

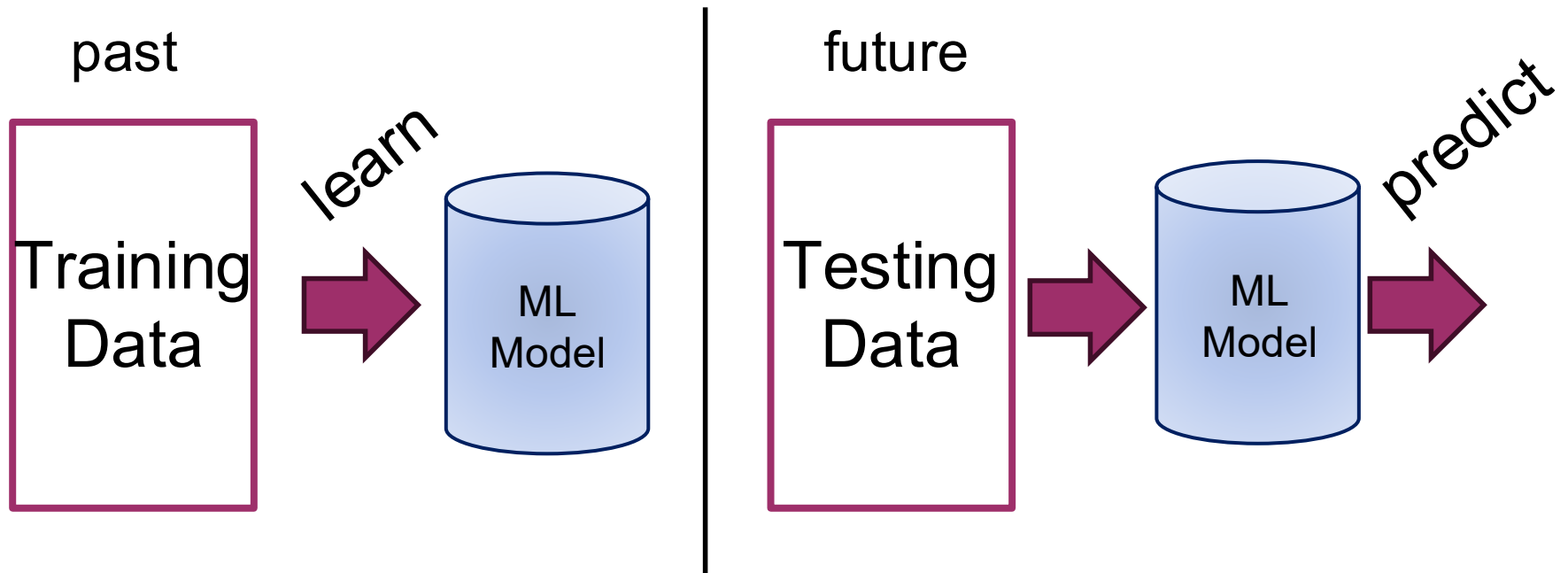
Lecture 12: Intro to Machine Learning

CS 181E

Fall 2026

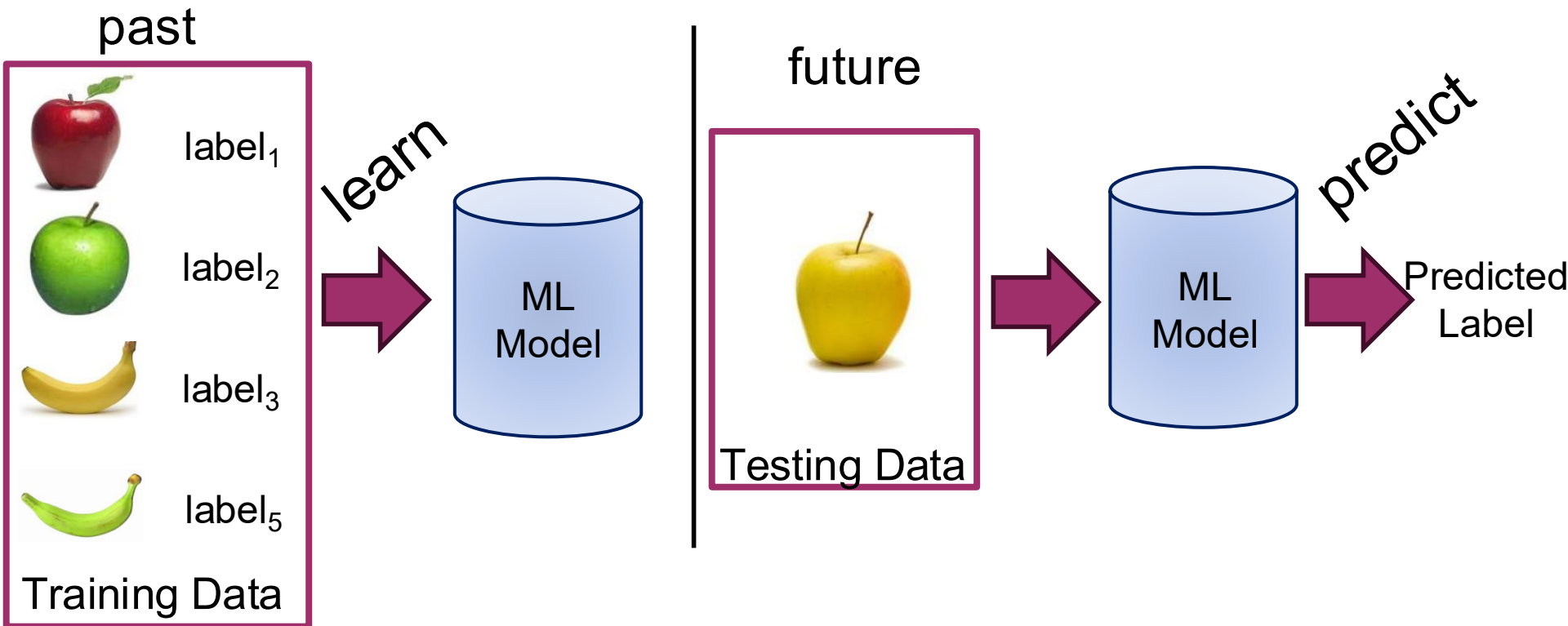
Machine Learning is...

Machine learning is about predicting the future based on the past.



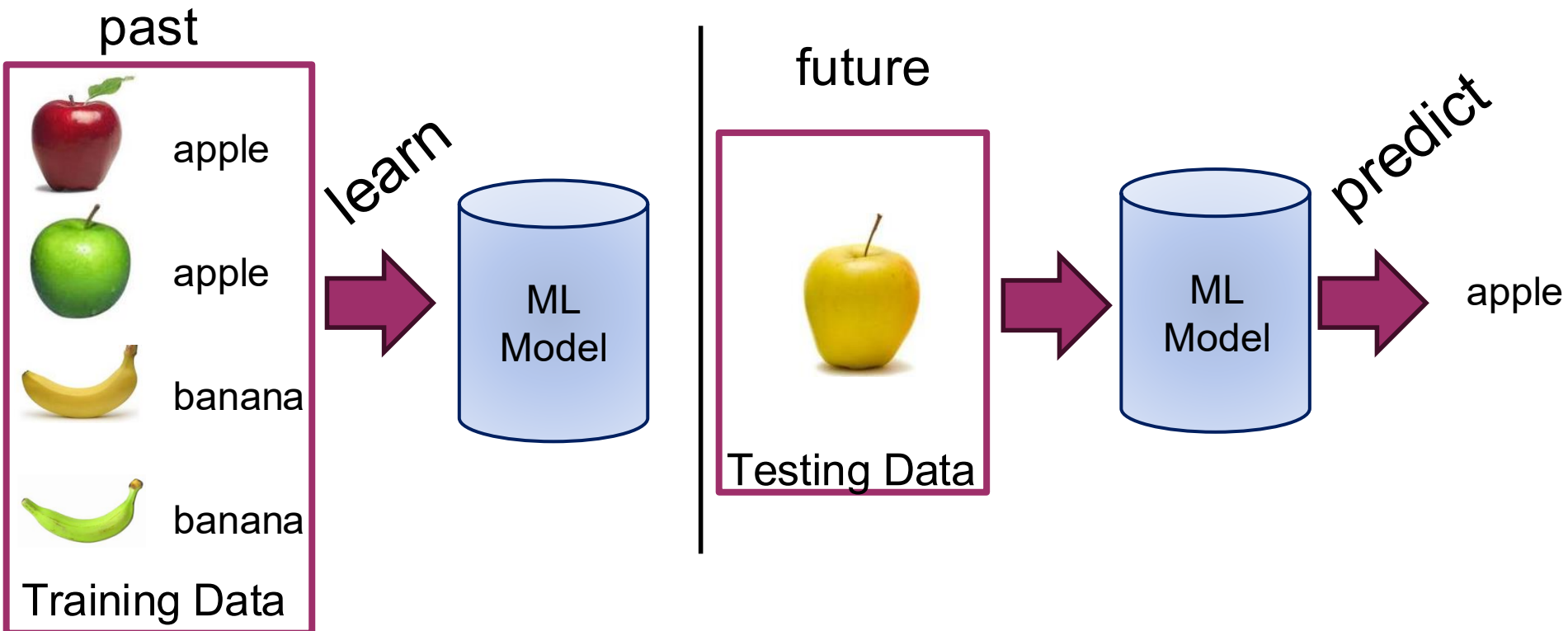
Type #1: Supervised Learning

- training data is labeled with correct answers



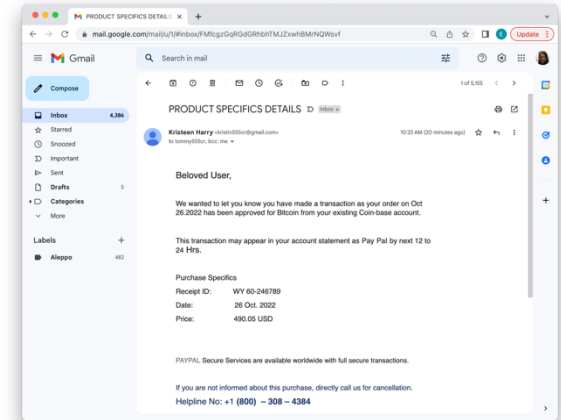
Type #1a: Supervised Classification

- training data is labeled with correct answers
- there is a finite, pre-defined set of possible labels
- ML task is to predict which of these labels applies



Examples of Classification

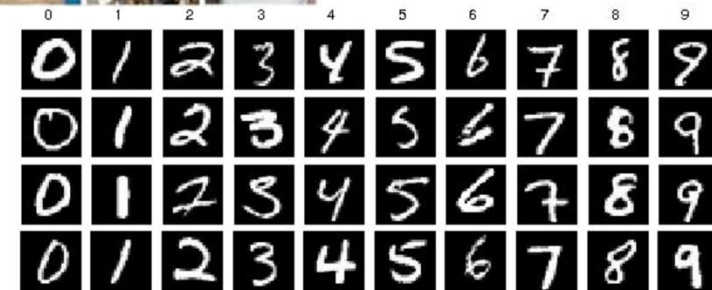
- Spam filters



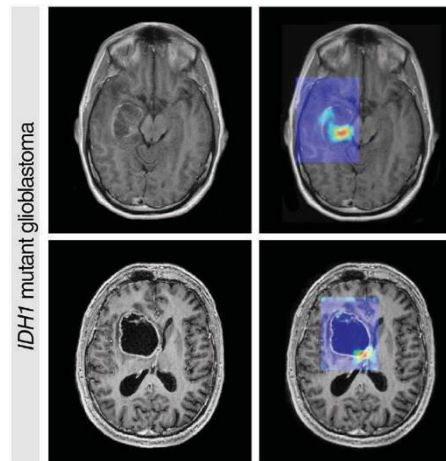
- Facial recognition



- Character recognition

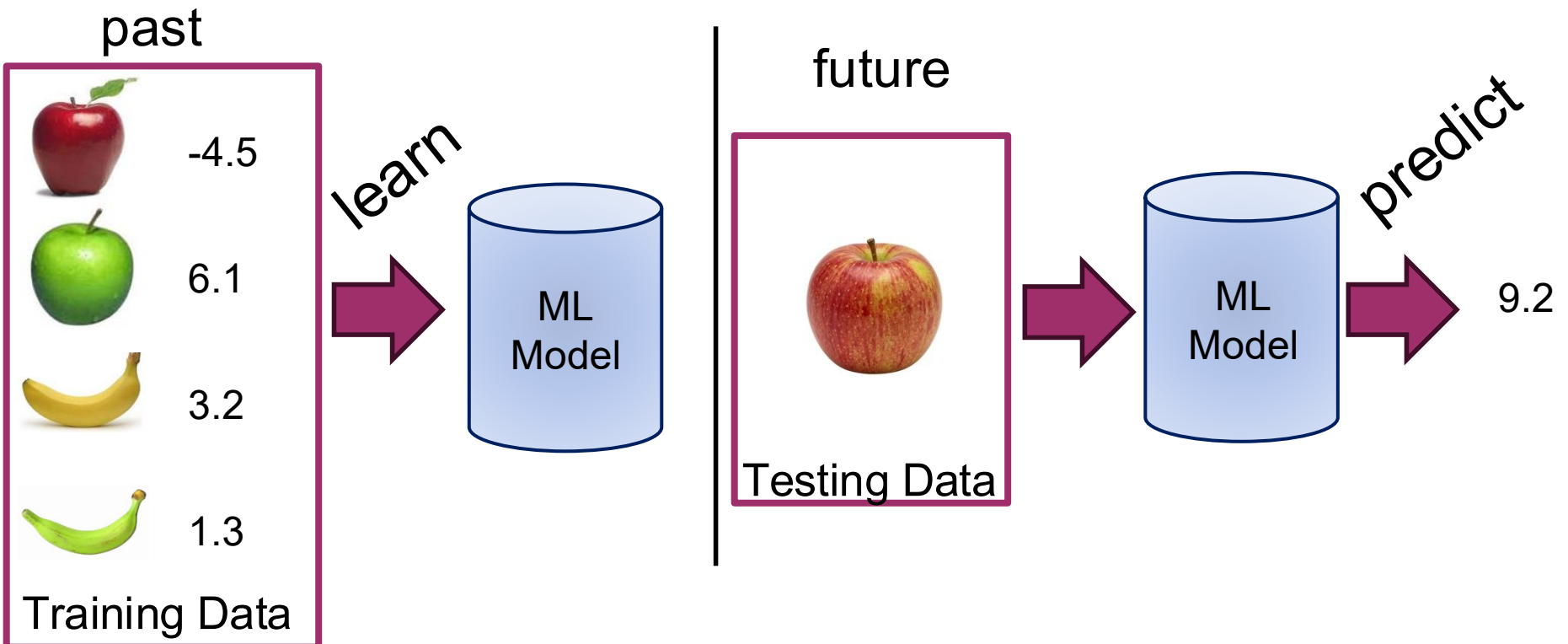


- Medical diagnosis



Type #1b: Supervised Regression

- training data is labeled with correct answers
- labels are real numbers
- ML task is to predict the right number

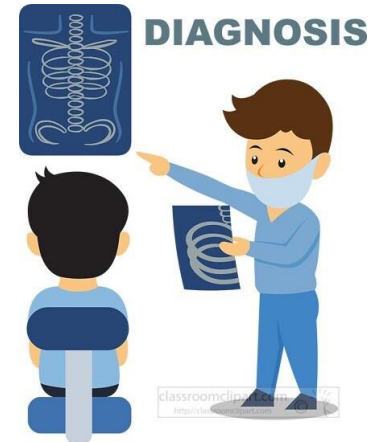


Examples of Regressions

- Online Advertising: predict the likelihood a user will click on an ad



- Epidemiology: predict likelihood/severity of a disease



- Car/Plane Navigation: predict angle of steering wheel, acceleration, etc

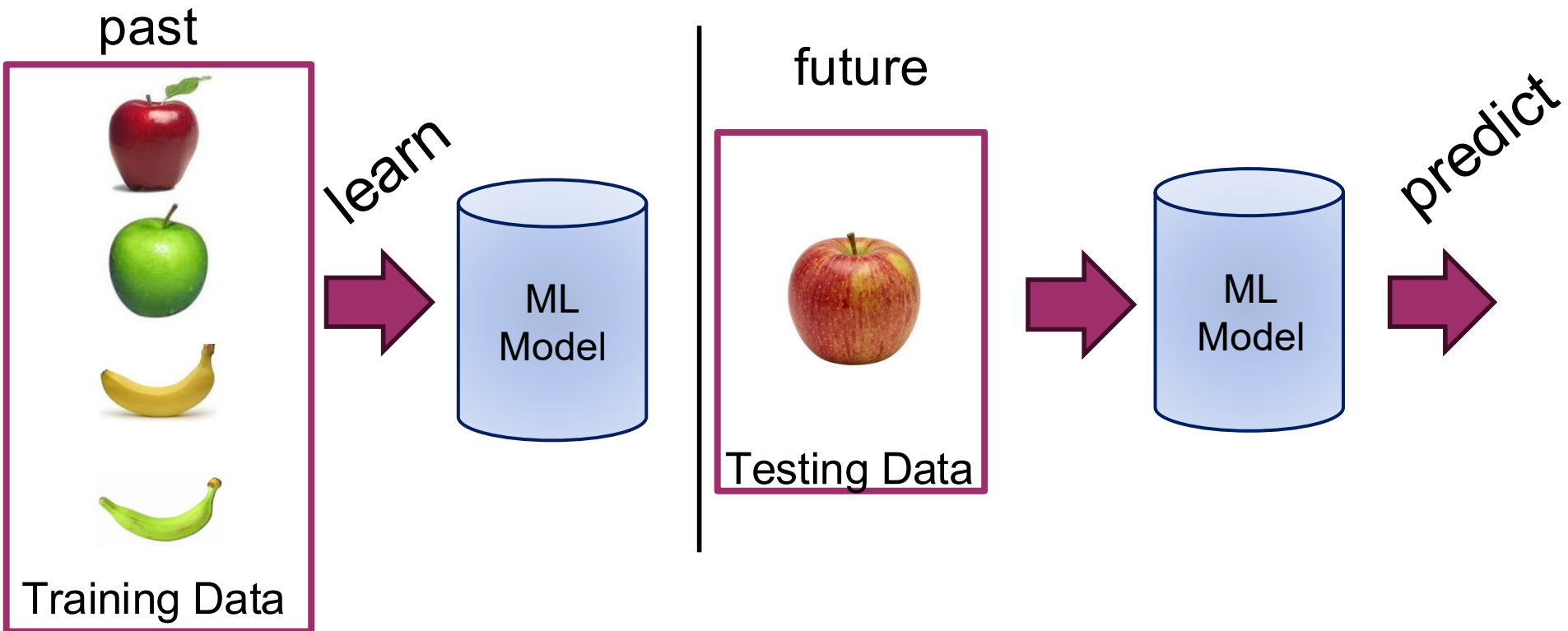


- Weather: predict temperature, likelihood of rain, etc



Type #2: Unsupervised Learning

- training data is unlabeled
- goal: discover patterns or structures in the data
- Examples: clustering, principal component analysis



Examples of Unsupervised Learning

- Clustering: Learn clusters/groups without any label
 - customer segmentation: which users “look-alike” in terms of buying habits
 - anomaly detection: which events are “normal” vs which are unusual
- Principal Component Analysis (PCA): reduce number of dimensions while keeping most important information
 - network intrusion detection: efficiently detect malicious network traffic
 - image compression: reduce image size (useful for optimizing facial recognition and image analysis pipelines)

Other learning variations

- What data is available:
 - Supervised, unsupervised
 - reinforcement learning, semi-supervised, active learning, ...
- How do we get the data:
 - online vs. offline learning
- Type of model:
 - generative vs. discriminative
 - parametric vs. non-parametric

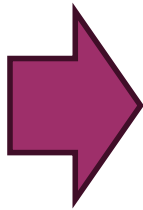
Example Supervised Learning Methods

Method	Classif./Regress.	Gen./Discr.	Param/Non
Naïve Bayes	Classification	Generative	Parametric
Linear Regression	Regression	Discriminative	Parametric
Logistic Regression	Classification	Discriminative	Parametric
Neural Networks	Both	Depends	Parametric
K-Nearest Neighbor	Classification	Generative	Nonparametric
Support Vector Machines	Both	Discriminative	Nonparametric

REPRESENTING DATA

Features

training data



features

$f_1, f_2, f_3, \dots, f_n$

$f_1, f_2, f_3, \dots, f_n$

$f_1, f_2, f_3, \dots, f_n$

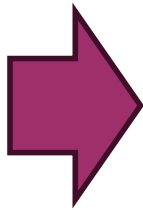
$f_1, f_2, f_3, \dots, f_n$

How our machine learning algorithms actually “view” the data

Features

training data

features



red, round, leaf, 3oz, ...

green, round, no leaf, 4oz, ...

yellow, curved, no leaf, 4oz, ...

green, curved, no leaf, 5oz, ...

How our machine learning algorithms actually “view” the data

Provided features

the data has been collected and the *relevant* features have been identified

- This is common in many domains (e.g. medicine, social science, etc.)
- we cannot collect more features from the examples (at least “core” features)

	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Age	Gender	Major	Year_of_Stud	AI_Usage_Fre	AI_Tools_Use	Primary_AI_P	Perceived_In	Study_Hours	Study_Hours	AI_Ethical_C	Stress_Level	Overall_Satis	Future_AI_Ado
2	22	Male	Computer Sc	Junior	Daily	GitHub Copil	Coding	Very Positive	25	20	Yes	Decreased	9	Very Likely
3	24	Female	Arts	Postgraduate	Sometimes	MidJourney	Creative proj	Positive	18	17	Yes	No change	8	Likely
4	20	Male	Business	Sophomore	Often	ChatGPT	Writing	Positive	30	22	Yes	Decreased	8	Very Likely
5	21	Female	Psychology	Junior	Often	ChatGPT	Research	Positive	28	25	Unsure	Increased	7	Likely
6	19	Male	Engineering	Sophomore	Daily	Multiple	Coding	Very Positive	35	28	Unsure	Decreased	10	Very Likely
7	23	Female	Medicine	Senior	Sometimes	Grammarly	Writing	Neutral	40	40	Yes	No change	6	Likely
8	26	Male	Law	Postgraduate	Often	Multiple	Research	Positive	32	29	Yes	Decreased	8	Very Likely

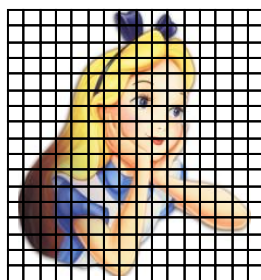
Raw data vs. features

In other domains, we are provided with the raw data, but must extract/identify features

For example

- image data
- text data
- audio data
- log data
- ...

Raw data

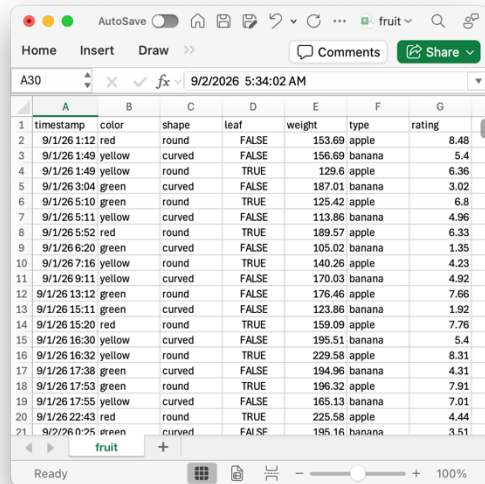


Clinton said banana
repeatedly last week on tv,
“banana, banana, banana”

Features

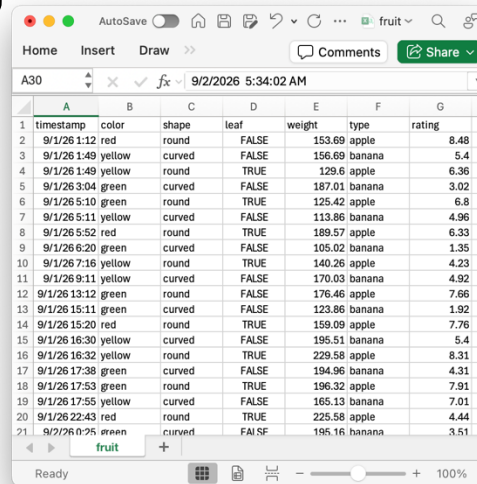
- pixels OR
- “patches”
- texture features
- shape features
- “bag of words”
- frequency
- bigram counts
- ...

Pre-processing



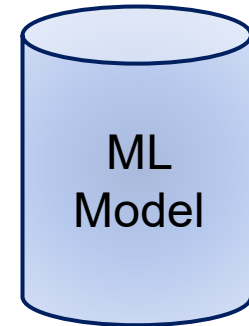
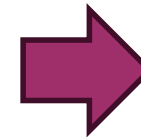
	A	B	C	D	E	F	G
1	timestamp	color	shape	leaf	weight	type	rating
2	9/1/26 1:12	red	round	FALSE	153.69	apple	8.48
3	9/1/26 1:49	yellow	curved	FALSE	156.69	banana	5.4
4	9/1/26 1:49	yellow	round	TRUE	129.6	apple	6.36
5	9/1/26 3:04	green	curved	FALSE	187.01	banana	3.02
6	9/1/26 5:10	green	round	TRUE	125.42	apple	6.8
7	9/1/26 5:11	yellow	curved	FALSE	113.86	banana	4.96
8	9/1/26 5:52	red	round	TRUE	189.57	apple	6.33
9	9/1/26 6:20	green	curved	FALSE	105.02	banana	1.35
10	9/1/26 7:16	yellow	round	TRUE	140.26	apple	4.23
11	9/1/26 9:11	yellow	curved	FALSE	170.03	banana	4.92
12	9/1/26 13:12	green	round	FALSE	176.46	apple	7.66
13	9/1/26 15:11	green	curved	FALSE	123.86	banana	1.92
14	9/1/26 15:20	red	round	TRUE	159.09	apple	7.76
15	9/1/26 16:30	yellow	curved	FALSE	195.51	banana	5.4
16	9/1/26 16:32	yellow	round	TRUE	229.58	apple	8.31
17	9/1/26 17:38	green	curved	FALSE	194.96	banana	4.31
18	9/1/26 17:53	green	round	TRUE	196.32	apple	7.91
19	9/1/26 17:55	yellow	curved	FALSE	165.13	banana	7.01
20	9/1/26 22:43	red	round	TRUE	225.58	apple	4.44
21	9/2/26 0:25	green	curved	FALSE	195.16	banana	3.51

pre-process



	A	B	C	D	E	F	G
1	timestamp	color	shape	leaf	weight	type	rating
2	9/1/26 1:12	red	round	FALSE	153.69	apple	8.48
3	9/1/26 1:49	yellow	curved	FALSE	156.69	banana	5.4
4	9/1/26 1:49	yellow	round	TRUE	129.6	apple	6.36
5	9/1/26 3:04	green	curved	FALSE	187.01	banana	3.02
6	9/1/26 5:10	green	round	TRUE	125.42	apple	6.8
7	9/1/26 5:11	yellow	curved	FALSE	113.86	banana	4.96
8	9/1/26 5:52	red	round	TRUE	189.57	apple	6.33
9	9/1/26 6:20	green	curved	FALSE	105.02	banana	1.35
10	9/1/26 7:16	yellow	round	TRUE	140.26	apple	4.23
11	9/1/26 9:11	yellow	curved	FALSE	170.03	banana	4.92
12	9/1/26 13:12	green	round	FALSE	176.46	apple	7.66
13	9/1/26 15:11	green	curved	FALSE	123.86	banana	1.92
14	9/1/26 15:20	red	round	TRUE	159.09	apple	7.76
15	9/1/26 16:30	yellow	curved	FALSE	195.51	banana	5.4
16	9/1/26 16:32	yellow	round	TRUE	229.58	apple	8.31
17	9/1/26 17:38	green	curved	FALSE	194.96	banana	4.31
18	9/1/26 17:53	green	round	TRUE	196.32	apple	7.91
19	9/1/26 17:55	yellow	curved	FALSE	165.13	banana	7.01
20	9/1/26 22:43	red	round	TRUE	225.58	apple	4.44
21	9/2/26 0:25	green	curved	FALSE	195.16	banana	3.51

learn



1. Encode data
 - hot-one encoding
 - dummy encoding

1. Throw out outlier examples
2. Remove noisy features
3. Pick “good” features
4. Normalize feature values
 - center data
 - scale data

Example: Pre-processing w/ Pandas

```
import pandas

# read csv file into pandas dataframe
data = pandas.read_csv("fruit.csv")

# ways to delete columns
data = data.drop(columns=["timestamp", "rating"])
type = data.pop("type")

# one-hot encode
cat_cols = ["color", "shape"]
encoded_data = pandas.get_dummies(data, columns=cat_cols)

# add column
encoded_data["type"] = type

# write pandas dataframe to csv file
encoded_data.to_csv("fruit-encoded.csv")
```

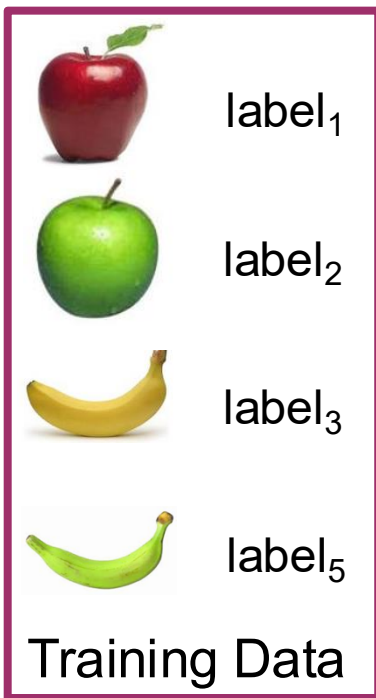
Exercise: Pre-processing Data

- Pre-process the file `example.csv` to make the following changes:
 - drop the `Student_ID` column and the columns about ethical concerns, overall satisfaction, and future AI intentions
 - one hot or dummy encode the remaining categorical columns (except the one about stress level)

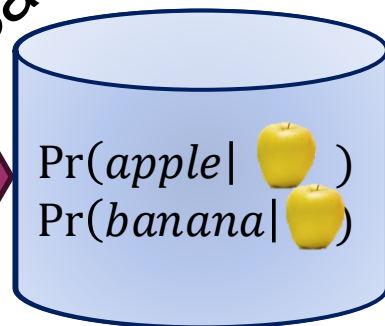
Naïve Bayes

- **Idea:** for each possible label, calculate probability that testing data would have that label, pick label with highest probability conditioned on that data
- training = figuring out those probabilities

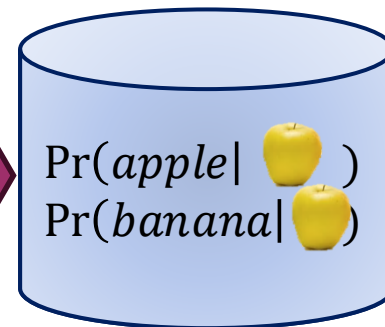
past



learn



future



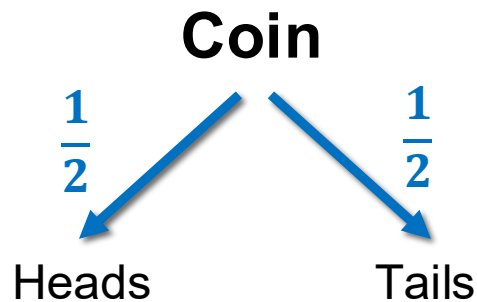
predict



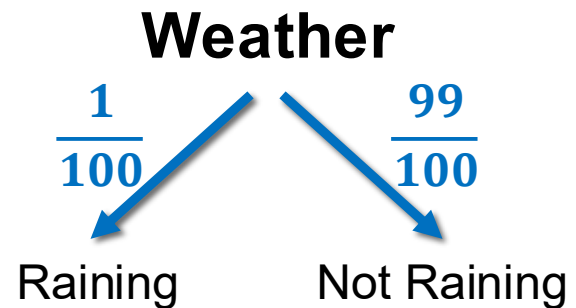
Pick
argmax

Review: Probability

- Probability is the math of describing how likely something is
- A **probability** is a number in the range $[0,1]$ that describes how likely something is.



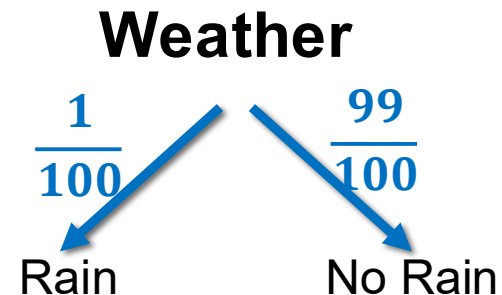
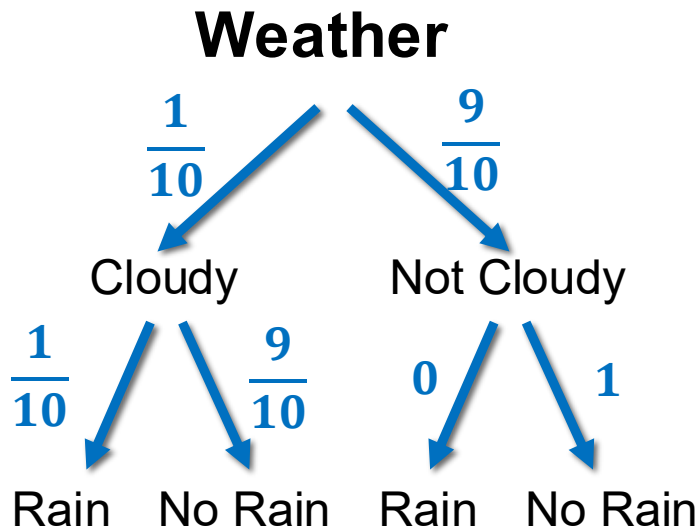
$$\Pr[\textit{coin} = \textit{Heads}] = \frac{1}{2}$$
$$\Pr[\textit{coin} = \textit{Tails}] = \frac{1}{2}$$



$$\Pr[\textit{weather} = \textit{Raining}] = \frac{1}{100}$$
$$\Pr[\textit{weather} = \textit{Not Raining}] = \frac{99}{100}$$

Review: Combining Probabilities

- In general, the probability of both of two events A and B occurring is $\Pr(A \text{ and } B) = \Pr(A) \cdot \Pr(B|A)$



$$\Pr(\text{Cloudy and No Rain}) = \Pr(\text{Cloudy}) \cdot \Pr(\text{No Rain}|\text{Cloudy}) = \frac{1}{10} \cdot \frac{9}{10} = \frac{9}{100}$$

Bayes Rule

$$\begin{aligned}\Pr(\text{data and label}) &= \Pr(\text{data}) \cdot \Pr(\text{label} | \text{data}) \\ &= \Pr(\text{label}) \cdot \Pr(\text{data} | \text{label})\end{aligned}$$

$$\Pr(\text{data}) \cdot \Pr(\text{label} | \text{data}) = \Pr(\text{label}) \cdot \Pr(\text{data} | \text{label})$$

$$\Pr(\text{label} | \text{data}) = \frac{\Pr(\text{label}) \cdot \Pr(\text{data} | \text{label})}{\Pr(\text{data})}$$

$$\Pr(\text{apple} | \text{🍏}) = \frac{\Pr(\text{apple}) \cdot \Pr(\text{🍏} | \text{apple})}{\Pr(\text{🍏})}$$

$$\Pr(\text{banana} | \text{🍏}) = \frac{\Pr(\text{banana}) \cdot \Pr(\text{🍏} | \text{banana})}{\Pr(\text{🍏})}$$

} Pick
argmax

$$\begin{aligned}&\Pr(\text{apple}) \cdot \Pr(\text{🍏} | \text{apple}) \\ &\Pr(\text{banana}) \cdot \Pr(\text{🍏} | \text{banana})\end{aligned} \quad \left. \vphantom{\begin{aligned} &\Pr(\text{apple}) \cdot \Pr(\text{🍏} | \text{apple}) \\ &\Pr(\text{banana}) \cdot \Pr(\text{🍏} | \text{banana}) \end{aligned}} \right\} \text{Pick argmax}$$

Naïve Bayes

$$\begin{aligned}\Pr(\textit{label}) \cdot \Pr(\text{🍏} | \textit{label}) &= \Pr(\textit{label}) \cdot \Pr(f_1, \dots, f_n | \textit{label}) \\ &= \Pr(\textit{label}) \cdot \prod_i \Pr(f_i | \textit{label})\end{aligned}$$

$$\begin{aligned}\Pr(\textit{apple}) \cdot \Pr(\text{🍏} | \textit{apple}) &= \Pr(\textit{apple}) \cdot \prod_i \Pr(f_i | \textit{apple}) \\ \Pr(\textit{banana}) \cdot \Pr(\text{🍏} | \textit{banana}) &= \Pr(\textit{banana}) \cdot \prod_i \Pr(f_i | \textit{banana})\end{aligned}$$

Pick
argmax

Estimating Probabilities

training
data

features

labels



red, leaf

apple



green, no leaf

apple



yellow, no leaf

banana



green, no leaf

banana

$$\Pr(\text{apple}) = \frac{2}{4}$$

$$\Pr(\text{banana}) = \frac{2}{4}$$

$$\Pr(f_1 = \text{red}|\text{apple}) = \frac{1}{2}$$

$$\Pr(f_1 = \text{green}|\text{apple}) = \frac{1}{2}$$

$$\Pr(f_1 = \text{yellow}|\text{apple}) = \frac{0}{2}$$

$$\Pr(f_1 = \text{red}|\text{banana}) = \frac{0}{2}$$

$$\Pr(f_1 = \text{green}|\text{banana}) = \frac{1}{2}$$

$$\Pr(f_1 = \text{yellow}|\text{banana}) = \frac{1}{2}$$

Making Predictions



red, no leaf

$$\begin{aligned} & \Pr(apple) \cdot \prod_i \Pr(f_i|apple) \\ &= \Pr(apple) \cdot \Pr(red|apple) \cdot \Pr(no\ leaf|apple) \\ &= \frac{2}{4} \cdot \frac{1}{2} \cdot \frac{1}{2} \\ &= \frac{1}{8} \end{aligned}$$



We'll go with apple

$$\begin{aligned} & \Pr(banana) \cdot \prod_i \Pr(f_i|banana) \\ &= \Pr(banana) \cdot \Pr(red|banana) \cdot \Pr(no\ leaf|banana) \\ &= \frac{2}{4} \cdot \frac{0}{2} \cdot \frac{2}{2} \\ &= 0 \end{aligned}$$

Making Predictions



yellow, no leaf

$$\begin{aligned} & \Pr(\text{apple}) \cdot \prod_i \Pr(f_i | \text{apple}) \\ &= \Pr(\text{apple}) \cdot \Pr(\text{yellow} | \text{apple}) \cdot \Pr(\text{no leaf} | \text{apple}) \\ &= \frac{2}{4} \cdot \frac{0}{2} \cdot \frac{1}{2} \\ &= 0 \end{aligned}$$

$$\begin{aligned} & \Pr(\text{banana}) \cdot \prod_i \Pr(f_i | \text{banana}) \\ &= \Pr(\text{banana}) \cdot \Pr(\text{yellow} | \text{banana}) \cdot \Pr(\text{no leaf} | \text{banana}) \\ &= \frac{2}{4} \cdot \frac{1}{2} \cdot \frac{2}{2} \\ &= \frac{1}{4} \end{aligned}$$



We'll go with banana

Exercise: Naïve Bayes

$$\Pr(\text{label}) \cdot \prod_i \Pr(f_i | \text{label})$$

	f_1, f_2	label
	red, curved	apple
	green, round	apple
	green, curved	banana
	yellow, curved	banana
	yellow, round	apple
	red, round.	apple
	yellow, round.	banana
	yellow, round	apple

Given the training set on the left:

1. Evaluate the likelihood that a new value (green, curved) is an apple
2. Evaluate the likelihood that a new value (green, curved) is a banana.
3. Evaluate the likelihood that a new value (green, round) is an apple.
4. Evaluate the likelihood that a new value (green, round) is a banana

Example: Training a NB model

```
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB, BernoulliNB

# define your training data and target labels
data = pandas.read_csv("fruit-encoded.csv")
labels = data.pop("type")

# split training + labels into training & test data
X_train, X_test, y_train, y_test = train_test_split(data, labels,
                                                    test_size=0.3, random_state=47)

# train model
model = BernoulliNB()
model.fit(X_train, y_train)
```



BernoulliNB for binary training data
GaussianNB for real-valued training data

Exercise: Training a NB model

- Train a Naïve Bayes model to predict whether student stress increased, decreased, or stayed the same post-AI.