Supervised Learning

Learning from Examples

Asmaa Elbadrawy

PhD, Lecturer

IFT Program, ASU

Arizona State University

In **Supervised Learning**, an agent reads in data in the form **(input, output)** pairs. It learns a **model** (function) that maps input to output.

Formulating the Learning Problem

- **Given:** a dataset that contains data points, described by some attributes x1, x2, … xn.

- Each data point has:
  - *Some attributes*, referred to as *X*
  - **A class label** , referred to as *y*

x1	x2	y

attributes | class labels

Formulating the Learning Problem

- **Required:** Find a function that can predict the class label **y** of any data point given its attribute values **X**.

$$y = f(X)$$

– That is, estimate **y** as a function of **x**.

We know y & X.
We want to find $f()$

x1	x2	y

attributes   class labels

Formulating the Learning Problem

$$y = f(X)$$

There are many options for *f()*.

f() can be linear or non-linear.
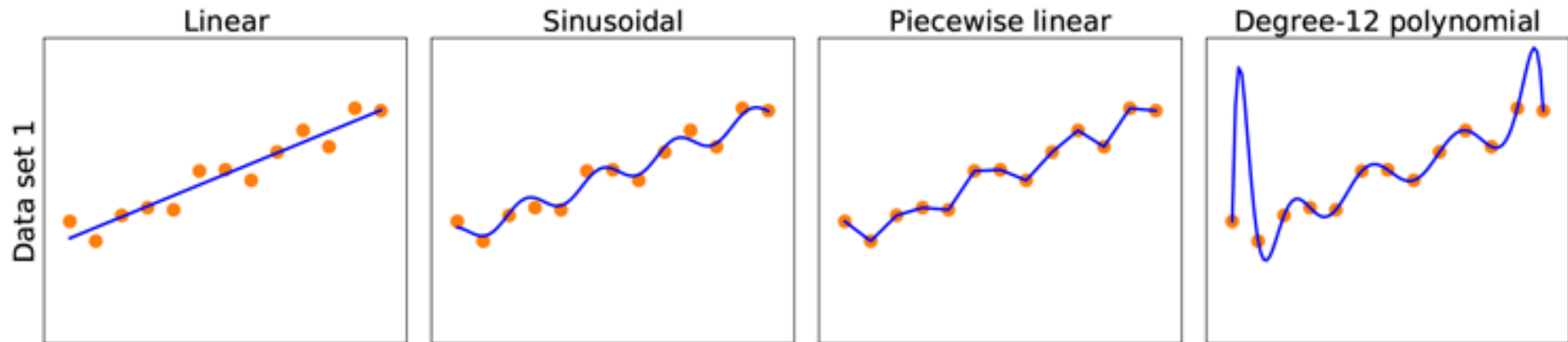
Fig 19.1, Russell & Norvig's Textbook

Example Applications

Object Classification in Images

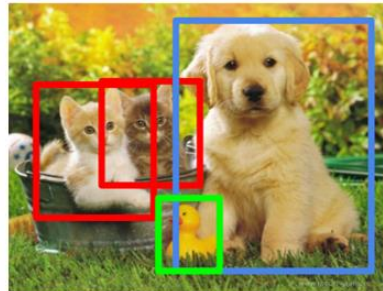- Given a dataset of images, classify the objects in the images within a set of ***predefined*** classes.

Image source: https://towardsdatascience.com/going-deep-into-object-detection-bed442d92b34
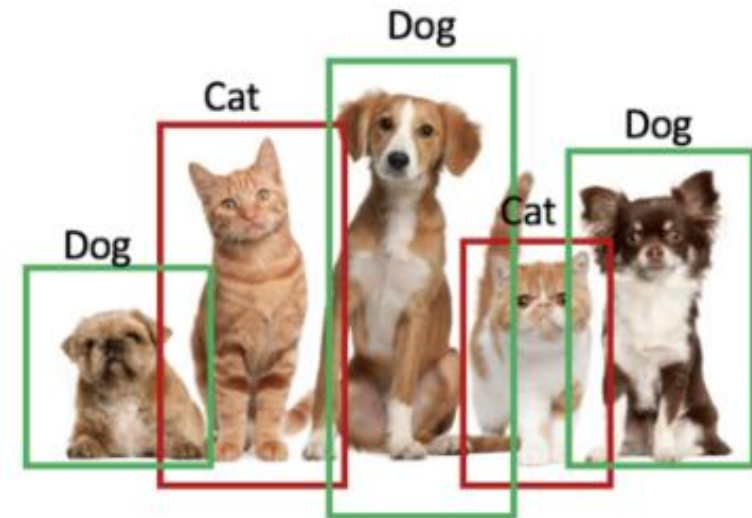
Image source: https://medium.com/@bluebirb/object-detection-current-popular-techniques-900d36ae9b08

Object Classification in Images

- Features used for cat/dog classification?
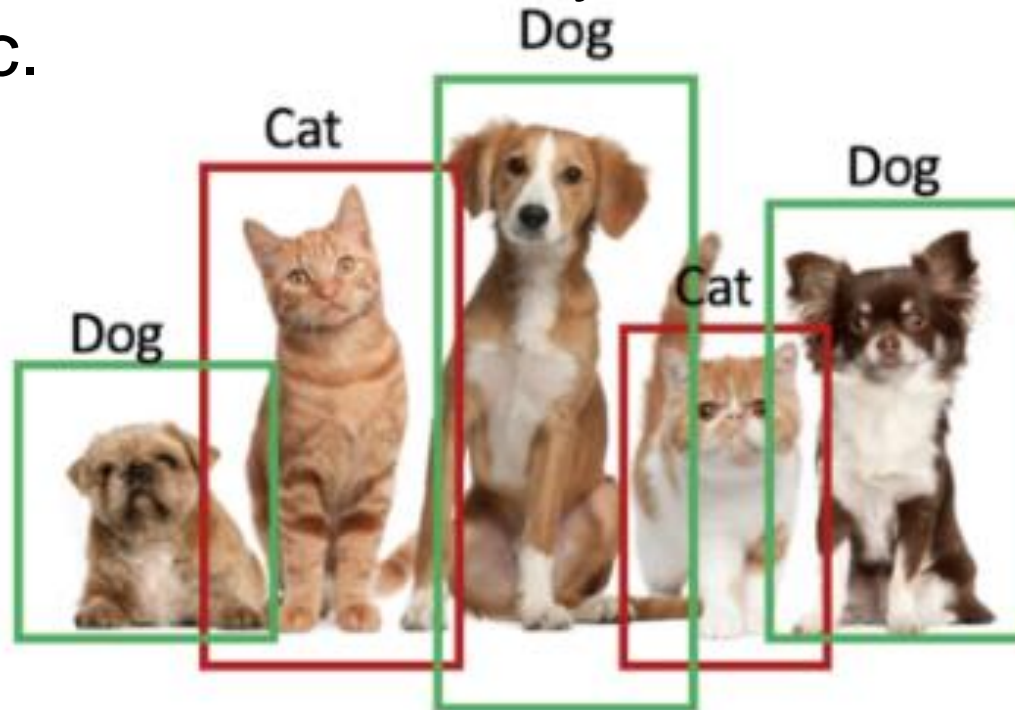 - Ear shape, face size w.r.t. body size, distance between nose and eyes, etc.

Tumor Cell Classification

- Given a dataset of images, each containing one or more cells, classify whether they are malignant or benign.
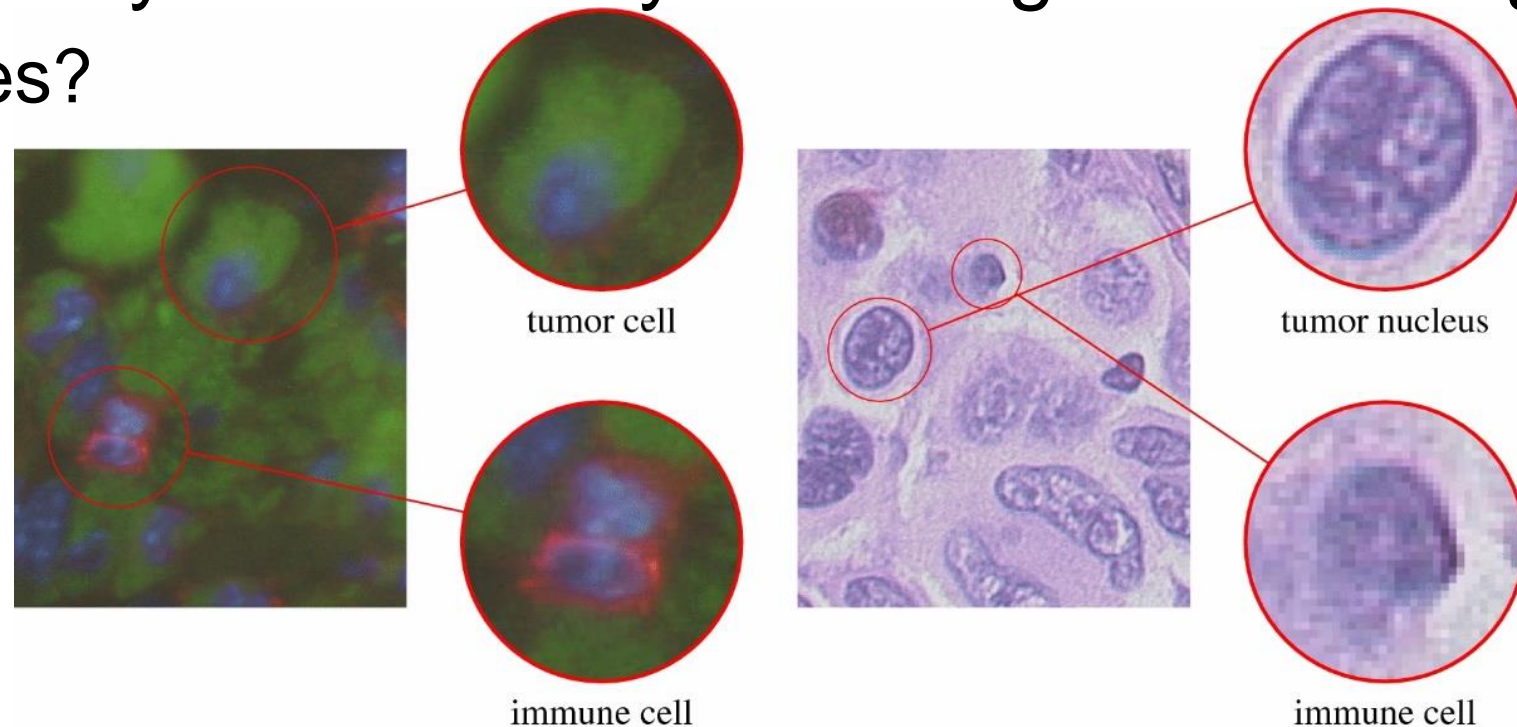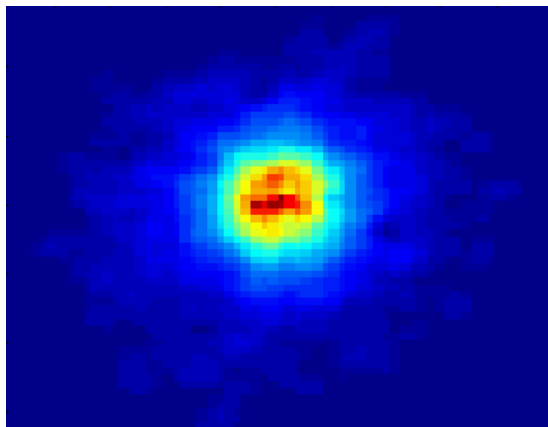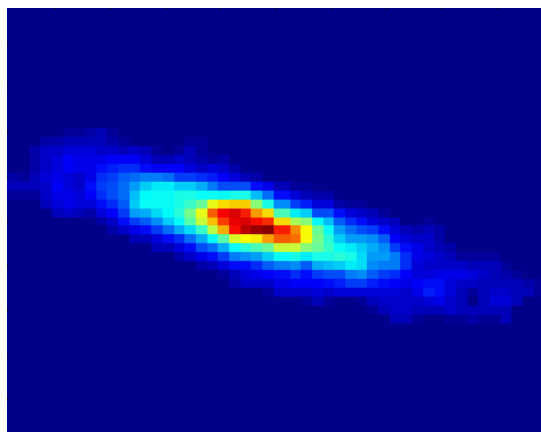 - Features?

tumor cell

immune cell

tumor nucleus

immune cell

Image source: https://link.springer.com/article/10.1007/s11548-019-01919-z
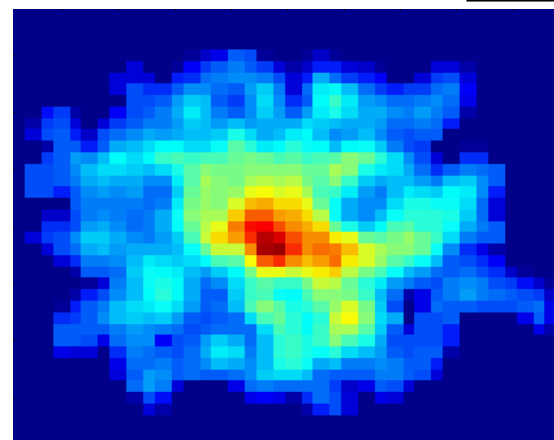
Galaxy Classification

Early

Intermediate

Late

Class:
• Stages of Formation

Attributes:
• Image features,
• Characteristics of light waves received, etc.

Courtesy: http://aps.umn.edu

Data Size:
• 72 million stars, 20 million galaxies
• Object Catalog: 9 GB
• Image Database: 150 GB

Predicting Stock prices given market conditions.

The Restaurant Waiting Problem

- Should we wait for a table at a restaurant?
- y ➜ Boolean variable deciding whether wait = True/False
- X ➜ 10 different attributes

Th

- S[...]
- y[...]lse
- X[...]

1. *Alternate*: whether there is a suitable alternative restaurant nearby.

2. *Bar*: whether the restaurant has a comfortable bar area to wait in.

3. *Fri/Sat*: true on Fridays and Saturdays.

4. *Hungry*: whether we are hungry right now.

5. *Patrons*: how many people are in the restaurant (values are *None, Some,* and *Full*).

6. *Price*: the restaurant's price range ($, $$, $$$).

7. *Raining:* whether it is raining outside.

8. *Reservation:* whether we made a reservation.

9. *Type:* the kind of restaurant (French, Italian, Thai, or burger).

10. *WaitEstimate:* host's wait estimate: 0–10, 10–30, 30–60, or >60 minutes.

The Restaurant Waiting Problem

- There are thousands of possible values for the 10 attributes.

Example	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	WillWait
x_1	Yes	No	No	Yes	Some	$$$	No	Yes	French	0–10	y_1 = Yes
x_2	Yes	No	No	Yes	Full	$	No	No	Thai	30–60	y_2 = No
x_3	No	Yes	No	No	Some	$	No	No	Burger	0–10	y_3 = Yes
x_4	Yes	No	Yes	Yes	Full	$	Yes	No	Thai	10–30	y_4 = Yes
x_5	Yes	No	Yes	No	Full	$$$	No	Yes	French	>60	y_5 = No
x_6	No	Yes	No	Yes	Some	$$	Yes	Yes	Italian	0–10	y_6 = Yes
x_7	No	Yes	No	No	None	$	Yes	No	Burger	0–10	y_7 = No
x_8	No	No	No	Yes	Some	$$	Yes	Yes	Thai	0–10	y_8 = Yes
x_9	No	Yes	Yes	No	Full	$	Yes	No	Burger	>60	y_9 = No
x_{10}	Yes	Yes	Yes	Yes	Full	$$$	No	Yes	Italian	10–30	y_{10} = No
x_{11}	No	No	No	No	None	$	No	No	Thai	0–10	y_{11} = No
x_{12}	Yes	Yes	Yes	Yes	Full	$	No	No	Burger	30–60	y_{12} = Yes

Fig 19.2, Russell & Norvig's Textbook