of Synth
Most Referenced · Apple Machine
MORE ABOUT
Apple
Journal of Machine Learning Research
You don't have to be a c
Iowa City Press Citizen · Jul 22, 20
1-16 of 20,570 results for "machi
Show results for
Books
AI & Machine Learning
Computers & Technology
Artificial Intelligence & Semantics
Data Processing
Probability & Statistics
Data Mining
Computer Science
Computer Programming
Machine Theory
Python Programming
See Less
Kindle Store
Computers & Technology
Computer Programming
Python Computer Programming
Computer Software
Probability & Statistics
Computer Databases
Mathematics
General Technology & Reference
Two-Hour Computers & Technology Short Reads
Business Software
See Less
Toys & Games
Home & Kitchen
Clothing, Shoes & Jewelry
Baby
Electronics
Office Products
Arts, Crafts & Sewing
Apps & Games
Degrees
Courses and Sp
Repositories 64K
Code
Commits 50K
Issues 23K
Wikis 22K
Users 8K
Advanced search
64,082 repository results
Sort: Best match
wepe/MachineLearning
Basic Machine Learning and Deep Learning
Python ★ 1.5k
Updated 2 days ago
udacity/machine-learning
Content for Udacity's Machine Learning curriculum
Jupyter Notebook ★ 1.3k
Updated 6 hours ago
robbiebarrat/rapping-neural-network
rap-song writing recurrent neural network
Python ★ 471
neural-network songs rap-songs lyrics
Updated on Mar 25
josephmisiti/awesome-machine-learning
A curated list of awesome Machine Learning frameworks, libraries and software.
Python ★ 24.4k
Updated 4 days ago
hangtwenty/dive-into-machine-learning
Dive into Machine Learning with Python Jupyter notebook and scikit-learn
★ 6.9k
data-science jupyter-notebook
Updated 5 days ago
quinnliu/machineLearning
supervised and unsupervised algorithms from
Matlab ★ 201
Languages
Python 13,908
HTML 10,561
Jupyter Notebook 7,498
Matlab 7,341
R 2,462
Java 2,310
C++ 1,210
JavaScript 954
TeX 523
Scala 393

what is machine learning?

‘learning’

<u>what</u> do we *learn*?

a description of the data

a ‘model’ that tells us how the data are distributed

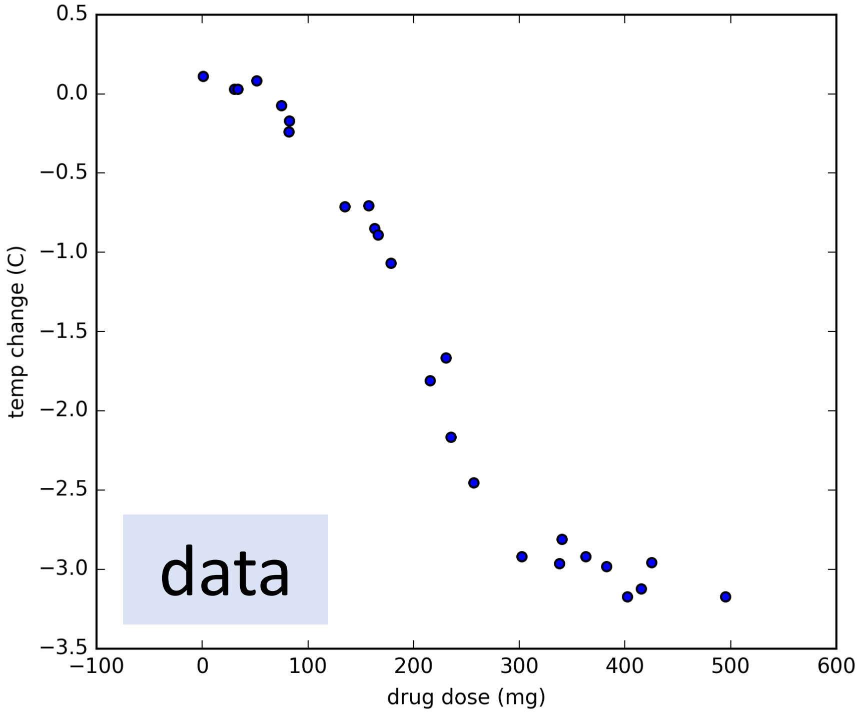
<u>why</u>?

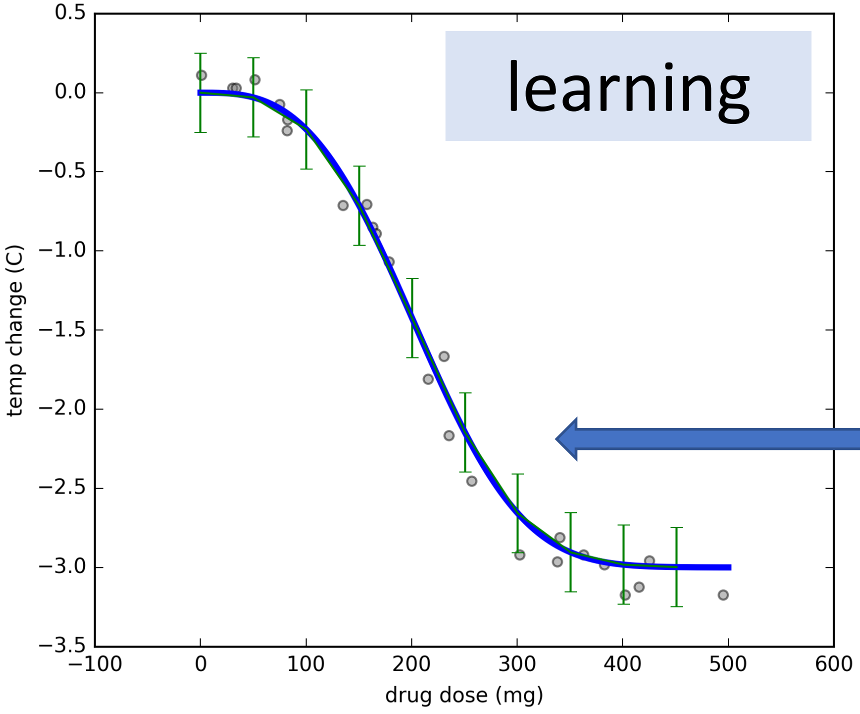
to make *predictions* and *decisions*

example

the patient's temperature has just exceeded 40C

we supply a drug and observe their temperature change after 2 hours





model

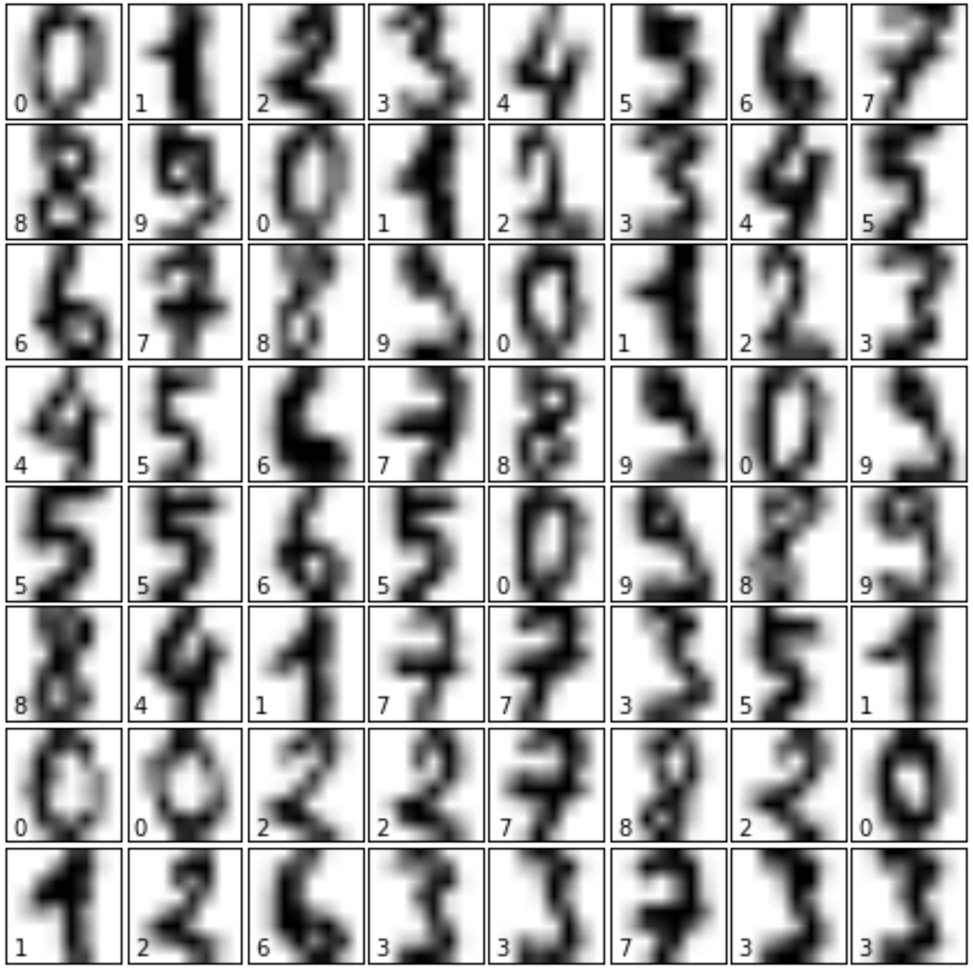
ok, we 'learned' - then what?

predict what happens to temperature if we supply 0.20mg?

decide minimum dose to achieve at least 2.5C temp drop?

can do with the model without the data

example

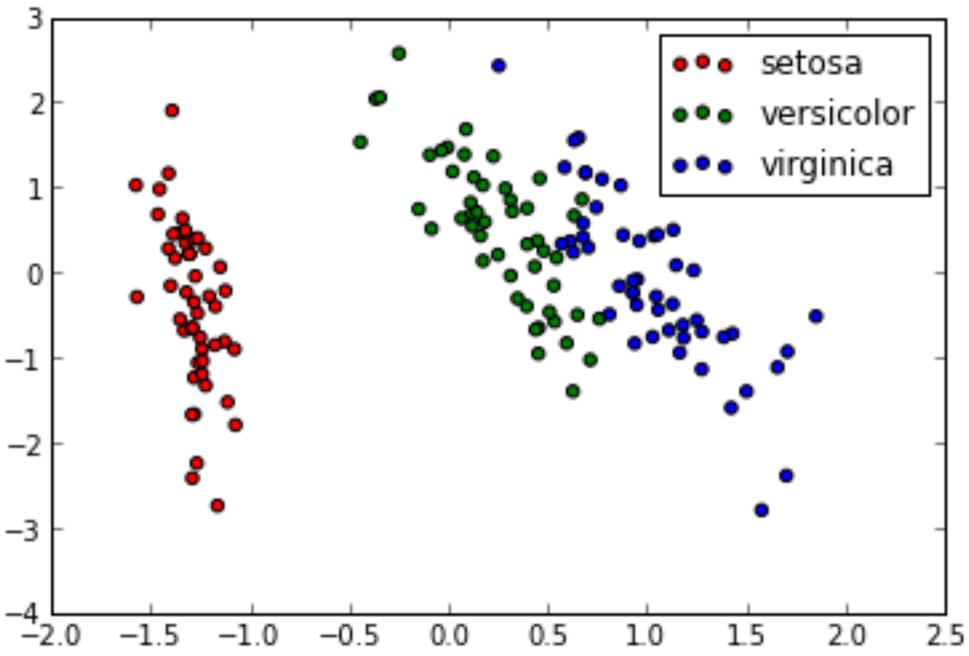


digit recognition

classification

example

Y



X

clustering

density estimation

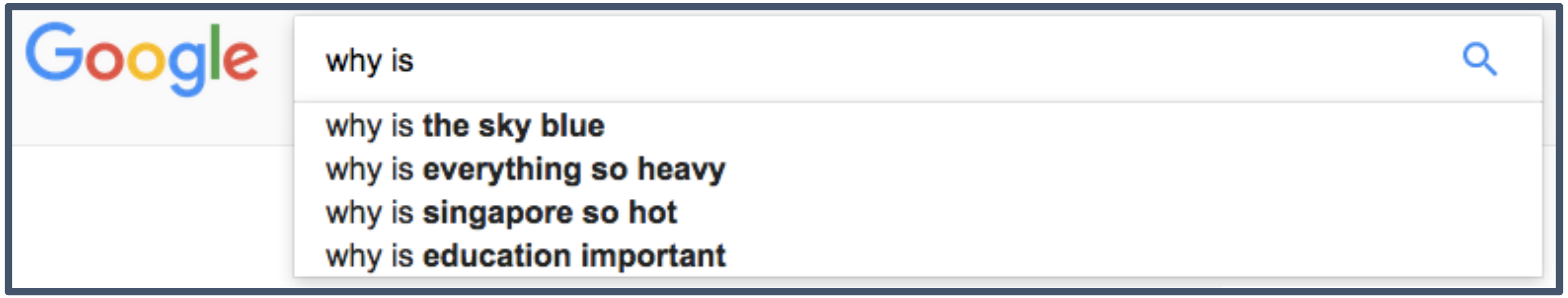
‘machine’ learning

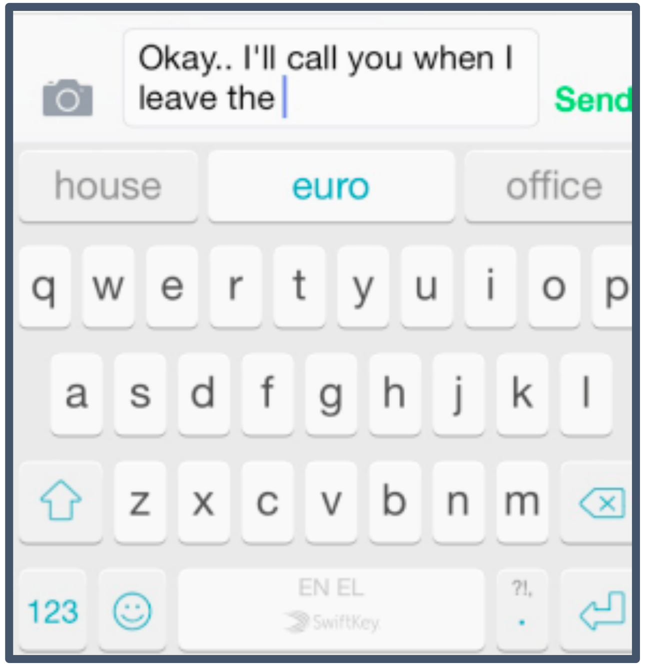
why do we need the machines?

to make learning
automated
and
efficient

big data
complex models

example: language





task: complete the sentence

language is complex

basic rules (syntax and grammar)

do not suffice for good predictions

requires complex models

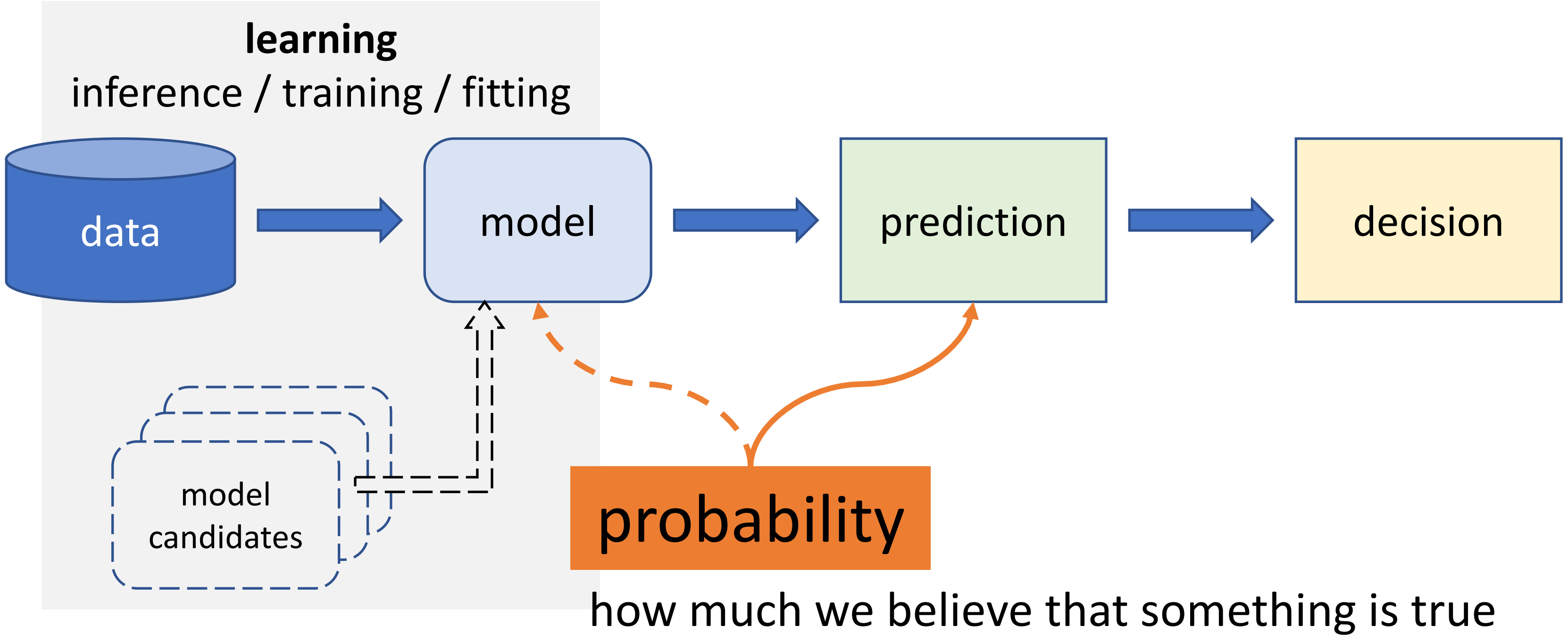
data

millions/billions of sentences/queries

user features

session attributes

ML pipeline


learning
inference / training / fitting
data
model
prediction
decision
model
candidates
probability
how much we believe that something is true

outline

- what is machine learning
  - examples of tasks: regression, classification, clustering
  - data, learning, prediction, decision; probability
- probability
- algorithms
  - regression
  - classification
  - clustering
- deep learning

probability

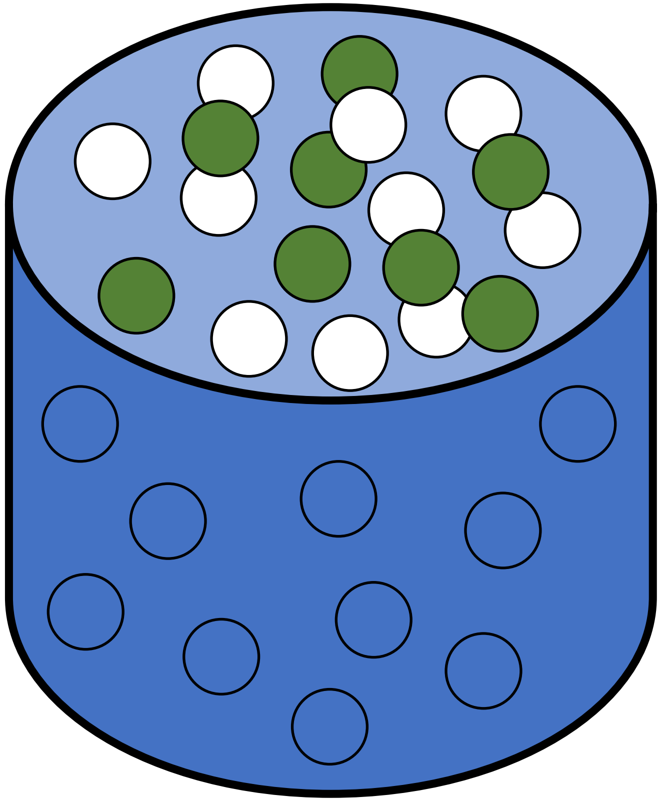
‘proposition’

how much we believe that something is true
GIVEN some information

VERY IMPORTANT!

0: impossible   1: certain

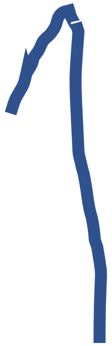
probability



a ball drops out of the box

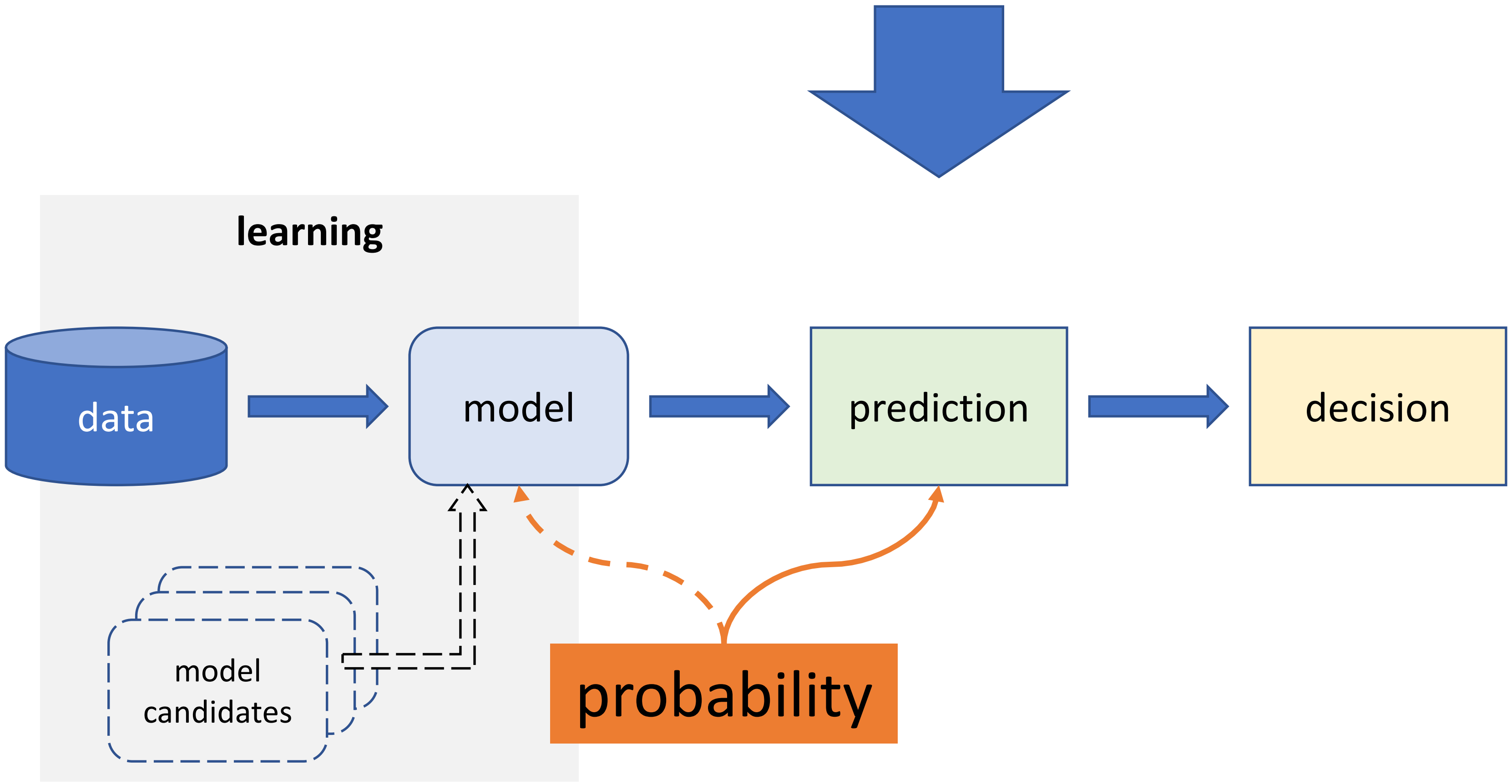
it is green

probability



this shape is ‘1’

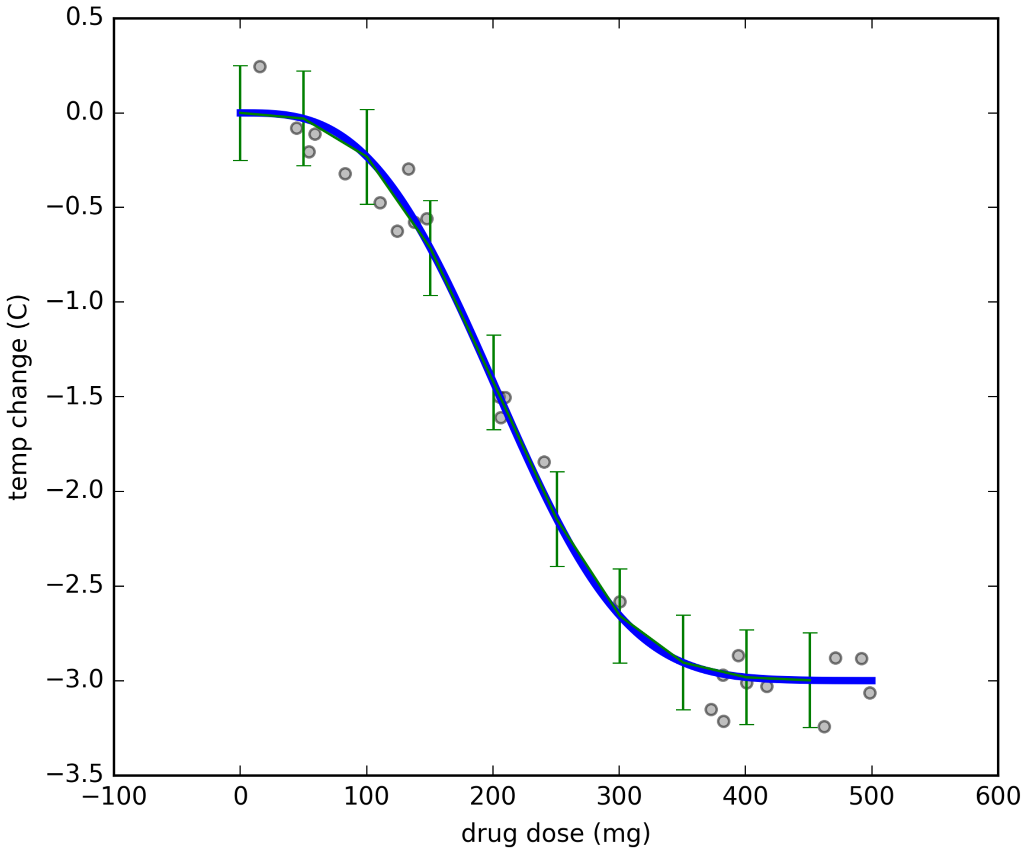
ML pipeline



assign probabilities to propositions

probability :: prediction

what is the probability that a dose of *300mg* drops temperature more than *2C* ?



probability of

temp. drop > 2C

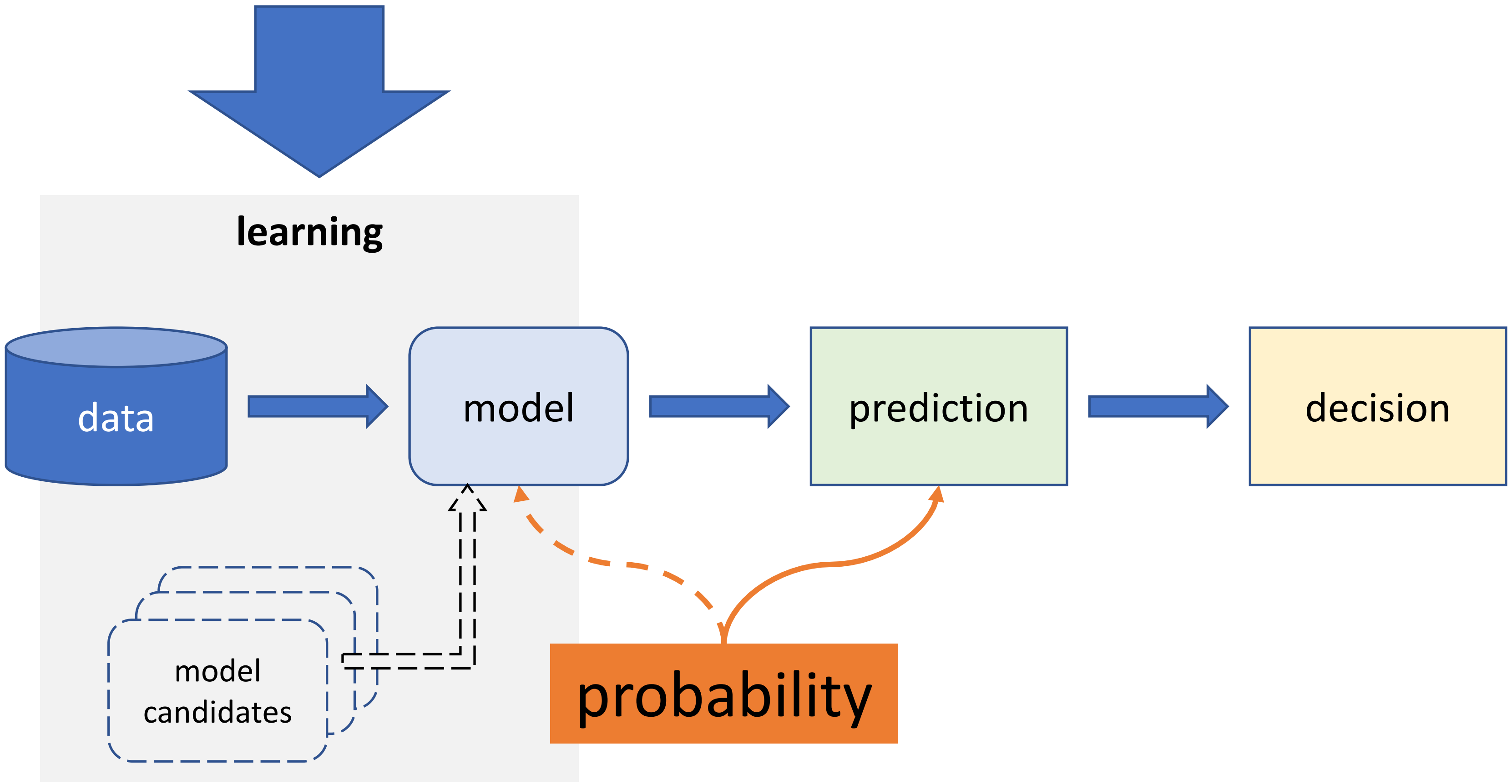
given

dose = 300 mg and model (see figure)

$$P(\overbrace{\text{temp. drop} > 2C}^{\text{proposition}} \mid \underbrace{\text{dose} = 300\text{ mg} \,;\, \text{model}}_{\text{given information}})$$

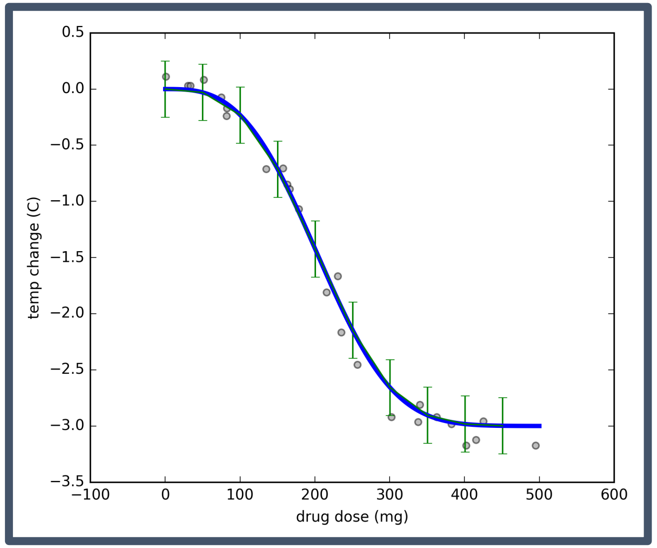
the value for this probability is provided by the model!

ML pipeline

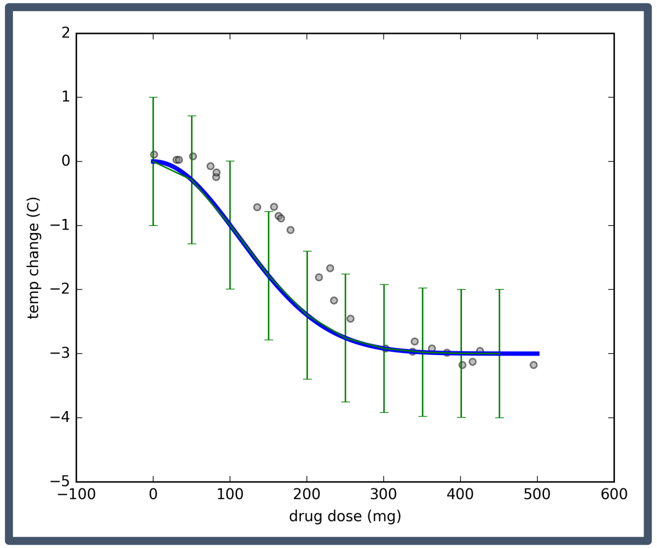


assign probabilities to models

probability :: learning



model M1

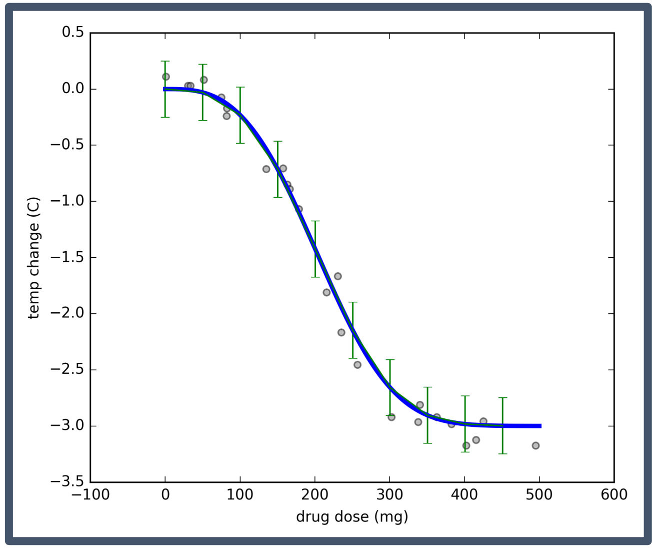


model M2

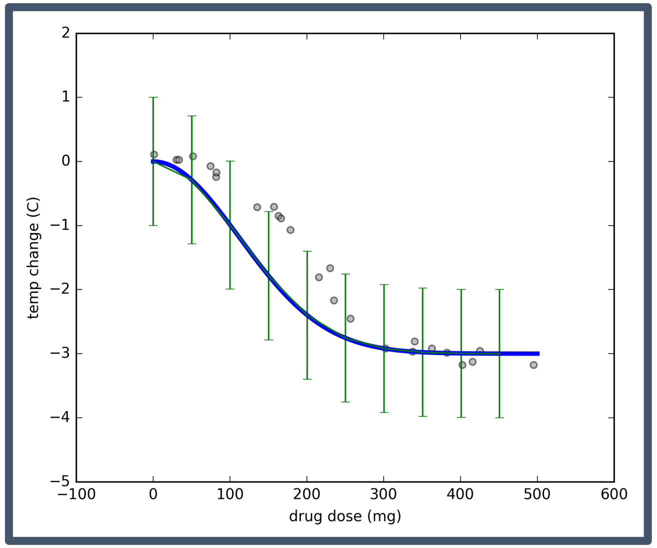
consider models where temperature drops
exponentially with dose
drop : dose^{-k} and error up to ε
what is the probability that the right model is M1 / M2 / … ?

probability of
model (k, ε)
given
data
k from -5 to + 5 and ε from -2 to +2

probability :: learning



model M1



model M2

p (model (k, ε) | data; k in [-5, + 5], ε in [-2, +2])

p (M | D ; I)

from Bayes' Rule, this is proportional to

p(data | M ; I) x p(M | I)

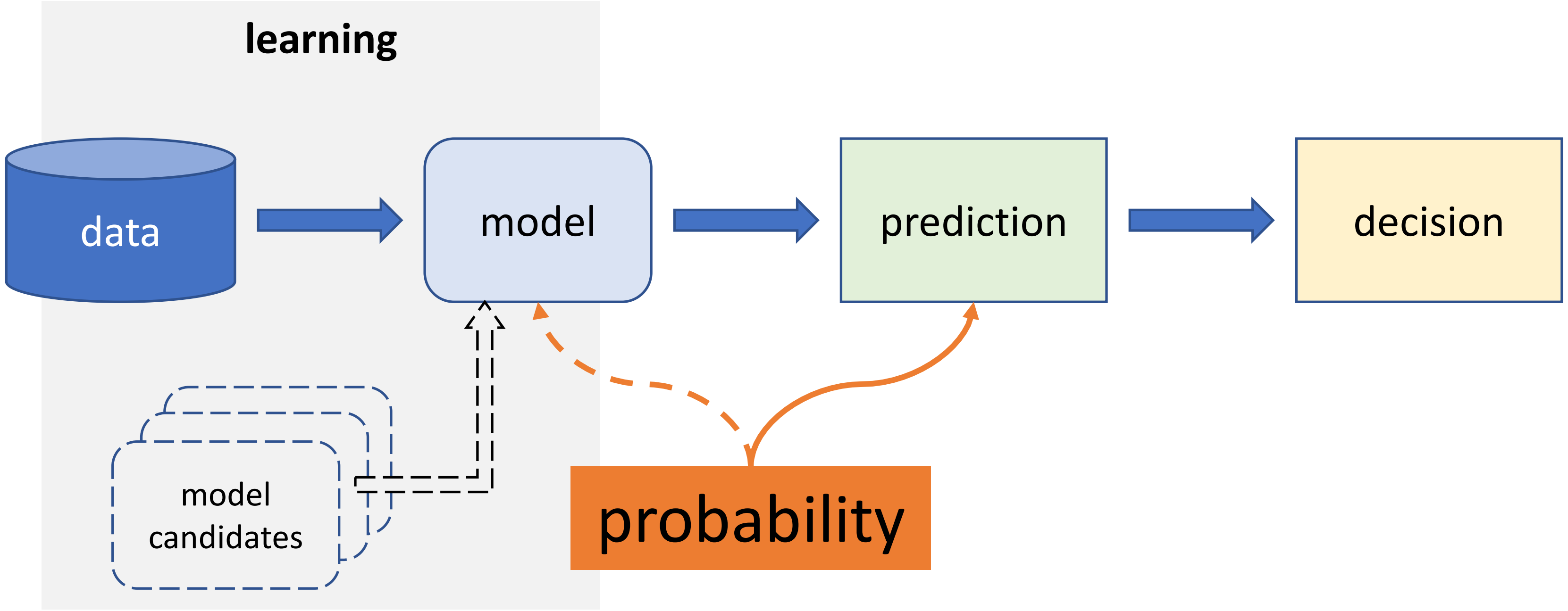
likelihood

prior

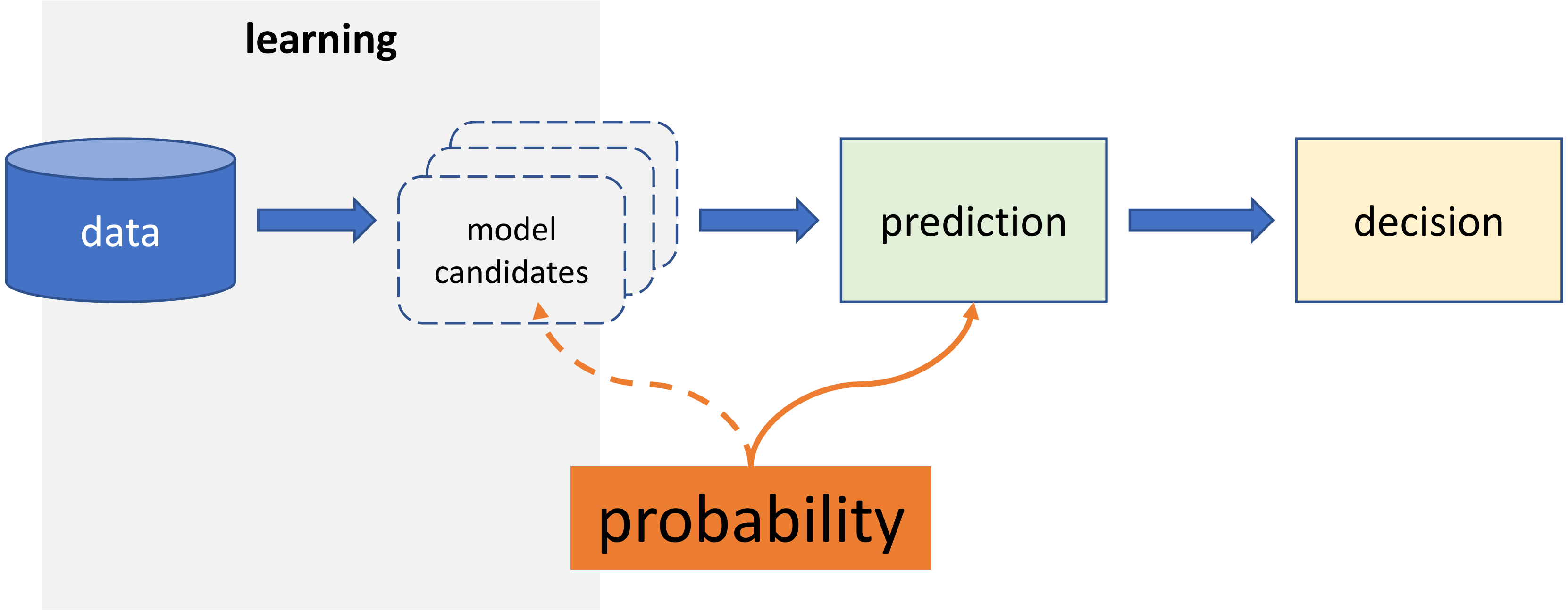
we choose the model of maximum probability

(do we have to?)

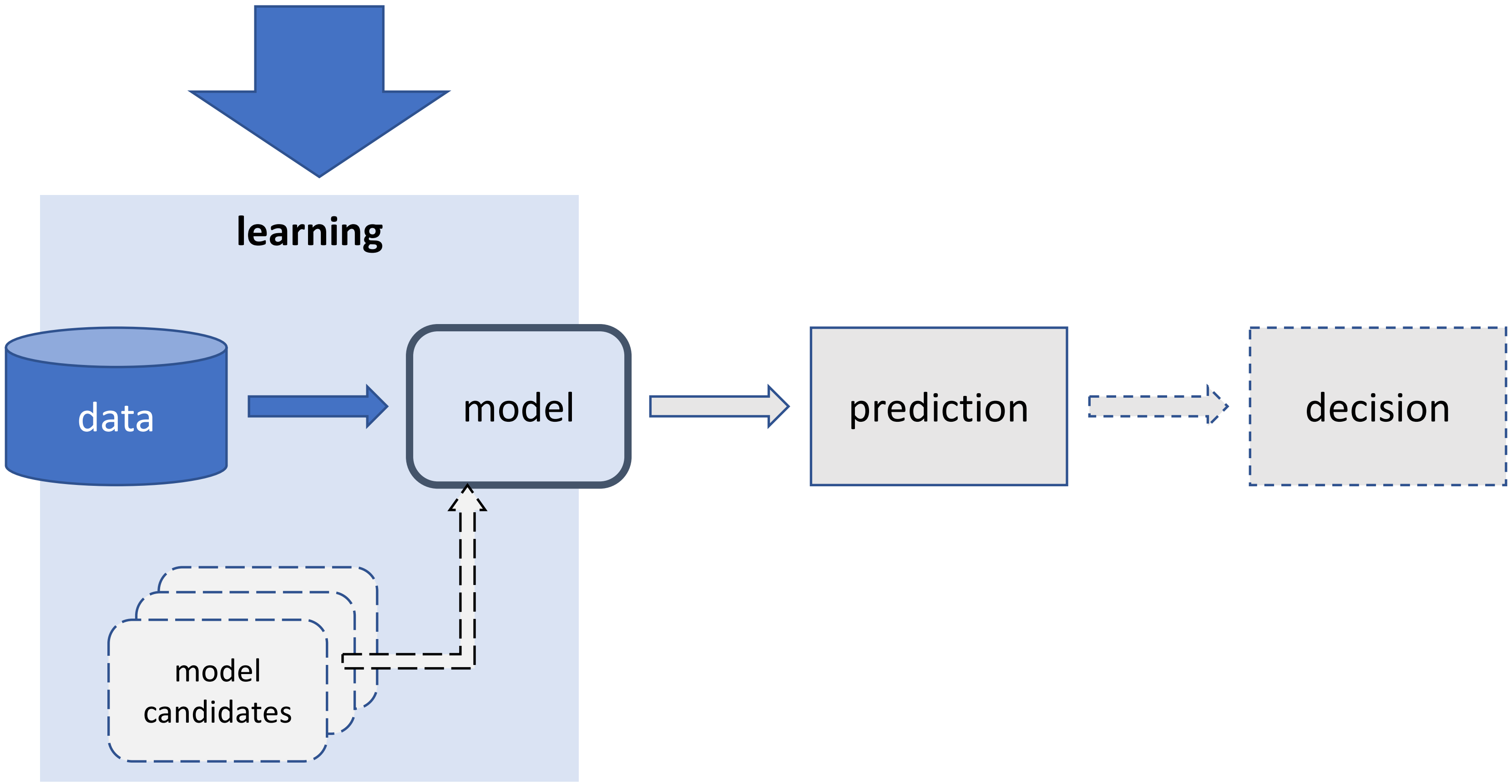
ML pipeline


learning
data
model
prediction
decision
model
candidates
probability

ML pipeline – the Bayesian way



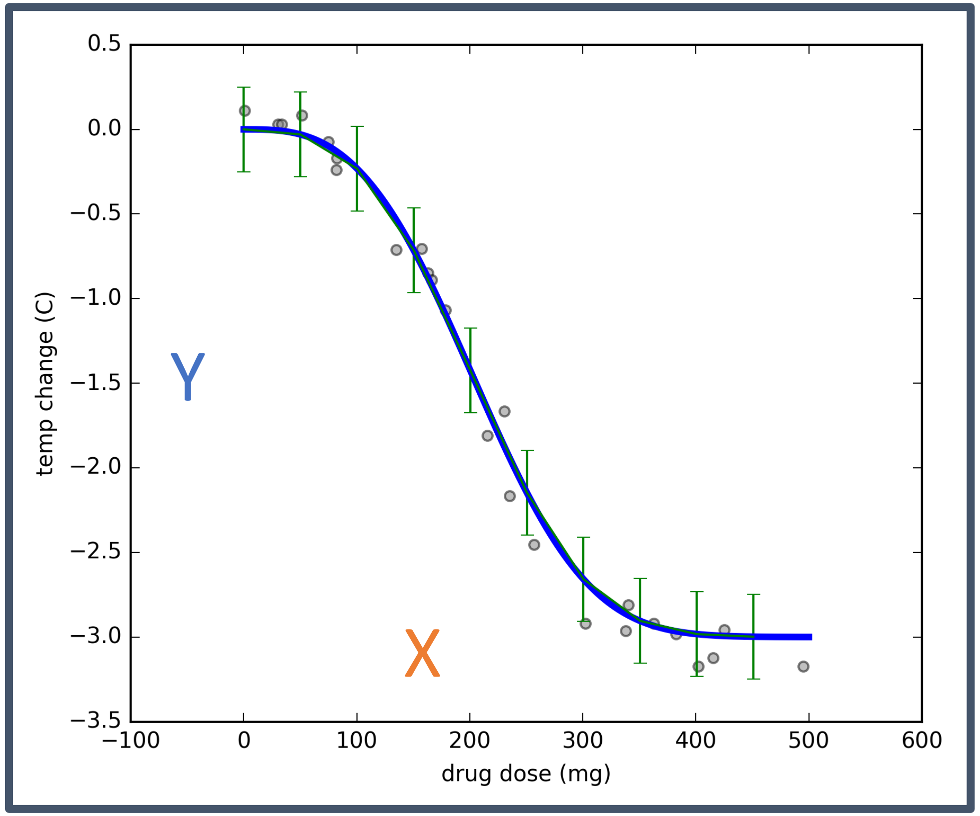
in what follows…



outline

- what is machine learning
  - examples of tasks: regression, classification, clustering
  - data, learning, prediction, decision; probability
- probability
- algorithms
  - regression
  - classification
  - clustering
- deep learning

regression

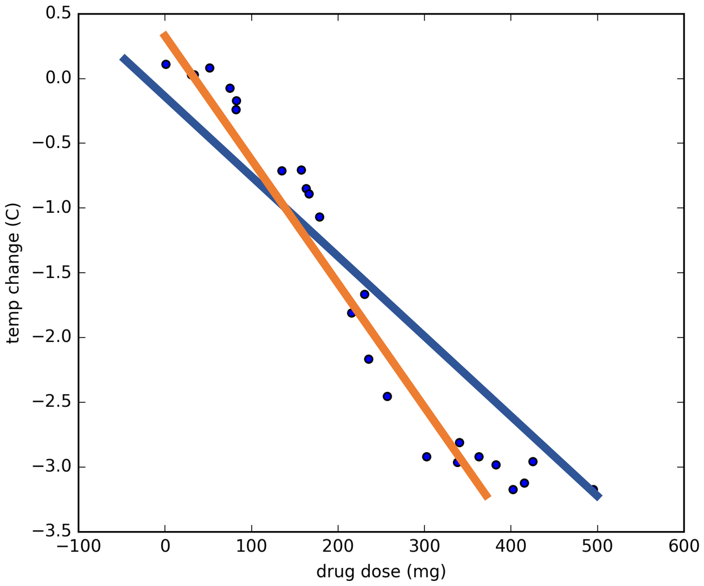


build model that provides
dp(Y = y | X = x; Model M)
for real-valued Y

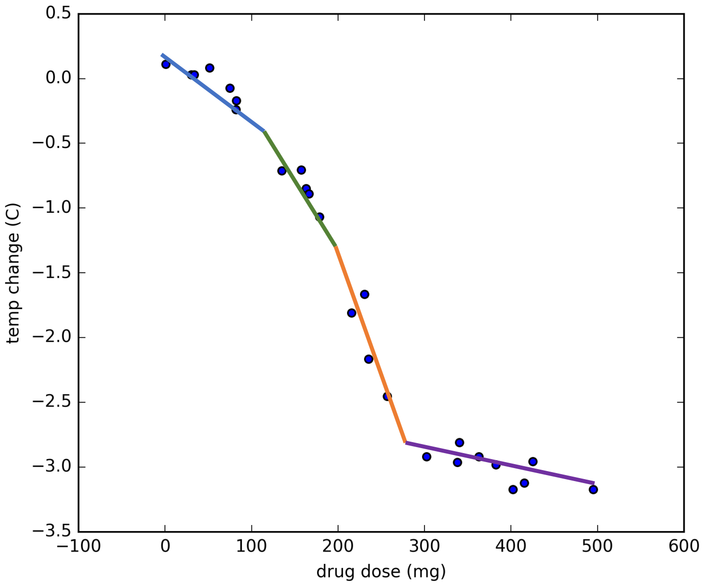
regression methods differ in
the set of model candidates
they consider

each method has corresponding
algorithm(s)
to search for best model

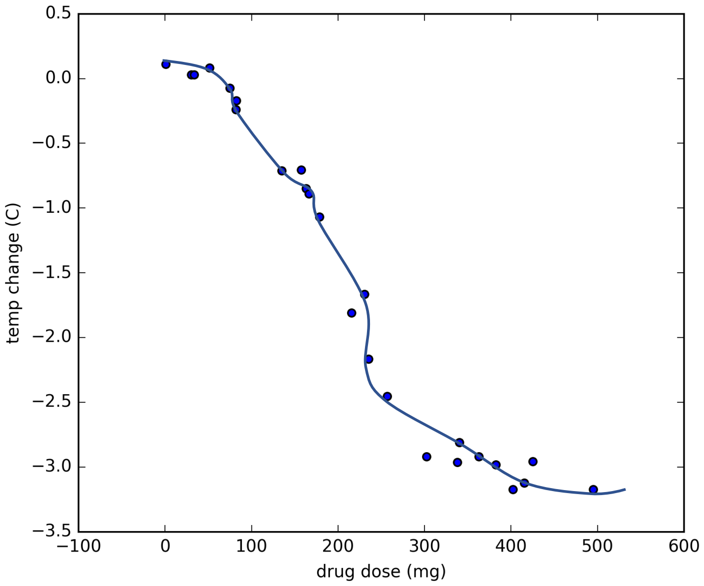
some regression methods



linear regression
line + error



segmented regression
k segments + errors



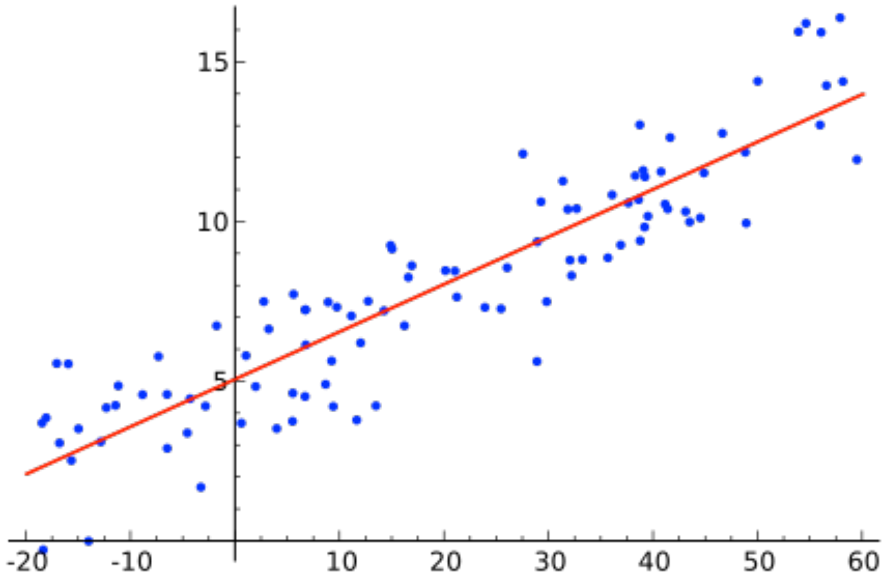
multinomial regression
curve + error

this is where methods differ

$$p(M \mid data ; I) \propto p(data \mid M ; I) \times p(M \mid I)$$

each model comes with its own

linear regression



$$
\begin{aligned}
Y &= \mathrm{E}(Y|X_1,\ldots,X_p)+\varepsilon \\
&= \beta_0+\sum_{j=1}^{p} X_j\beta_j+\varepsilon,
\end{aligned}
$$

solved with linear algebra if the data points are more than the dimensions

$$
\mathrm{RSS}(\beta) = \sum_{i=1}^{N}(y_i - f(x_i))^2
$$

ridge regression

ridge regression

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \Big(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \Big)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}.$$

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} \Big(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \Big)^2,$$
$$\text{subject to } \sum_{j=1}^{p} \beta_j^2 \le t,$$

lasso

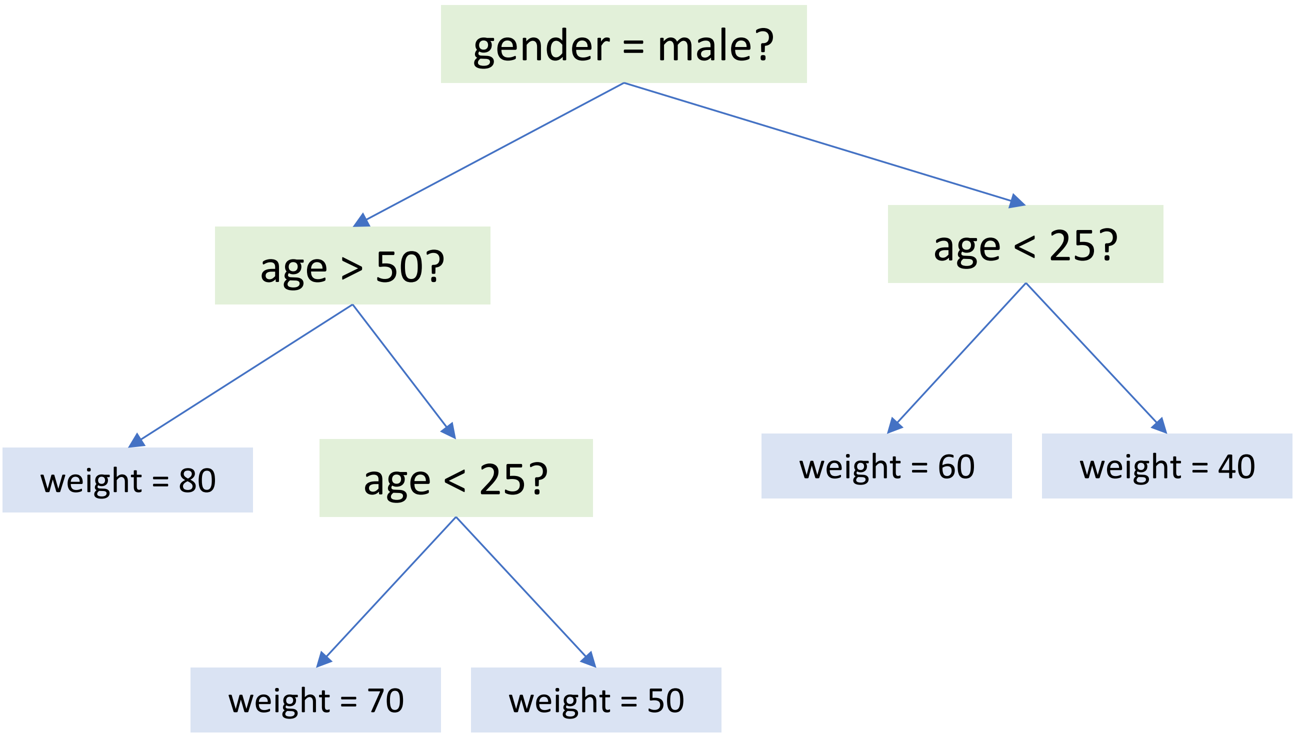
$$\boxed{\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}}$$

$$\boxed{\begin{aligned} \hat{\beta}^{\text{lasso}} &= \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 \\ & \text{subject to } \sum_{j=1}^{p} |\beta_j| \le t. \end{aligned}}$$

linear regression with shrinkage

the penalty on the size of β expresses a prior

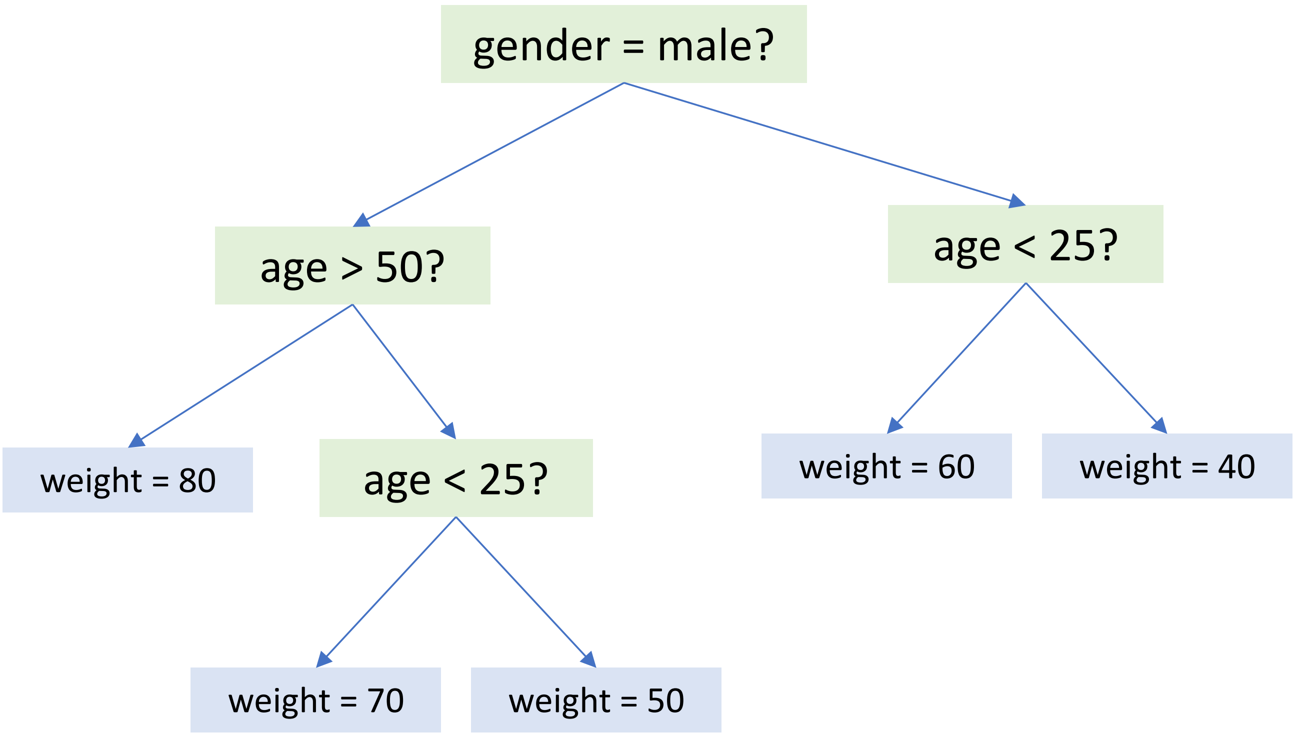
decision trees



predict the value of the leaf

build tree to minimize error
(e.g., square error)
subject to restrictions

random forests



build many decision trees on
random subset of data
random subset of features

combine predictions

neural networks

idea

apply a linear model on a non-linear transformation of the input

$$h_i = g(\boldsymbol{x}^\top \boldsymbol{W}_{:,i} + c_i)$$

$$g(z) = \max\{0, z\}$$

$$f(\boldsymbol{x}; \boldsymbol{W}, \boldsymbol{c}, \boldsymbol{w}, b) = \boldsymbol{w}^\top \max\{0, \boldsymbol{W}^\top \boldsymbol{x} + \boldsymbol{c}\} + b.$$

outline

- what is machine learning
  - examples of tasks: regression, classification, clustering
  - data, learning, prediction, decision; probability
- probability
- algorithms
  - regression
  - classification
  - clustering
- deep learning

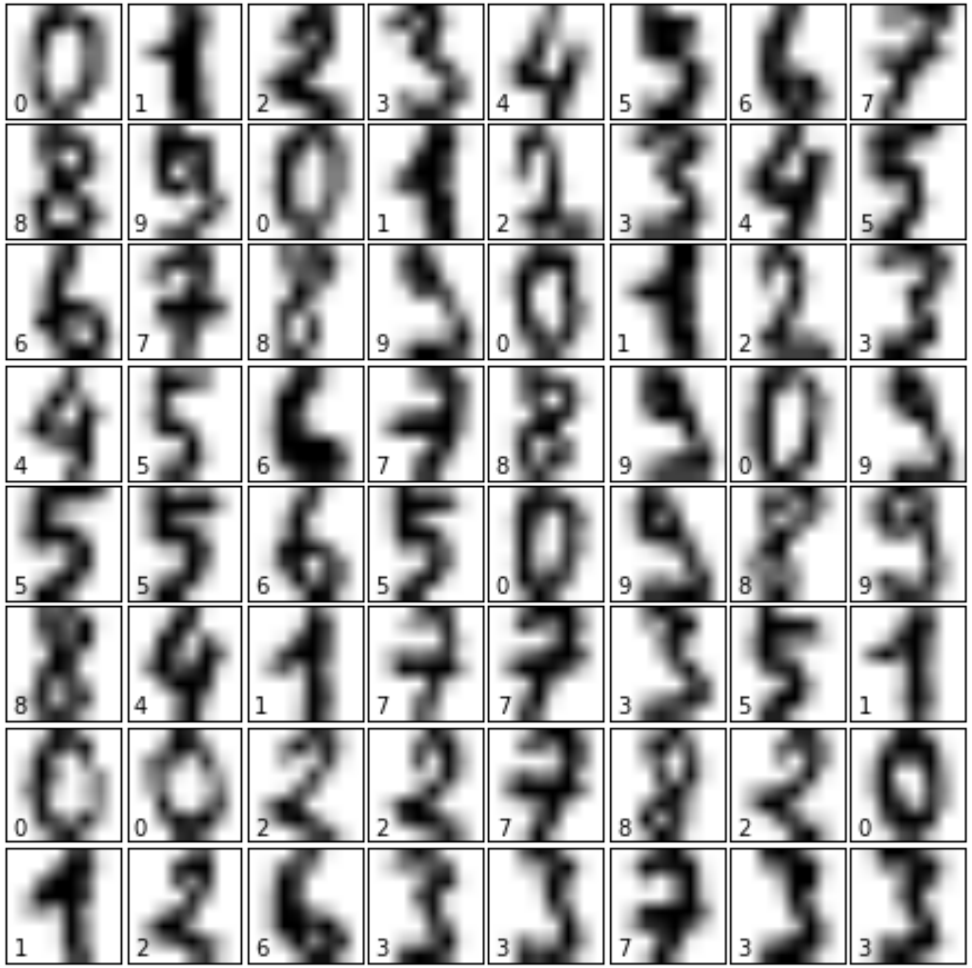
naïve-bayes

within each class, features are distributed independently

p(X | C = j) = $f_j(X) = \prod_{k=1}^{p} f_{jk}(X_k).$

$$
\begin{aligned}
\operatorname{logit} \frac{\Pr(G=\ell|X)}{\Pr(G=J|X)} &= \log \frac{\pi_\ell f_\ell(X)}{\pi_J f_J(X)} \\
&= \log \frac{\pi_\ell \prod_{k=1}^{p} f_{\ell k}(X_k)}{\pi_J \prod_{k=1}^{p} f_{Jk}(X_k)} \\
&= \log \frac{\pi_\ell}{\pi_J} + \sum_{k=1}^{p} \log \frac{f_{\ell k}(X_k)}{f_{Jk}(X_k)} \\
&= \alpha_\ell + \sum_{k=1}^{p} g_{\ell k}(X_k).
\end{aligned}
$$

classification



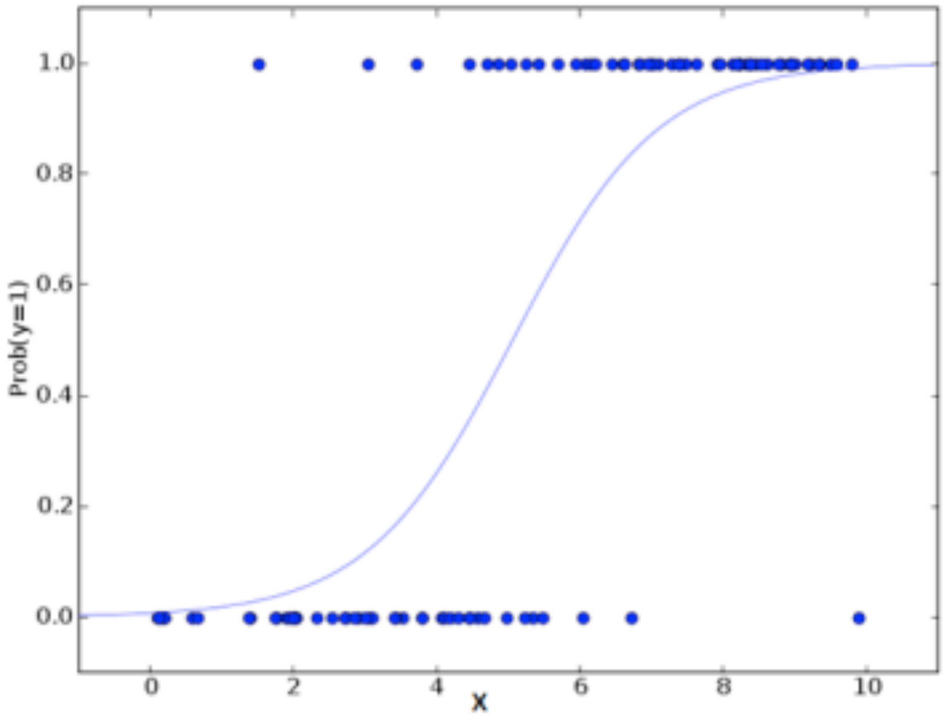
what is X and Y for digit recognition?

build model that provides
$p(Y = y \mid X = x; \text{Model } M)$
for categorically-valued Y

classification methods differ in
the set of model candidates
they consider

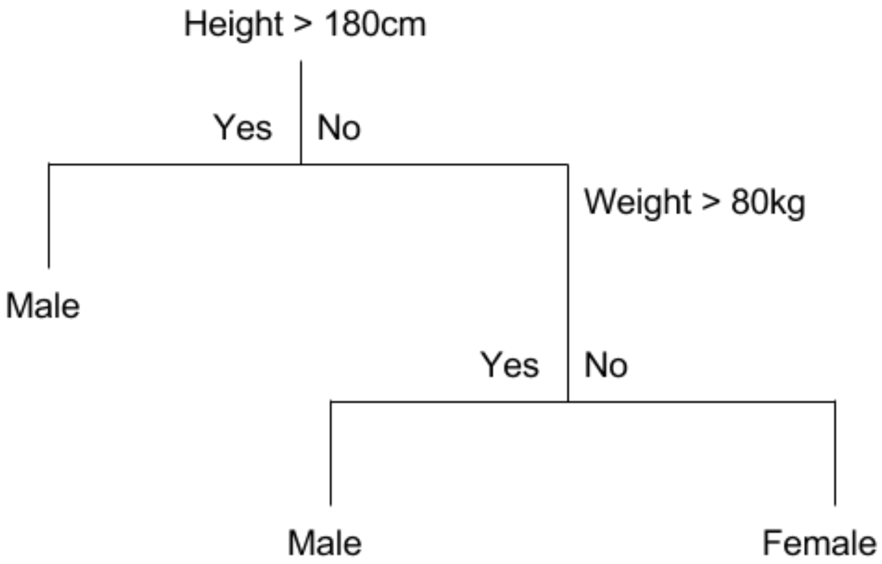
each method has corresponding
algorithm(s)
to search for best model

logistic regression



$$\Pr(Y_i = c) = \frac{e^{\boldsymbol{\beta}_c \cdot \mathbf{X}_i}}{\sum_{k=1}^{K} e^{\boldsymbol{\beta}_k \cdot \mathbf{X}_i}}$$

decision trees & random forests



very similar to regression methods
leafs assign probabilities to classes

neural networks

idea

apply a linear model on a non-linear transformation of the input

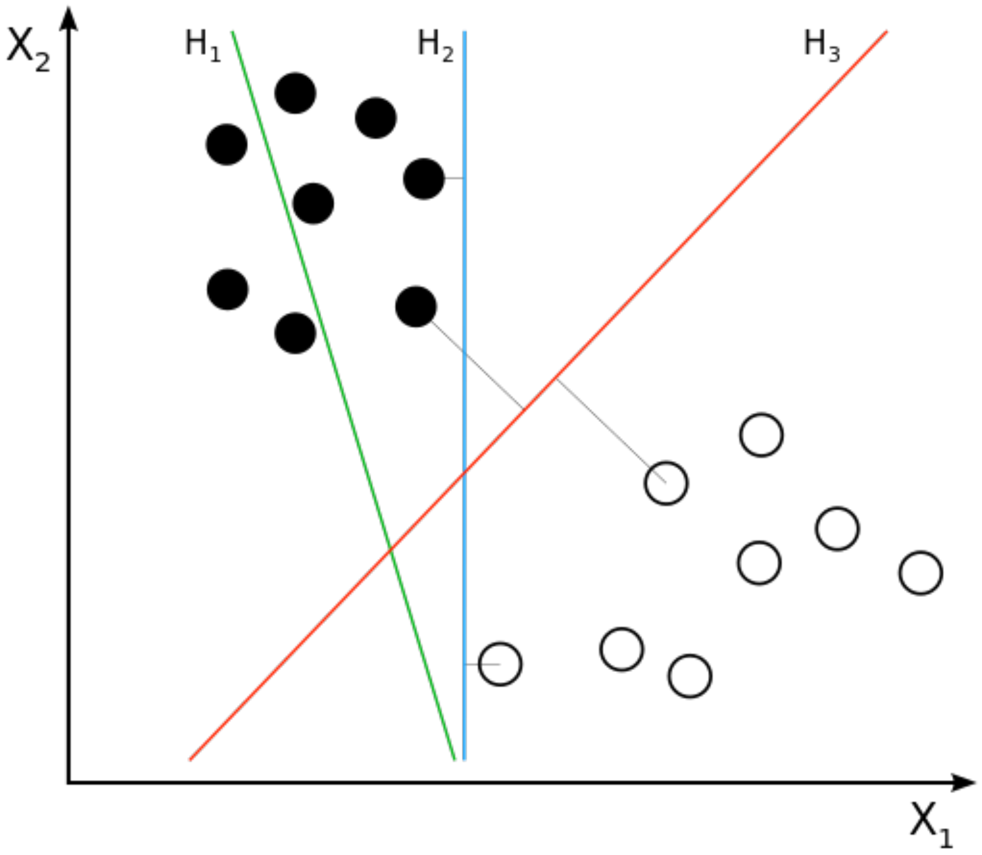
$$h_i = g(\boldsymbol{x}^\top \boldsymbol{W}_{:,i} + c_i)$$

$$g(z) = \max\{0, z\}$$

$$f(\boldsymbol{x}; \boldsymbol{W}, \boldsymbol{c}, \boldsymbol{w}, b) = \boldsymbol{w}^\top \max\{0, \boldsymbol{W}^\top \boldsymbol{x} + \boldsymbol{c}\} + b.$$

evidence for one class

support-vector machines



separate the classes
with hyperplanes

outline

- what is machine learning
  - examples of tasks: regression, classification, clustering
  - data, learning, prediction, decision; probability
- probability
- algorithms
  - regression
  - classification
  - clustering
- deep learning

supervised and unsupervised learning

the methods we saw for
regression and classification
are cases of ‘supervised’ learning

build model that provides
$p(Y = y \mid X = x; \text{Model } M)$

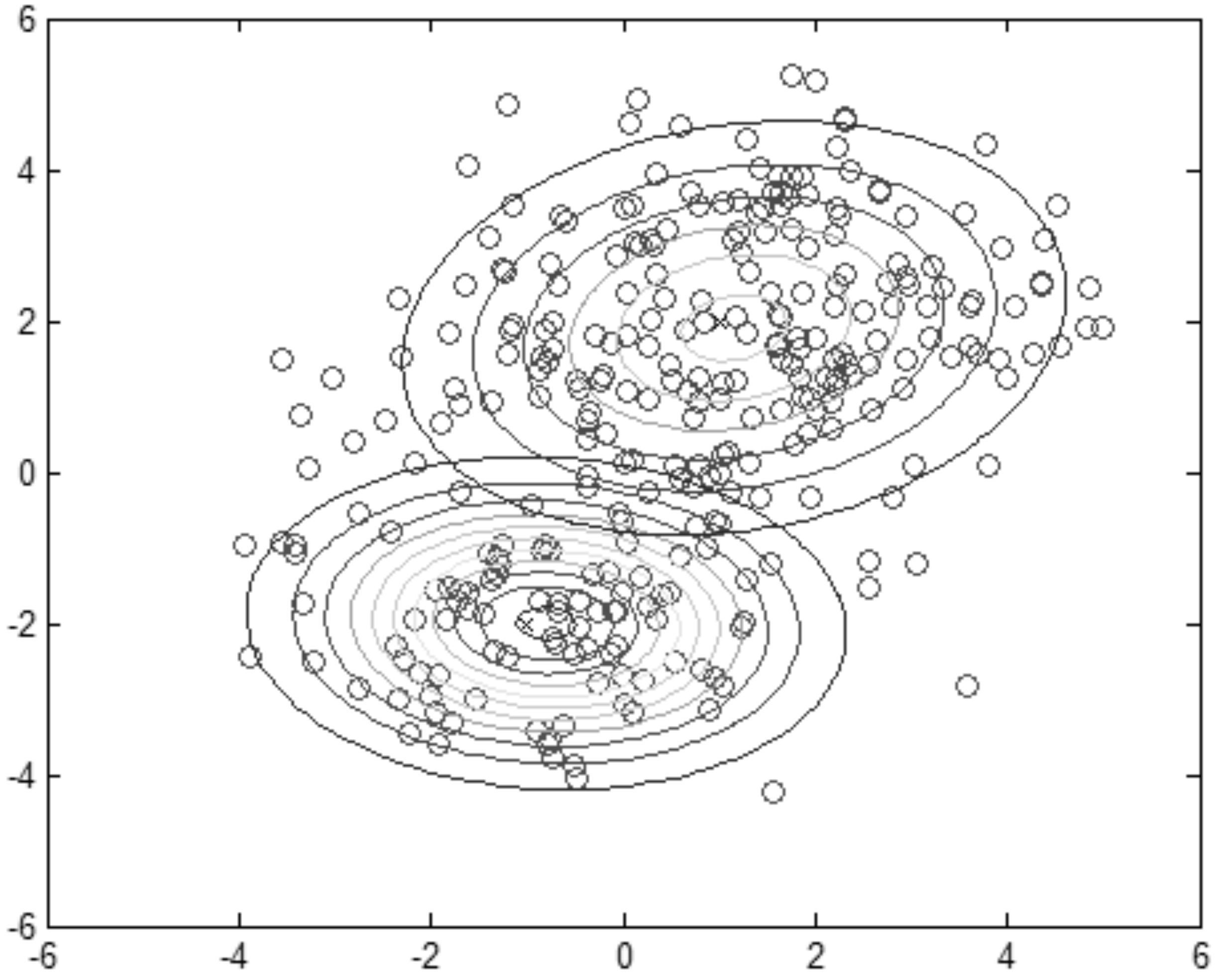
other data features

some data features

build model that provides
$p(X = x, Y = y; \text{Model } M)$

‘unsupervised’ learning

unsupervised learning



Y

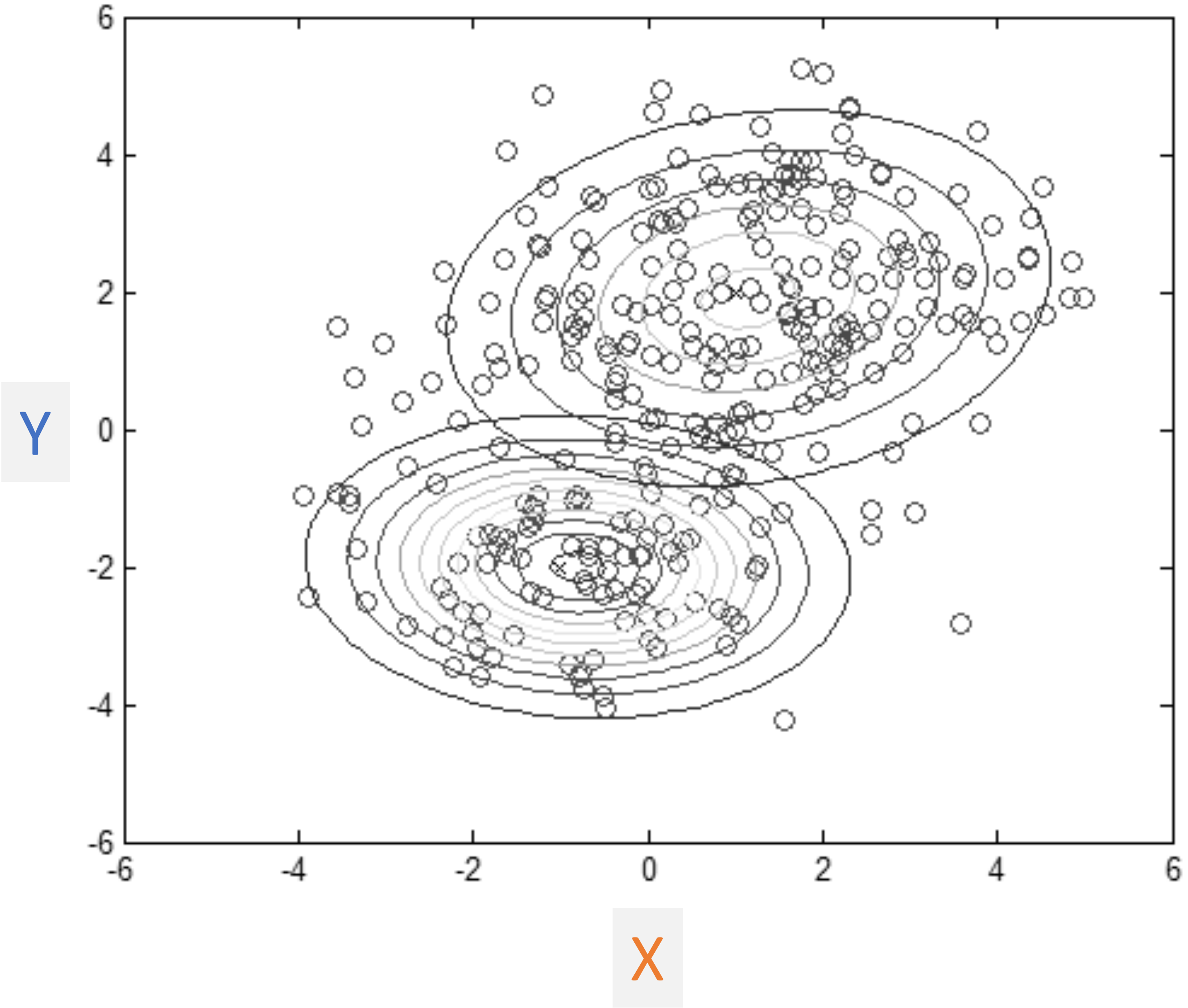
X

build model that provides

p(X = x, Y = y; Model M)

find structure in the data

k-means clustering

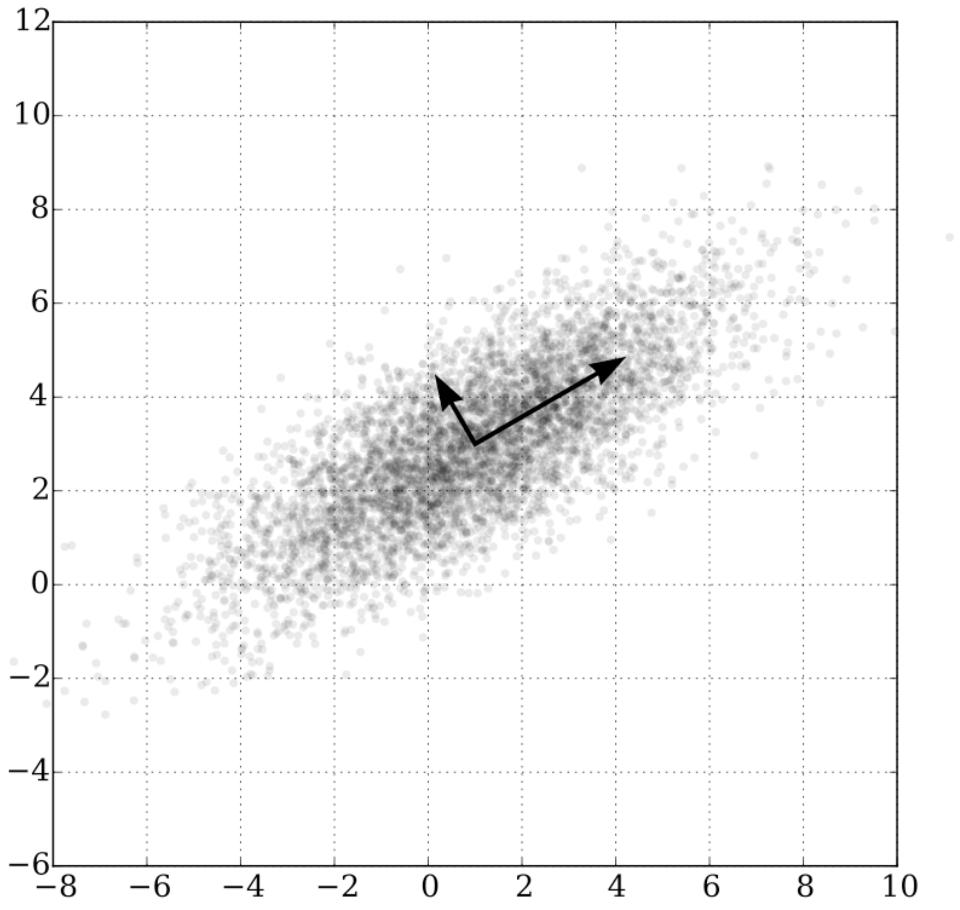


clustering

assign points to clusters
so that total distance from cluster center is minimized

k-means
assign points to cluster of nearest center
compute centers from assigned points
repeat

PCA



project the data on orthogonal system
so that successive dimensions maximize
remaining variance

outline

- what is machine learning
  - examples of tasks: regression, classification, clustering
  - data, learning, prediction, decision; probability
- probability
- algorithms
  - regression
  - classification
  - clustering
- deep learning

deep learning

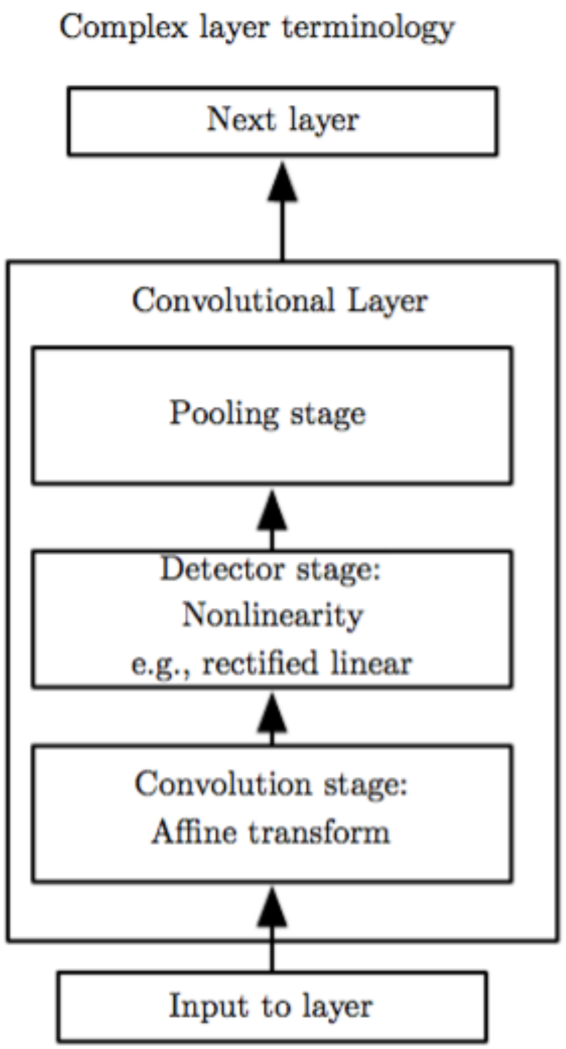
basically another name for ‘neural networks’
with many layers
and generalized structure

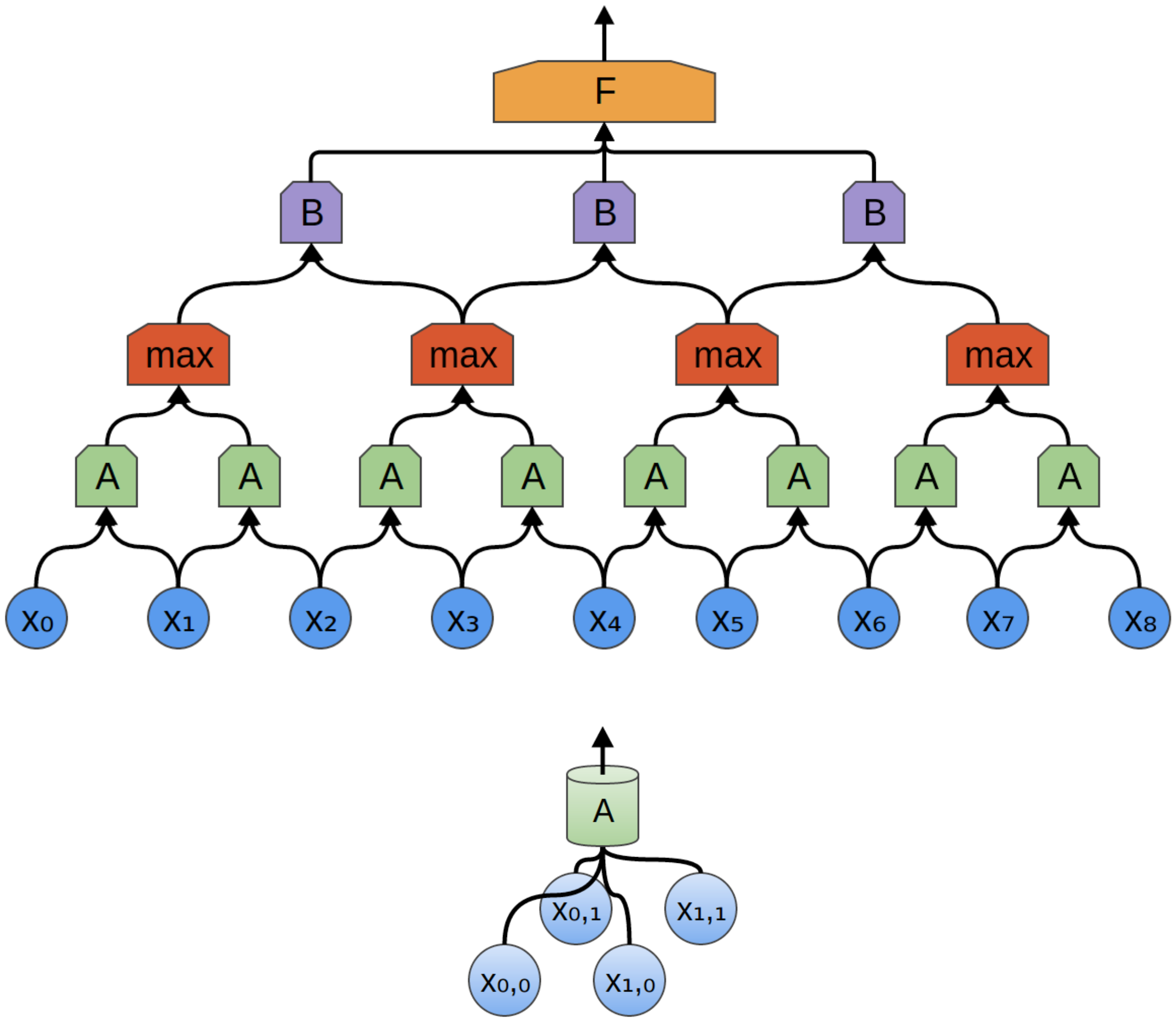
rebranded due to
efficiency and good results on difficult tasks

language (translation, sentence completion)
image recognition

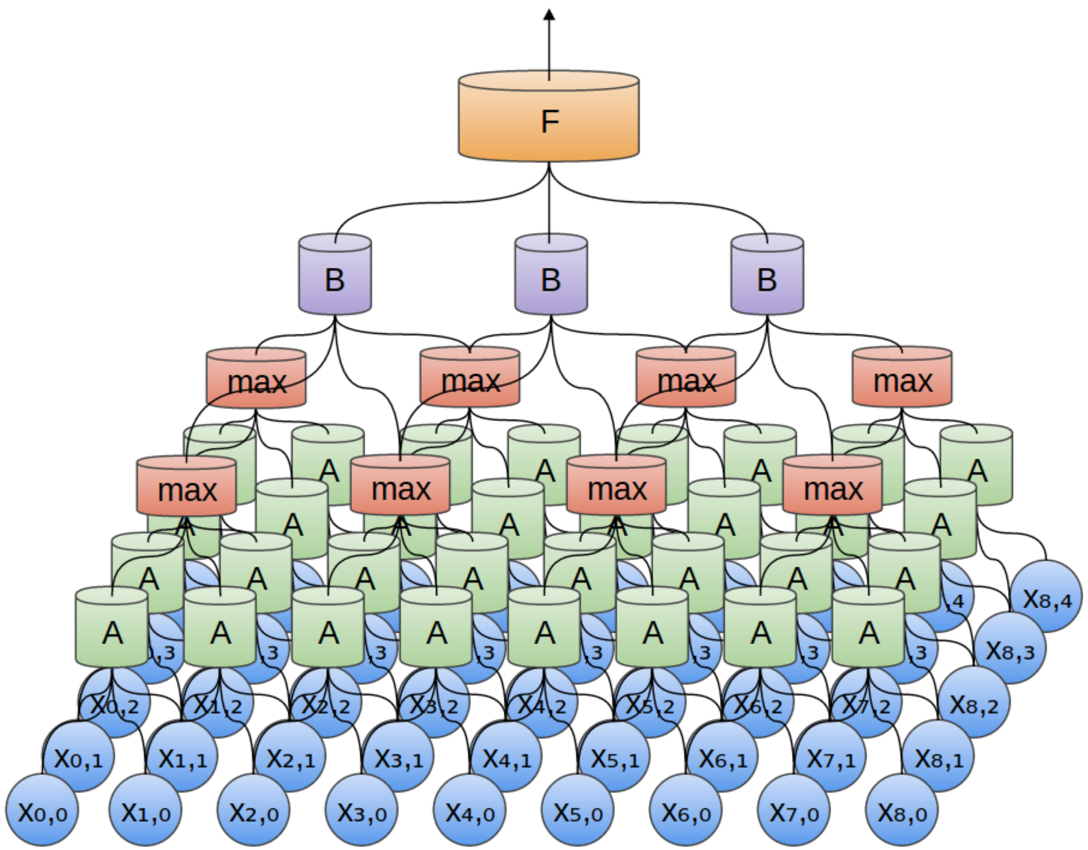
recurrent neural networks

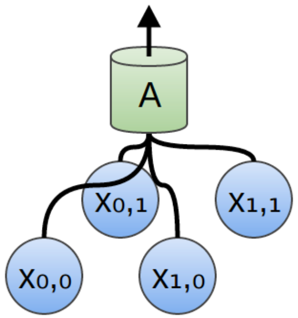
convolutional neural networks





convolutional neural networks





part 2: platforms and software

outline

- scikit-learn
- deep-learning libraries
- ML on the cloud
  - amazon, azure, google
- apache spark

scikit-learn

python ML library
on top of scipy stack

many general ML algorithms
standardized pipeline
ideal for fast prototyping
on moderate datasets

deep learning :: tensorflow

TensorFlow

deep learning library
uses computation graphs

deep learning :: other

torch.ch
open source machine learning library
scientific framework, programming language (Lua)
used by Facebook Research

theano
http://deeplearning.net/software/theano/
deep learning with efficient numerical operations

microsoft cognitive toolkit (cntk)
https://cntk.ai/
tensorflow alternative

keras
simpler tensorflow, theano, cntk in python

cloud :: google

Google Cloud Platform

Cloud ML Engine

basically offers the ML pipeline
with Deep Learning models
as implemented in Tensorflow

other services
trained models for other applications
speech, video or image tagging, translation
https://cloud.google.com/products/machine-learning/
pricing: about 0.5$ per hour

cloud :: other

amazon aws

classification and regression

with logistic and linear regression

microsoft azure

'cortana intelligence'

ML pipeline

apache spark

machine learning algorithms
on top of Spark

iterative optimization

part 3: hands-on session

outline

- scikit-learn
- tensorflow

scikit-learn

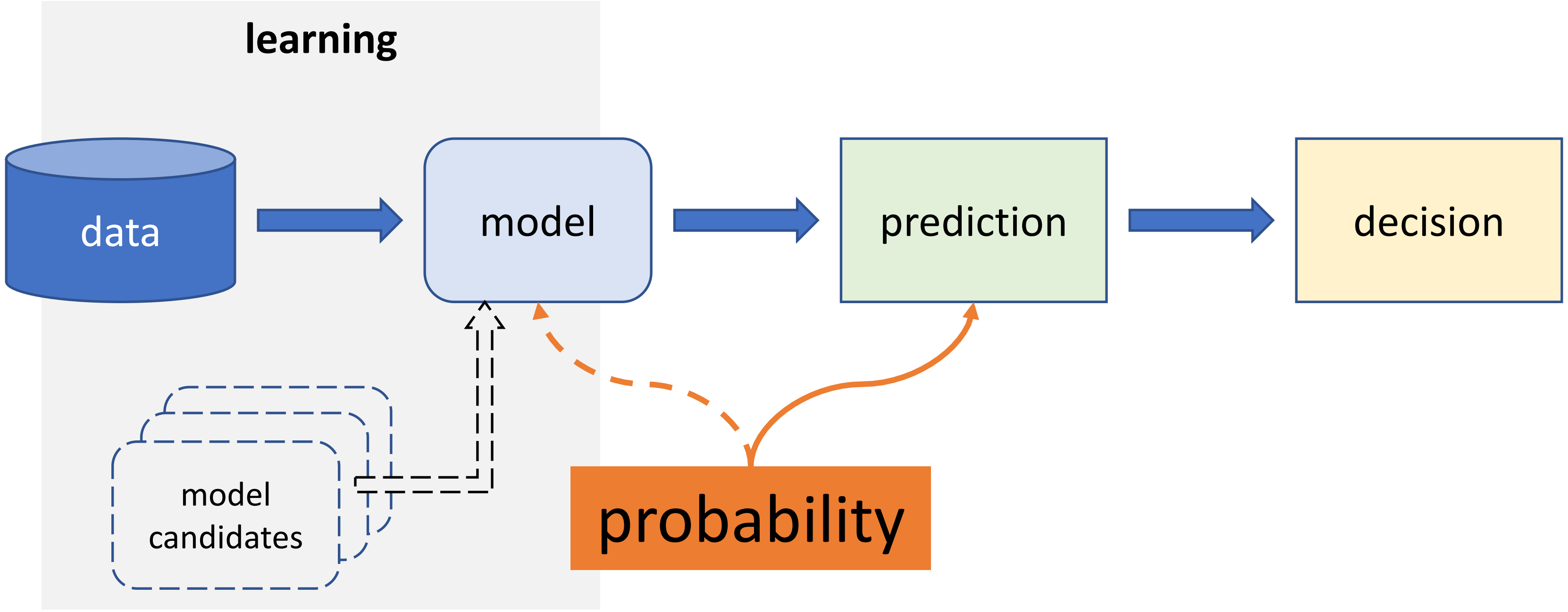
http://scikit-learn.org/

tensorflow

https://www.tensorflow.org/

the end…

ML pipeline


learning
data
model
prediction
decision
model
candidates
probability