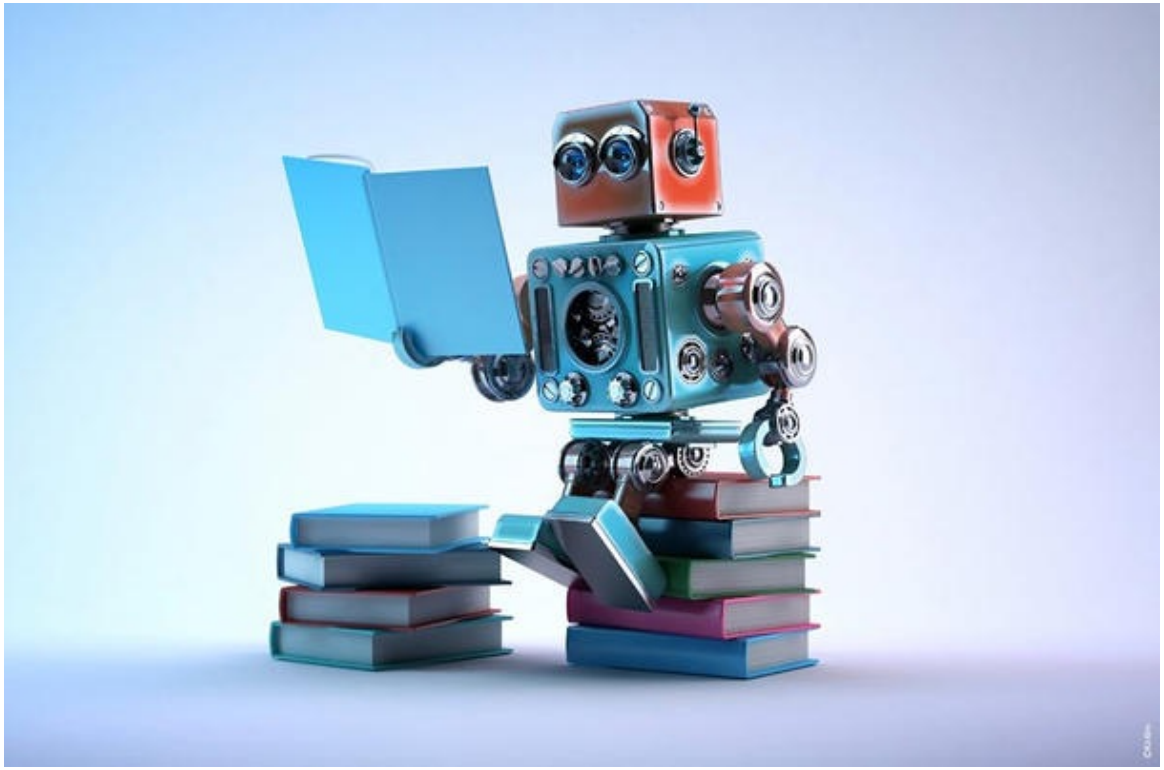
Intro to Machine learning



Sanja Lazarova-Molnar

Machine learning

- Definition
  - Getting a computer to do well on a task without explicitly programming it
  - Improving performance on a task based on experience

Example: Image classification

input	desired output
	apple
	pear
	tomato
	cow
	dog
	horse

Learning for episodic tasks

- The agent gets a series of unrelated problem instances and has to make some decision or inference about each of them
- In this case, “experience” comes in the form of *training data*

Training data

apple

pear

tomato

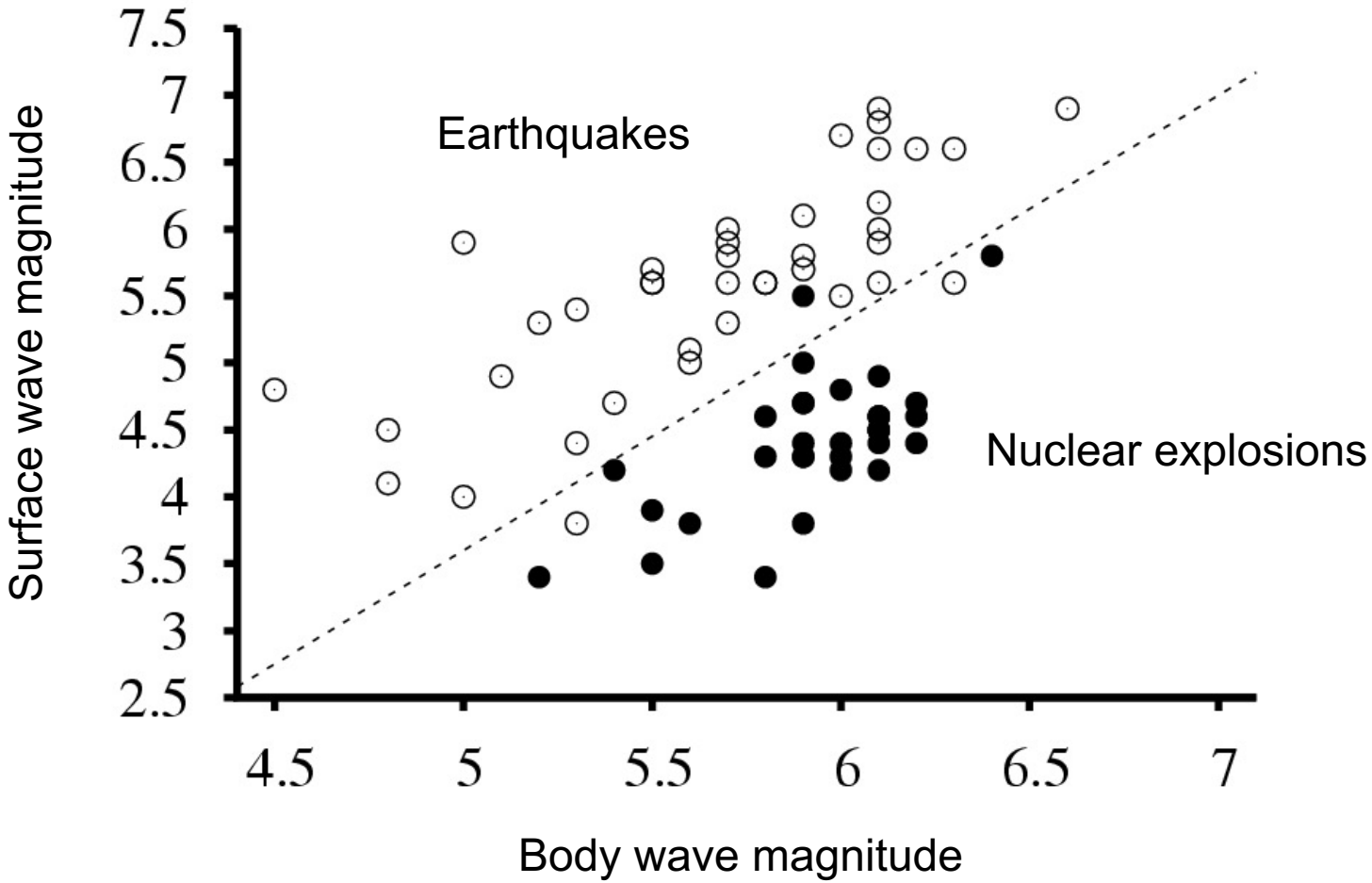
cow

dog

horse

- Key challenge of learning: *generalization* to unseen examples

Example 2: Seismic data classification



Example 3: Spam filter

Dear Sir.

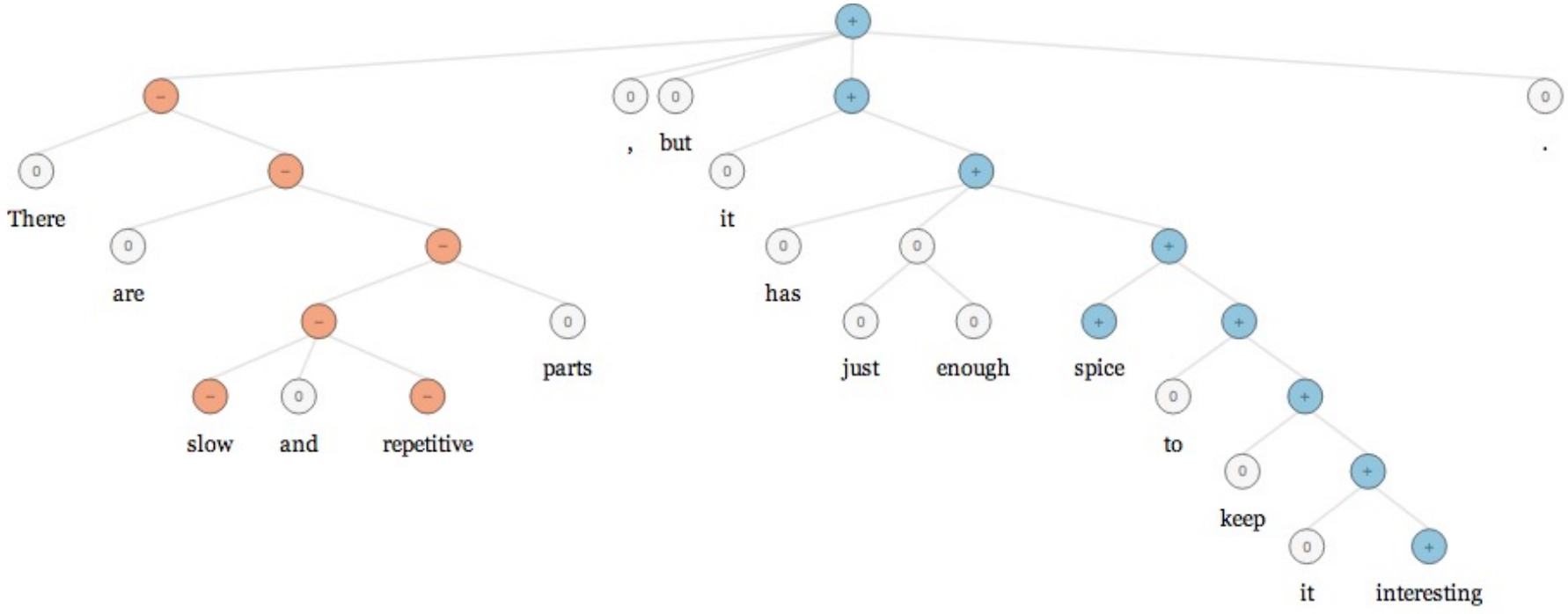
First, I must solicit your confidence in this transaction, this is by virture of its nature as being utterly confidencial and top secret. …

TO BE REMOVED FROM FUTURE MAILINGS, SIMPLY REPLY TO THIS MESSAGE AND PUT "REMOVE" IN THE SUBJECT.

99 MILLION EMAIL ADDRESSES
FOR ONLY $99

Ok, Iknow this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power nothing happened.

Example 4: Sentiment analysis

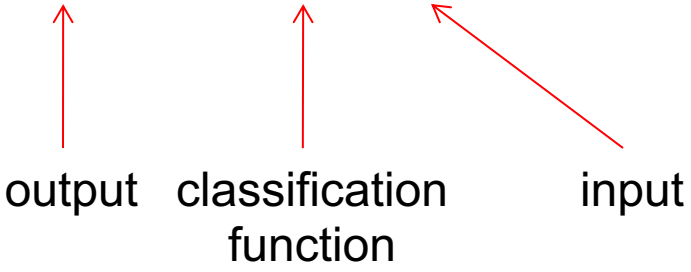


http://gigaom.com/2013/10/03/stanford-researchers-to-open-source-model-they-say-has-nailed-sentiment-analysis/

http://nlp.stanford.edu:8080/sentiment/rntnDemo.html

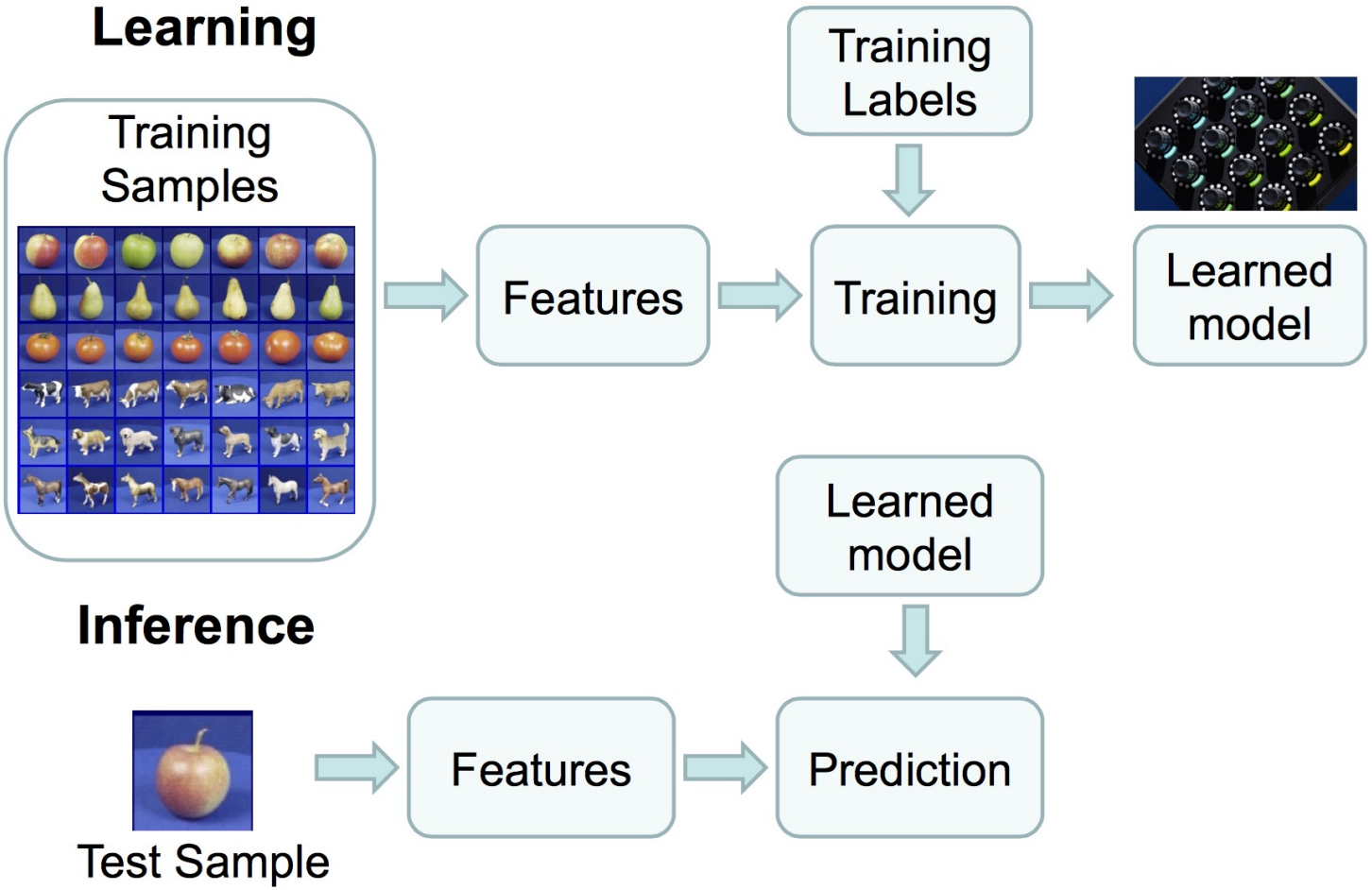
The basic machine learning framework

$$y = f(\mathbf{x})$$



- **Learning:** given a *training set* of labeled examples $\{(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_N,y_N)\}$, estimate the parameters of the prediction function f
- **Inference:** apply f to a never-before-seen *test example* $\mathbf{x}$ and output the predicted value $y = f(\mathbf{x})$

Learning and inference pipeline



Naïve Bayes classifier

$$f(\boldsymbol{x}) = \arg\max_y P(y \mid \boldsymbol{x})$$

$$\propto \arg\max_y P(y) P(\boldsymbol{x} \mid y)$$

$$= \arg\max_y P(y) \prod_d P(x_d \mid y)$$

A single dimension or attribute of **x**

Example of Naïve Bayes Classifier

[Optional]

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

A: attributes

M: mammals

N: non-mammals

$$P(A \mid M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} \times \frac{2}{7} = 0.06$$

$$P(A \mid N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042$$

$$P(A \mid M)P(M) = 0.06 \times \frac{7}{20} = 0.021$$

$$P(A \mid N)P(N) = 0.004 \times \frac{13}{20} = 0.0027$$

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

P(A|M)P(M) > P(A|N)P(N)

=> Mammals

Decision tree classifier

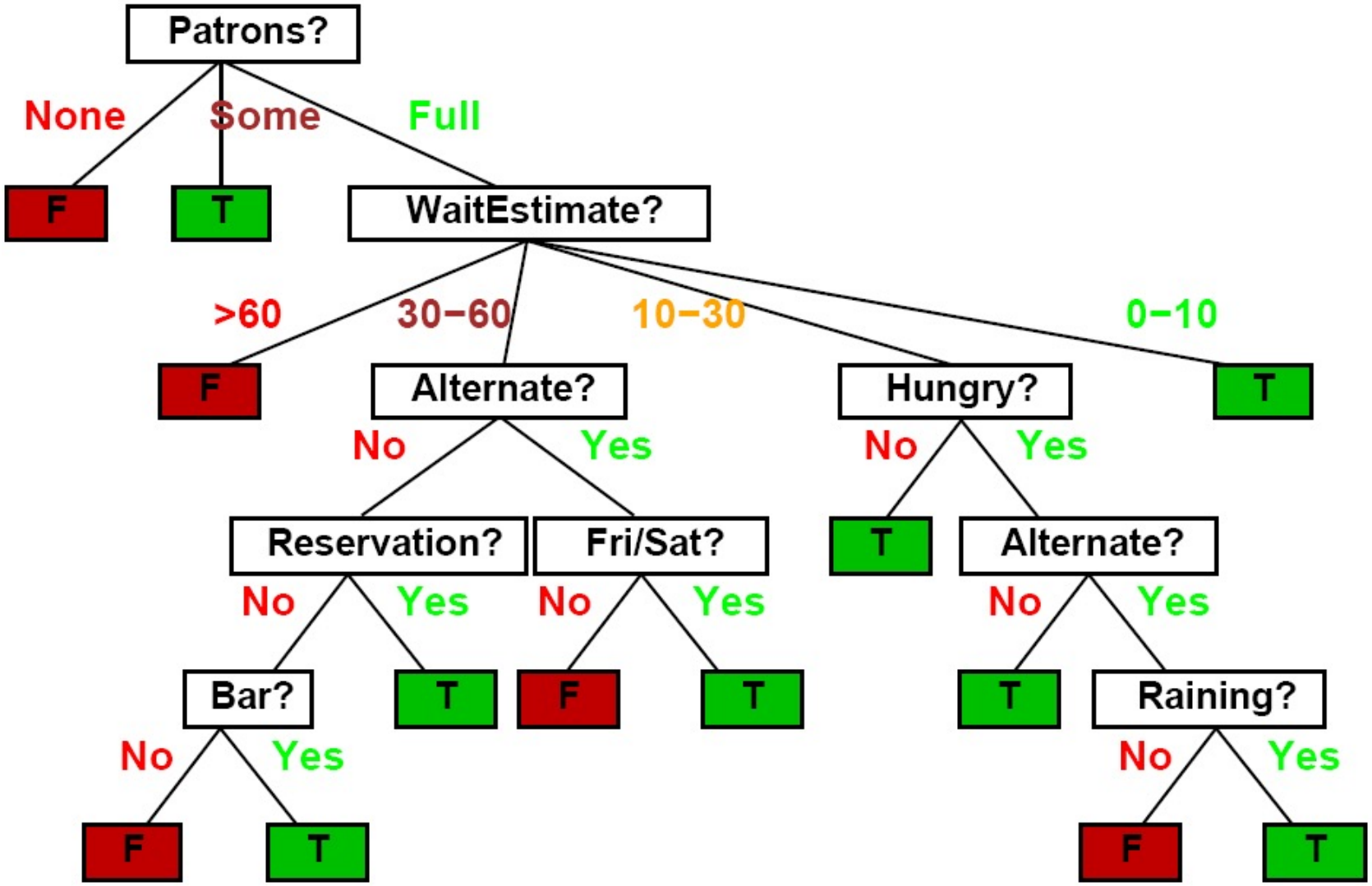
Example problem - decide whether to wait for a table at a restaurant, based on the following attributes:

1. **Alternate:** is there an alternative restaurant nearby?
2. **Bar:** is there a comfortable bar area to wait in?
3. **Fri/Sat:** is today Friday or Saturday?
4. **Hungry:** are we hungry?
5. **Patrons:** number of people in the restaurant (None, Some, Full)
6. **Price:** price range ($, $$, $$$)
7. **Raining:** is it raining outside?
8. **Reservation:** have we made a reservation?
9. **Type:** kind of restaurant (French, Italian, Thai, Burger)
10. **WaitEstimate:** estimated waiting time (0-10, 10-30, 30-60, >60)

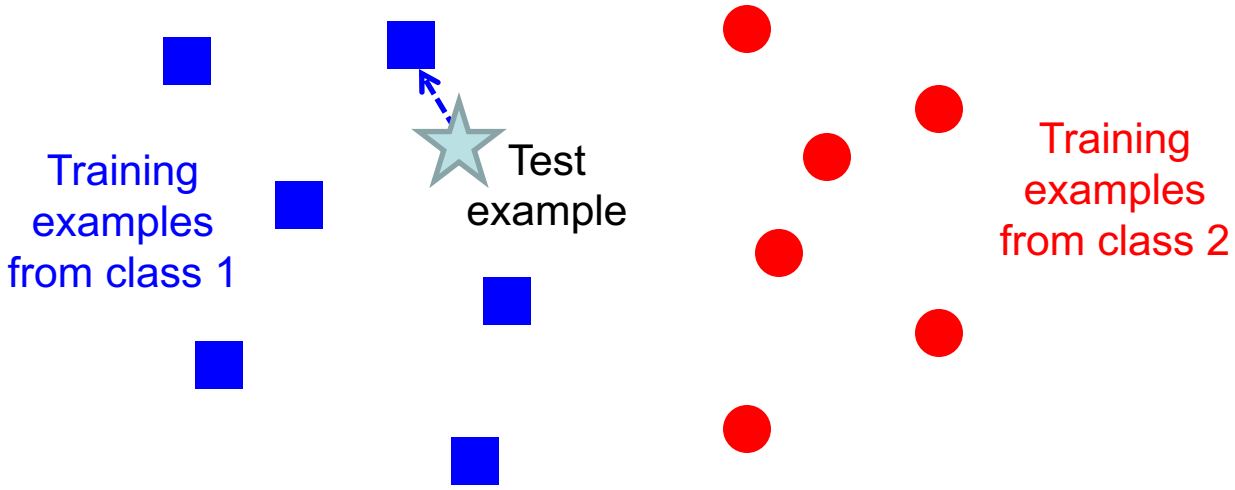
Decision tree classifier

Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*Wait*
X_1	T	F	F	T	Some	$$$	F	T	French	0–10	T
X_2	T	F	F	T	Full	$	F	F	Thai	30–60	F
X_3	F	T	F	F	Some	$	F	F	Burger	0–10	T
X_4	T	F	T	T	Full	$	F	F	Thai	10–30	T
X_5	T	F	T	F	Full	$$$	F	T	French	>60	F
X_6	F	T	F	T	Some	$$	T	T	Italian	0–10	T
X_7	F	T	F	F	None	$	T	F	Burger	0–10	F
X_8	F	F	F	T	Some	$$	T	T	Thai	0–10	T
X_9	F	T	T	F	Full	$	T	F	Burger	>60	F
X_{10}	T	T	T	T	Full	$$$	F	T	Italian	10–30	F
X_{11}	F	F	F	F	None	$	F	F	Thai	0–10	F
X_{12}	T	T	T	T	Full	$	F	F	Burger	30–60	T

Decision tree classifier



Nearest neighbor classifier

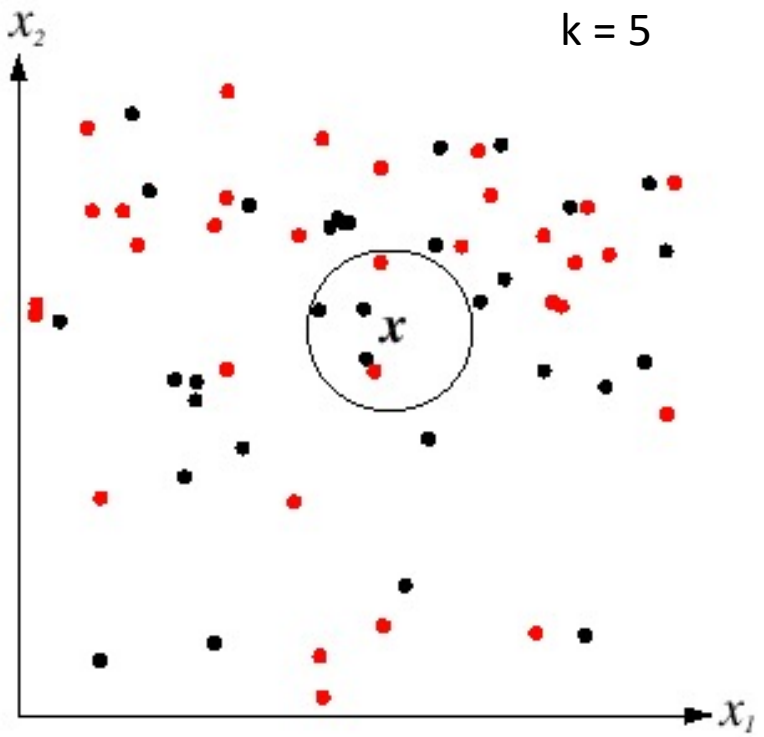


f(**x**) = label of the training example nearest to **x**

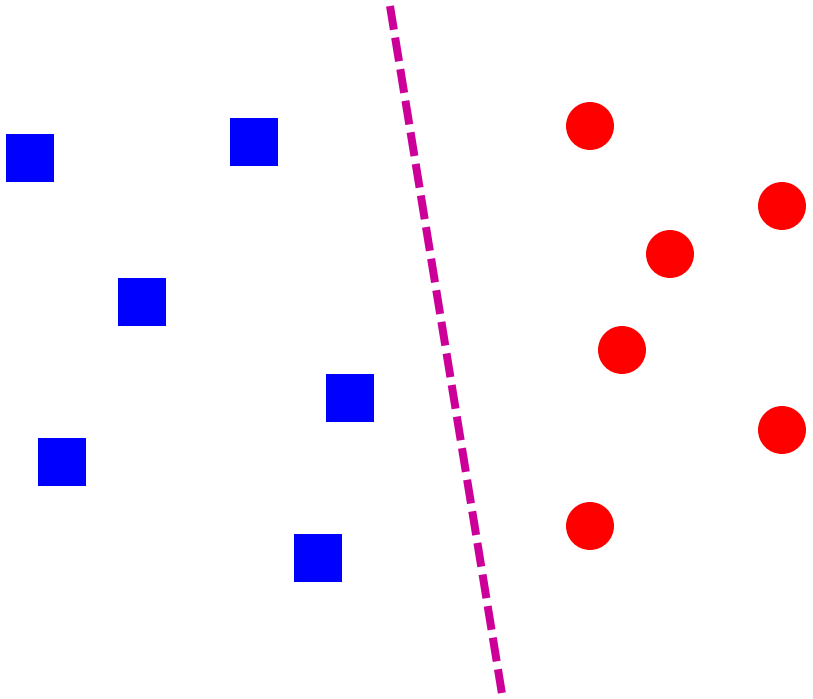
- All we need is a distance function for our inputs
- No training required!

K-nearest neighbors classifier

- For a new point, find the k closest points from training data
- Vote for class label with labels of the k points



Linear classifier



- Find a *linear function* to separate the classes

$$f(\mathbf{x}) = \text{sgn}(w_1 x_1 + w_2 x_2 + \ldots + w_D x_D) = \text{sgn}(\mathbf{w} \cdot \mathbf{x})$$

EXAM INFORMATION

Homework / Exercise Folder

- Submission Deadline – May 9, 2021
- Needs to be complete, i.e., include all exercises and homework, organized
- Should be obvious that you have attempted to completely solve the problem
  - Few lines of added code is not that

Exam Details

- Multiple choice
- Only move forward
- Questions & Answers may be randomized
  - Possible that questions will be different for different students
- Duration: 2 hours

Exam Code Clarification

- Any additional comments in the code, beyond what is typical and directly related to the code, is not allowed in the exam.
- Code may only be run for the two exam questions related to the homework/exercise, and the code should be corresponding to the description of the exercise/homework. You won't be asked anything beyond that.

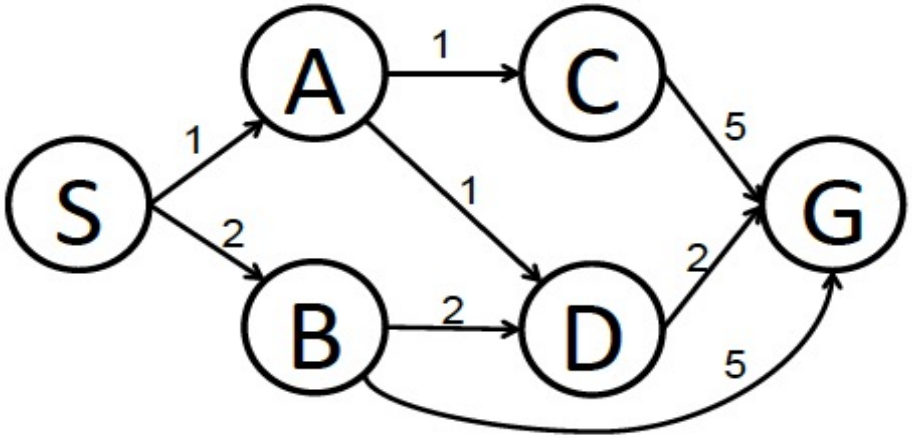
Example Exam Question

What kind of environment has a crossword puzzle?

A. Dynamic
*B. Static
C. None of the mentioned
D. Stochastic

Example Exam Question

Consider the search problem shown in the figure. S is the start-state, G is the (only) goal-state. Break any ties alphabetically. For the questions that ask for a path, please give your answers in the form `S - A - D - G .' (2 points)



What path would depth-first graph search return for this search problem?

A. S – A – D – G

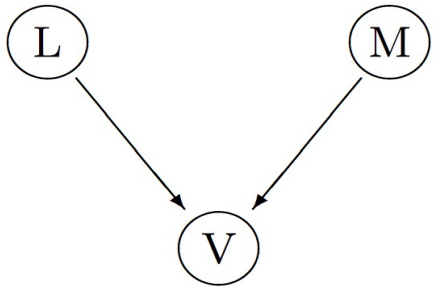
B. S – B – G

*C. S – A – C – G

D. S – B – D – G

Exam Question Example

The Bayesian Network LMV below has three nodes for boolean variables, L, M and V. The probabilities for L and M are: P (M = true) = 0.2 and P (L = true) = 0.7. The conditional probabilities for variable V are as shown in the table below. (3 points)



L	M	$P(V = true \mid L, M)$
true	true	0.9
true	false	0.5
false	true	0.3
false	false	0.05

Which value is closest to the value of P(V = false | L = false) ?

A. 0.3

B. 0.7

*C. 0.9

D. 0.1

Homework Example Exam Question

The homework question is related to: Exercise on Bayesian Networks

Using your homework from Exercise 7, calculate the conditional probability of the fuel tank leaking given that there is a high consumption.

If needed, the Bayesian network is shown on the next slide.

What is the conditional probability of the fuel tank leaking given that there is a high consumption? (4 points)

A. approximately 87 %

B. approximately 3 %

C. approximately 42 %

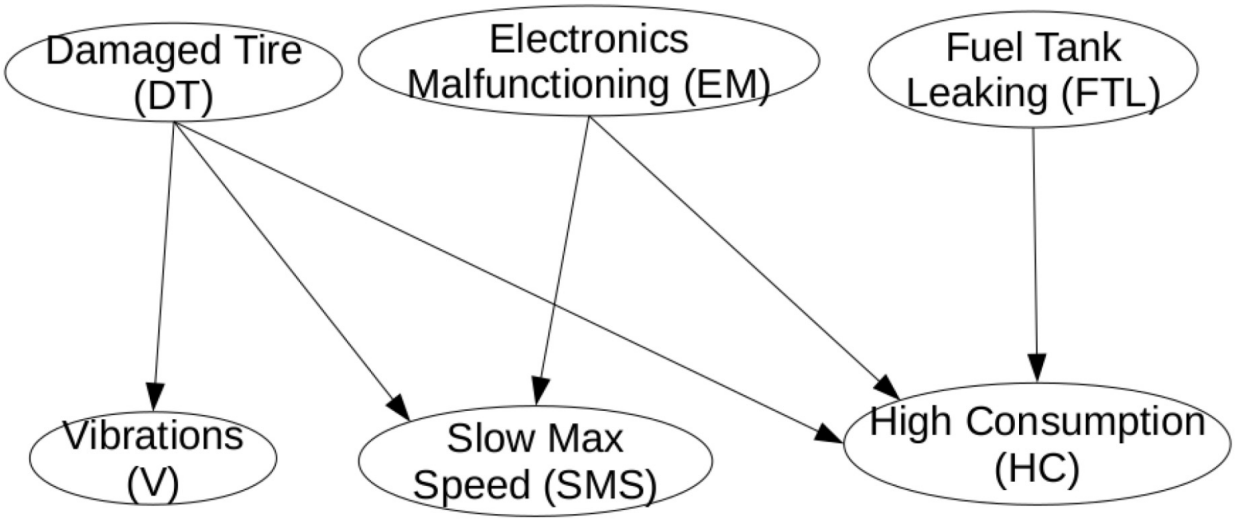
D. approximately 11 %

*E. approximately 62 %

DT	P(DT)
T	0.3
F	0.7

EM	P(EM)
T	0.3
F	0.7

FTL	P(FTL)
T	0.2
F	0.8



DT	P(V | DT)
T	0.7
F	0.1

DT	EM	P(SMS=T | DT,EM)
T	T	0.05
T	F	0.6
F	T	0.3
F	F	0.7

DT	FTL	EM	P(HC=T | DT,FTL,EM)
T	T	T	0.9
T	T	F	0.8
T	F	T	0.3
T	F	F	0.2
F	T	T	0.6
F	T	F	0.5
F	F	T	0.1
F	F	F	0.01


That's all Folks!