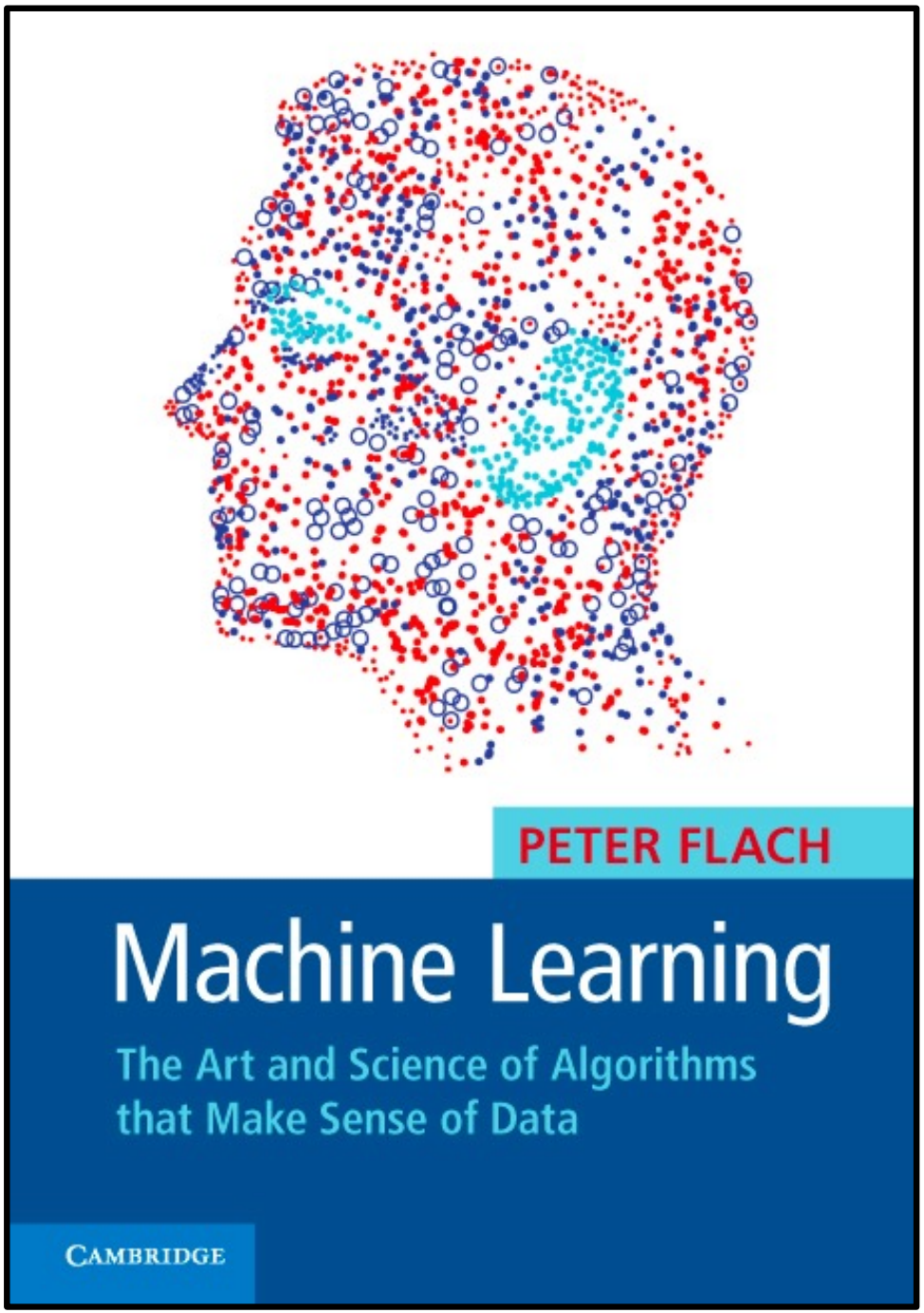
6.S079 MACHINE LEARNING 1

MARCH 7, 2024
MIKE CAFARELLA

THANKS TO TIM KRASKA FOR SLIDES



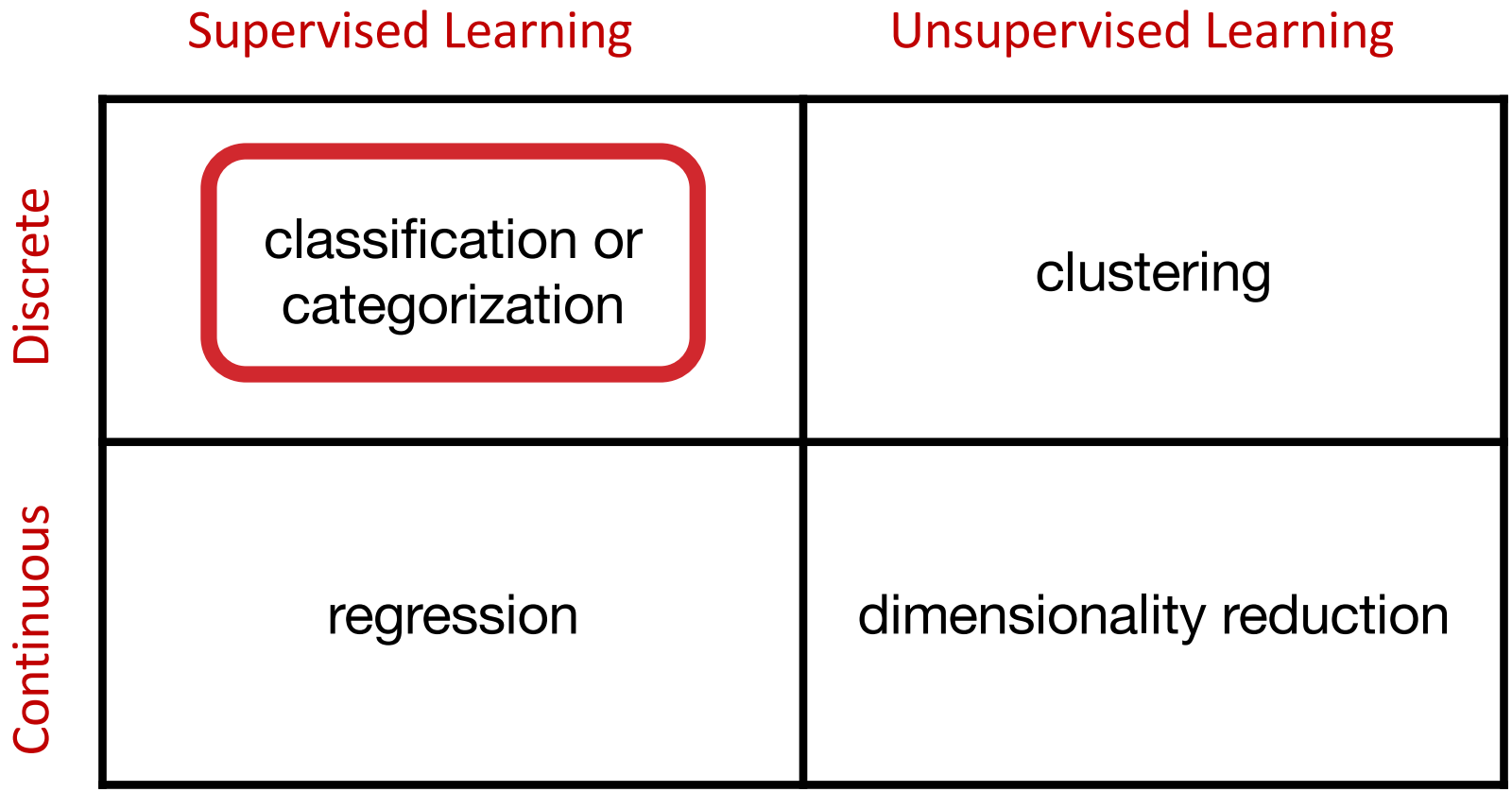
PDF available via MIT Library online

Readings posted on class site

MACHINE LEARNING PROBLEMS

(Boosted-) Decision Trees

K-Means
Agglomerative clustering
DBScan



(Boosted-) Decision Trees

PCA

WHAT IS A CLASSIFIER?

Apply a prediction function to a feature representation of an image/data-set to get the desired output:

f() = “apple”

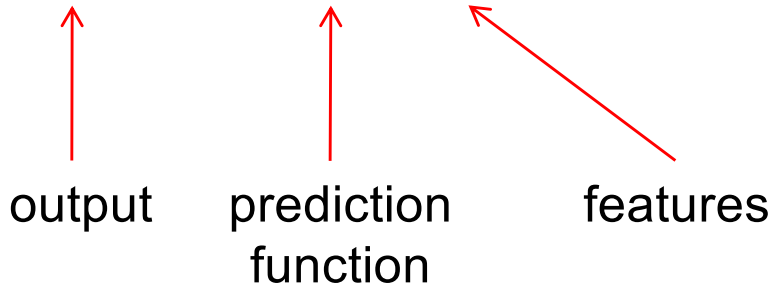
f() = “tomato”

f() = “cow”

Slide credit: L. Lazebnik

THE MACHINE LEARNING FRAMEWORK

$$y = f(x)$$

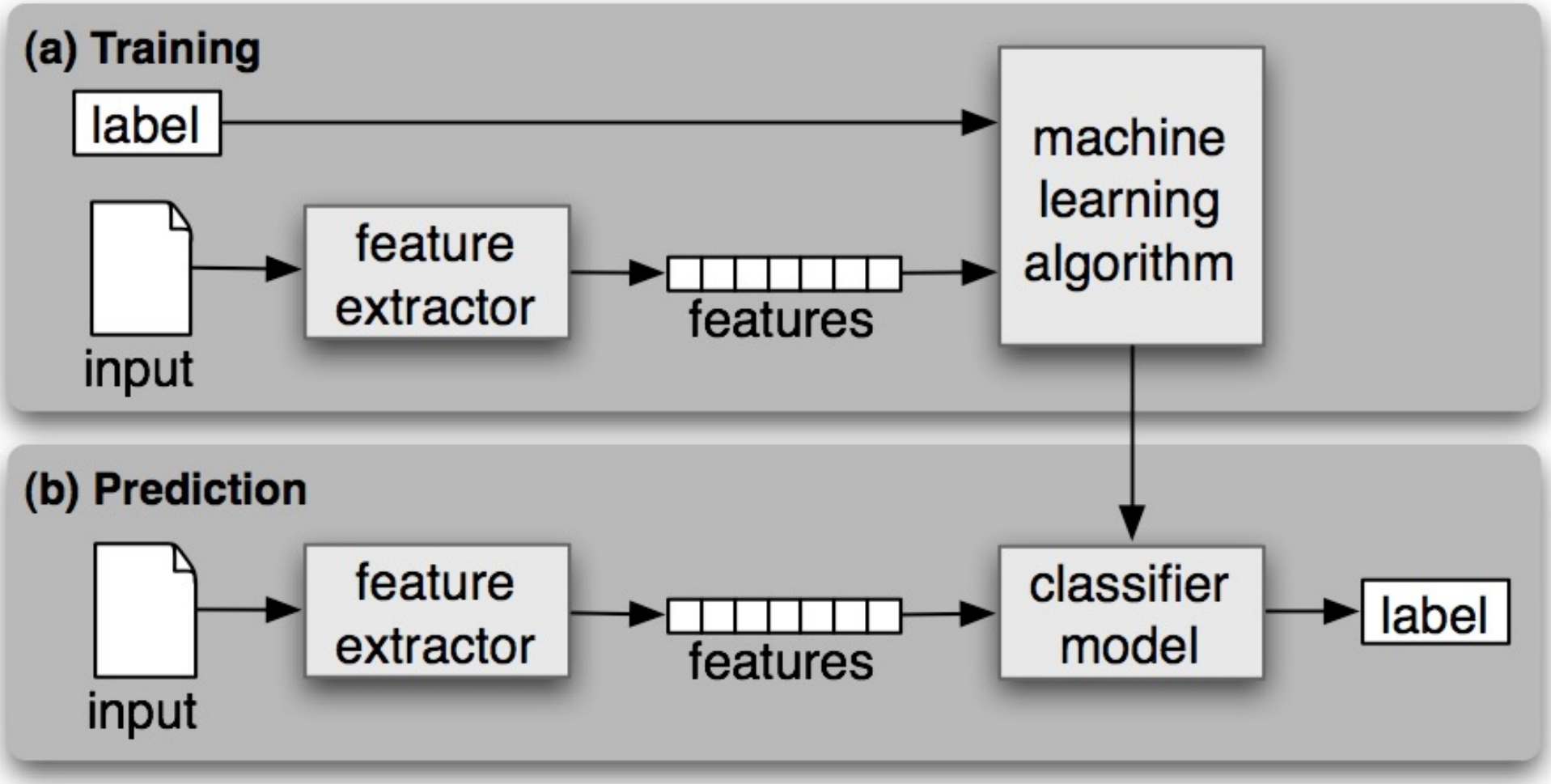


Training: given a *training set* of labeled examples $\{(x_1,y_1), \ldots, (x_N,y_N)\}$, estimate the prediction function f by minimizing the prediction error on the training set

Testing: apply f to a never before seen *test example* x and output the predicted value $y = f(x)$

Slide credit: L. Lazebnik

ML PIPELINE (SUPERVISED)



MANY CLASSIFIERS TO CHOOSE FROM

K-nearest neighbor

Support Vector Machines

Decision Trees

Random Forest

(Gradient) Boosted Decision Trees

Logistic Regression

Naïve Bayes

Bayesian network

RBMs

….

Which is the best one?

Slide credit: D. Hoiem

MANY CLASSIFIERS TO CHOOSE FROM

K-nearest neighbor

Support Vector Machines

Decision Trees

Random Forest

(Gradient) Boosted Decision Trees

Logistic Regression

Naïve Bayes

Bayesian network

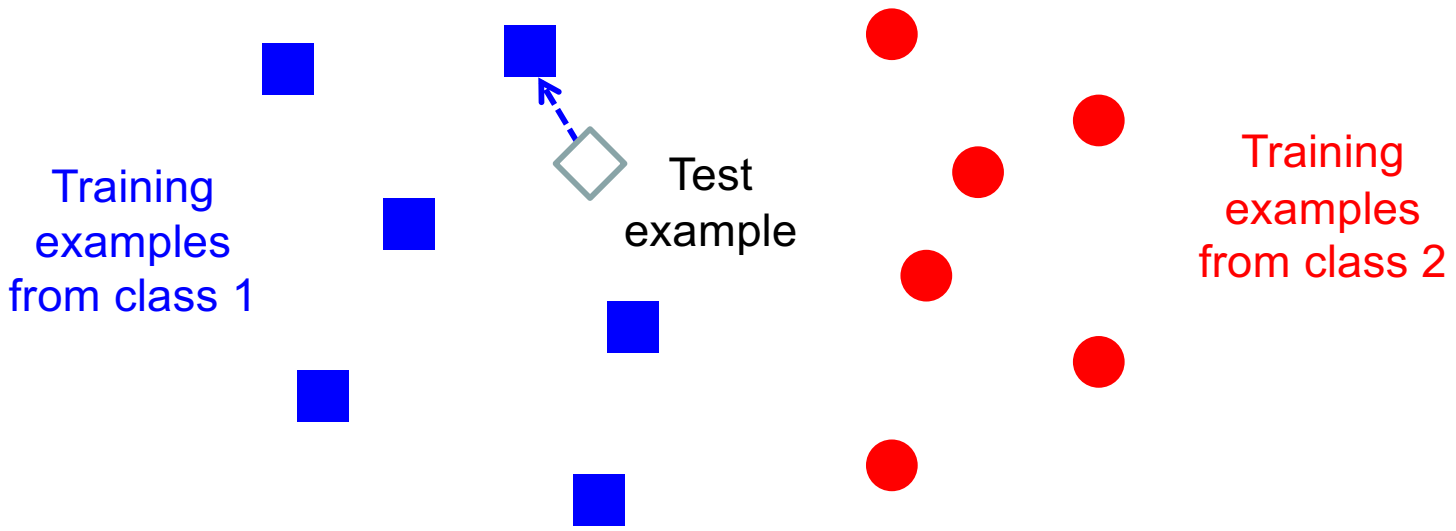
RBMs

….

Which is the best one?

Slide credit: D. Hoiem

CLASSIFIERS: NEAREST NEIGHBOR

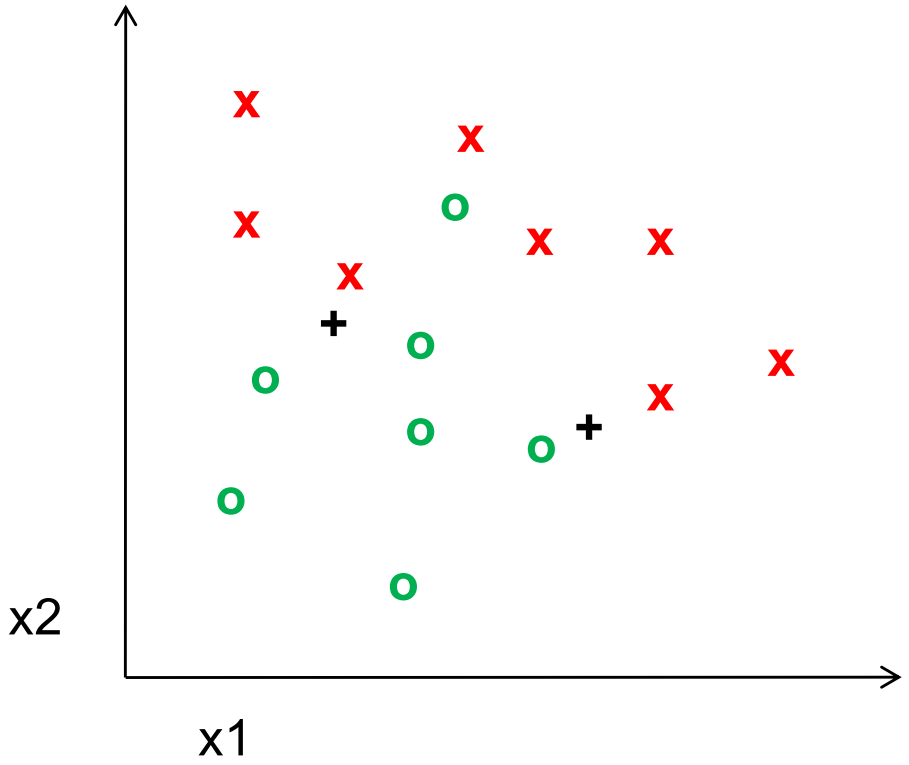


f(**x**) = label of the training example nearest to **x**

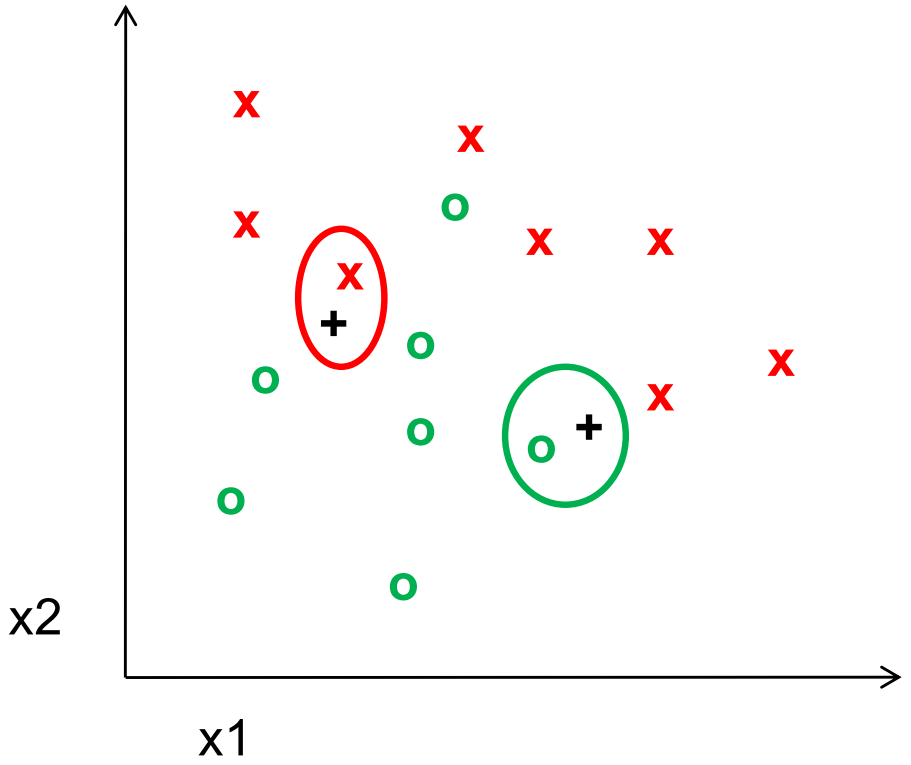
- All we need is a distance function for our inputs
- No training required!

Slide credit: L. Lazebnik

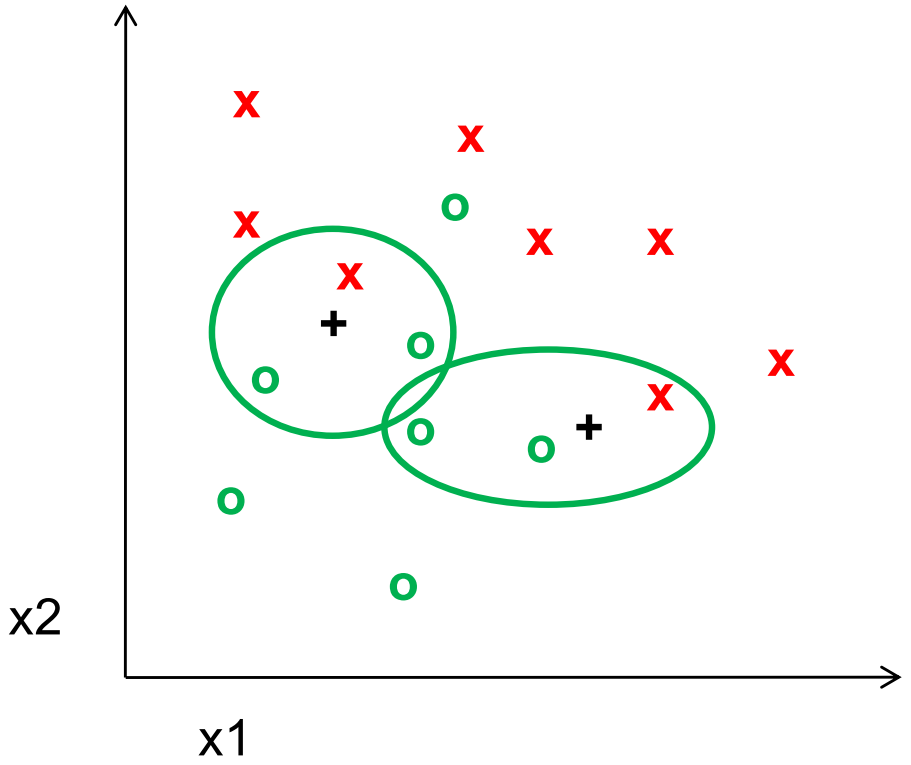
K-NEAREST NEIGHBOR



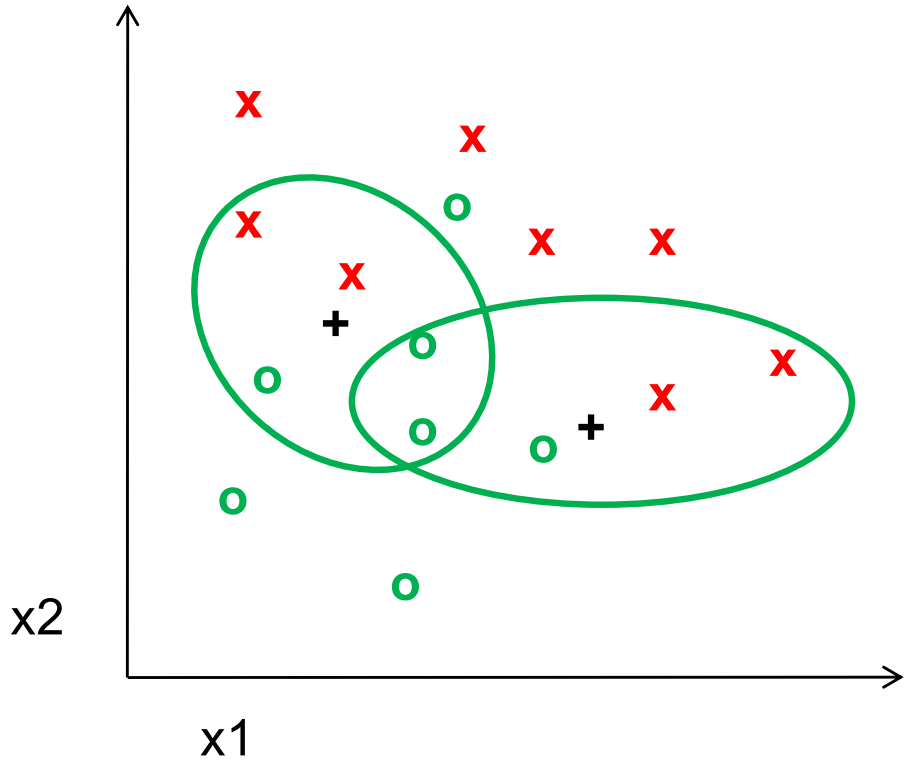
1-NEAREST NEIGHBOR



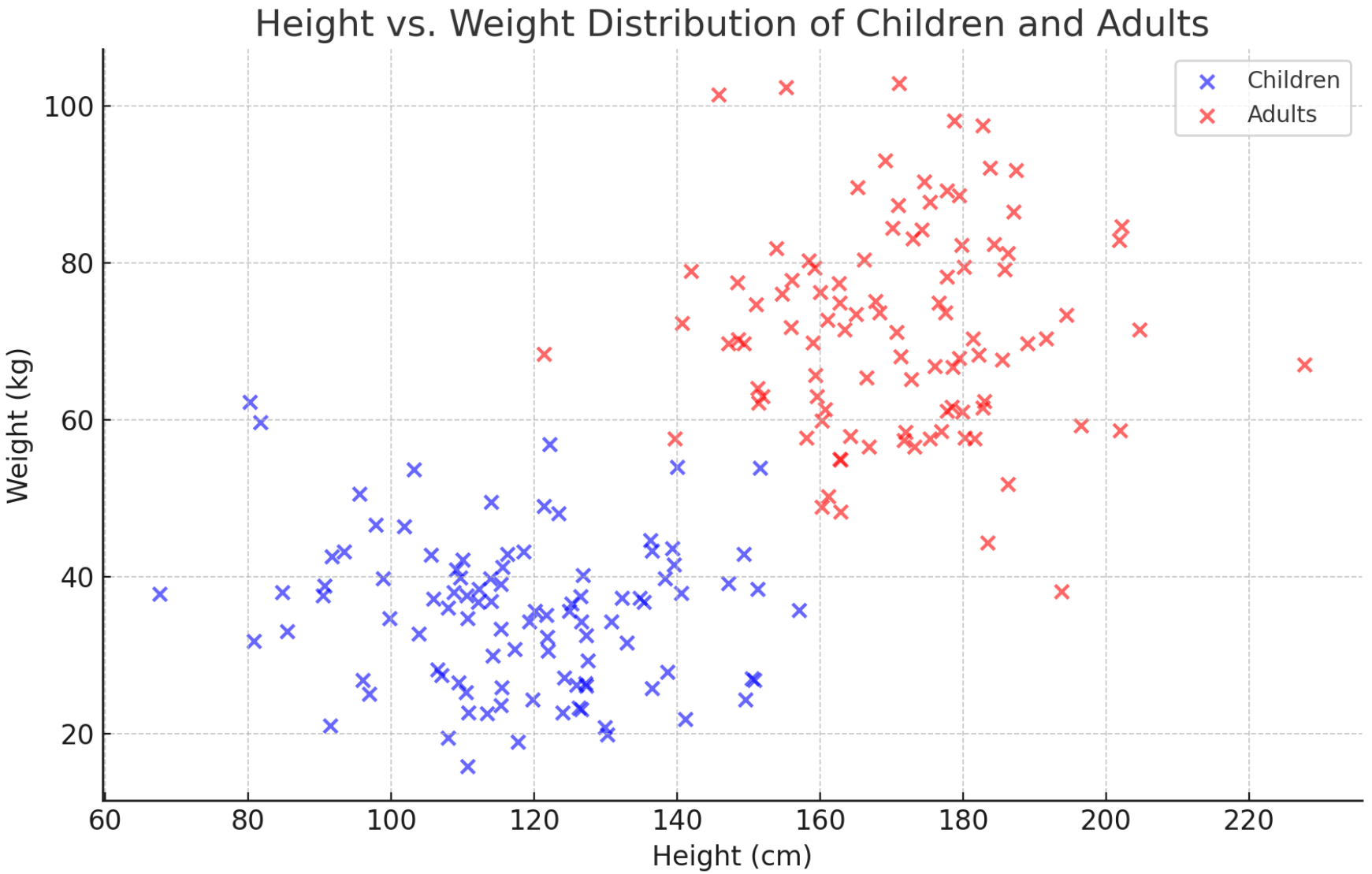
3-NEAREST NEIGHBOR



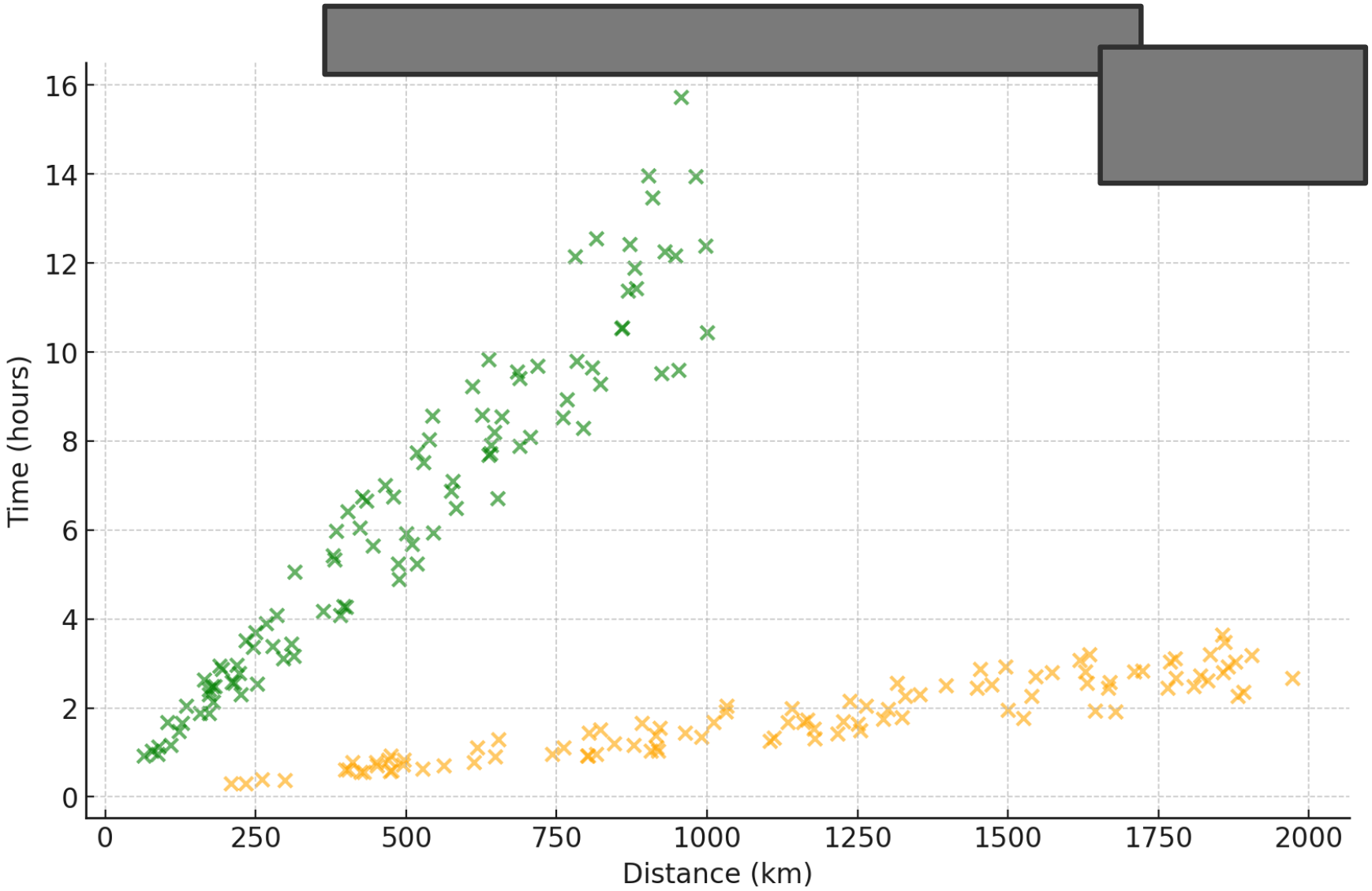
5-NEAREST NEIGHBOR



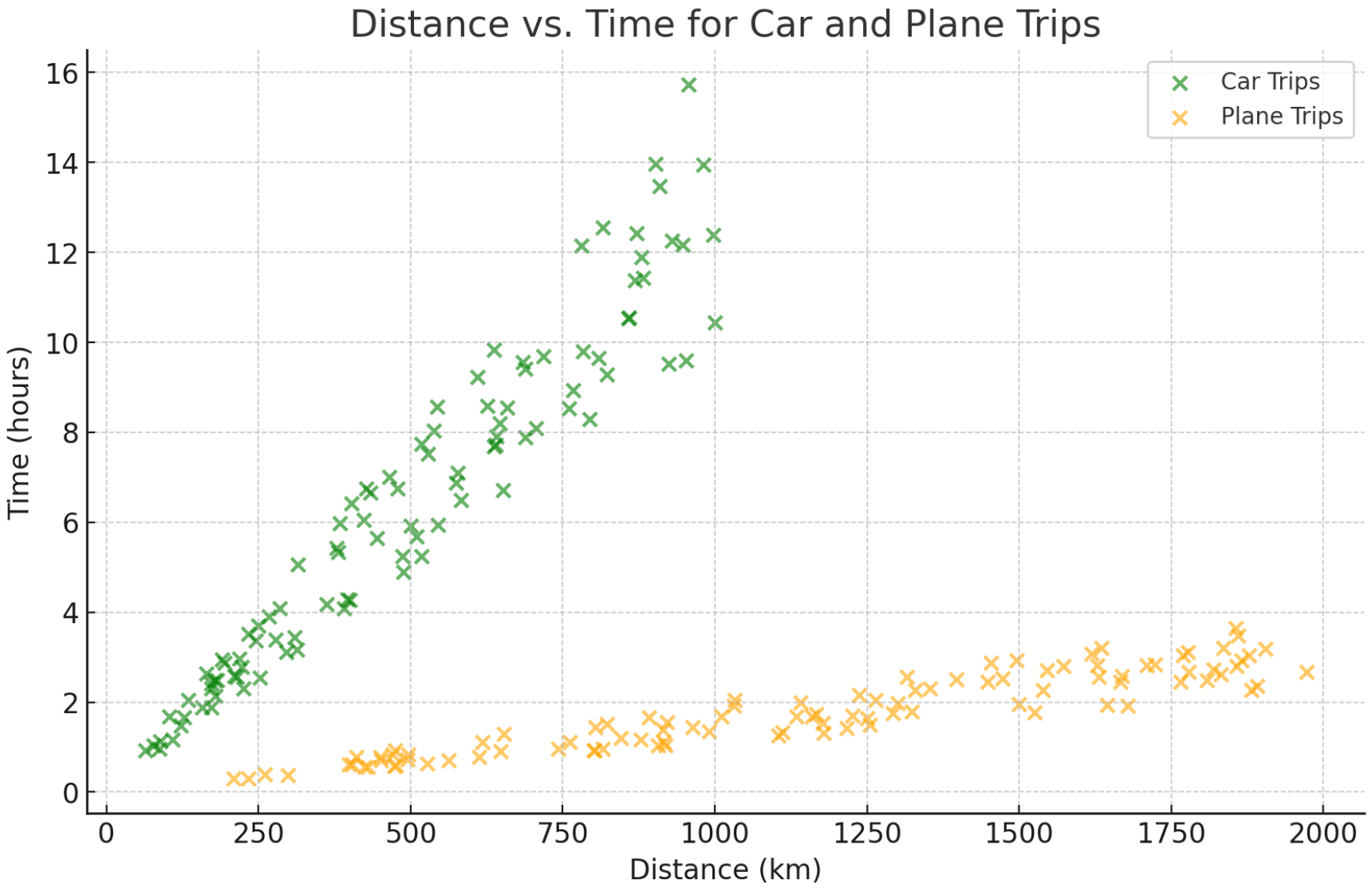
EXAMPLES


Height vs. Weight Distribution of Children and Adults
Children
Adults
Weight (kg)
Height (cm)
20
40
60
80
100
60
80
100
120
140
160
180
200
220

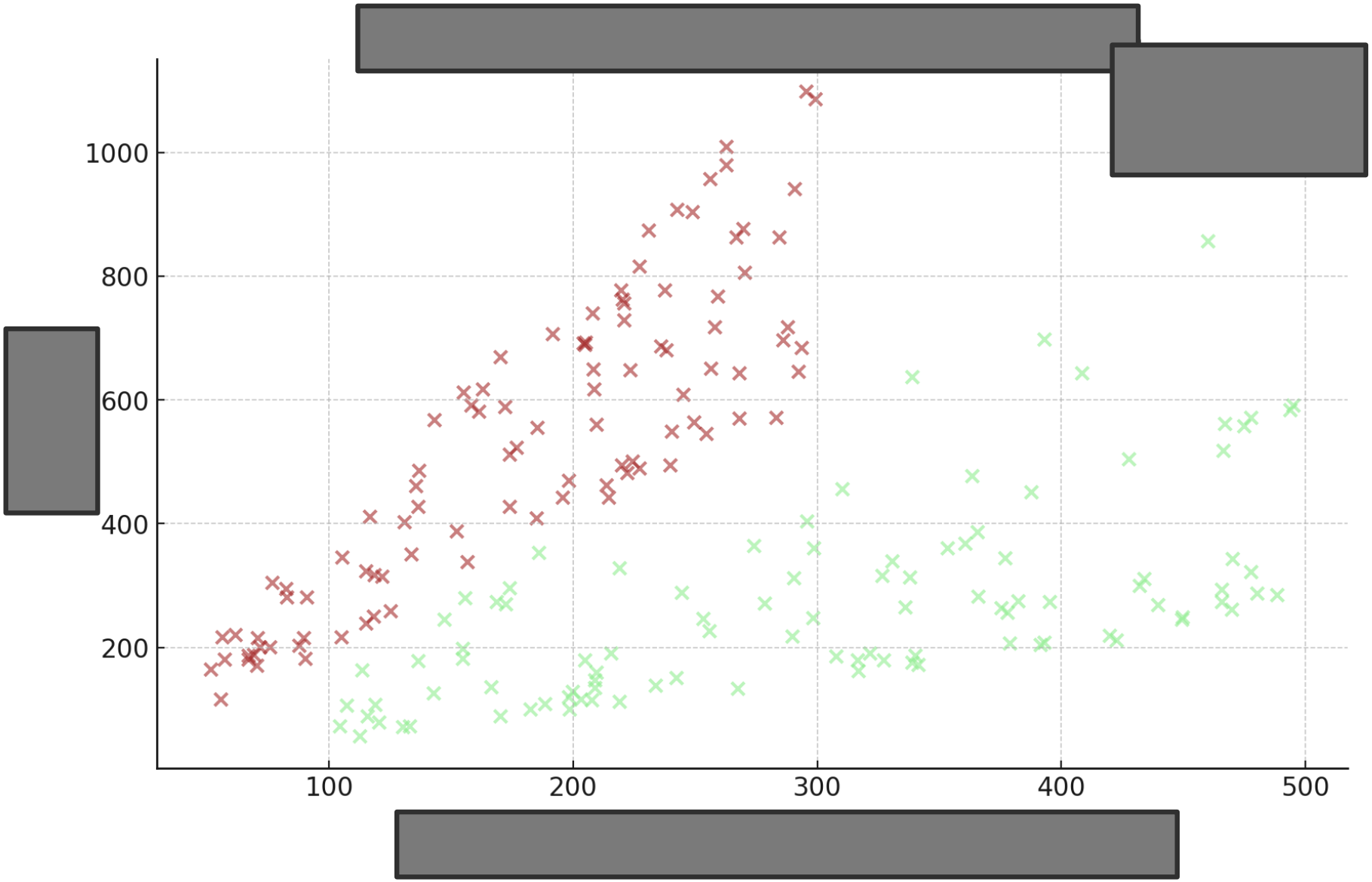
EXAMPLES


Time (hours)
0
2
4
6
8
10
12
14
16
0
250
500
750
1000
1250
1500
1750
2000
Distance (km)

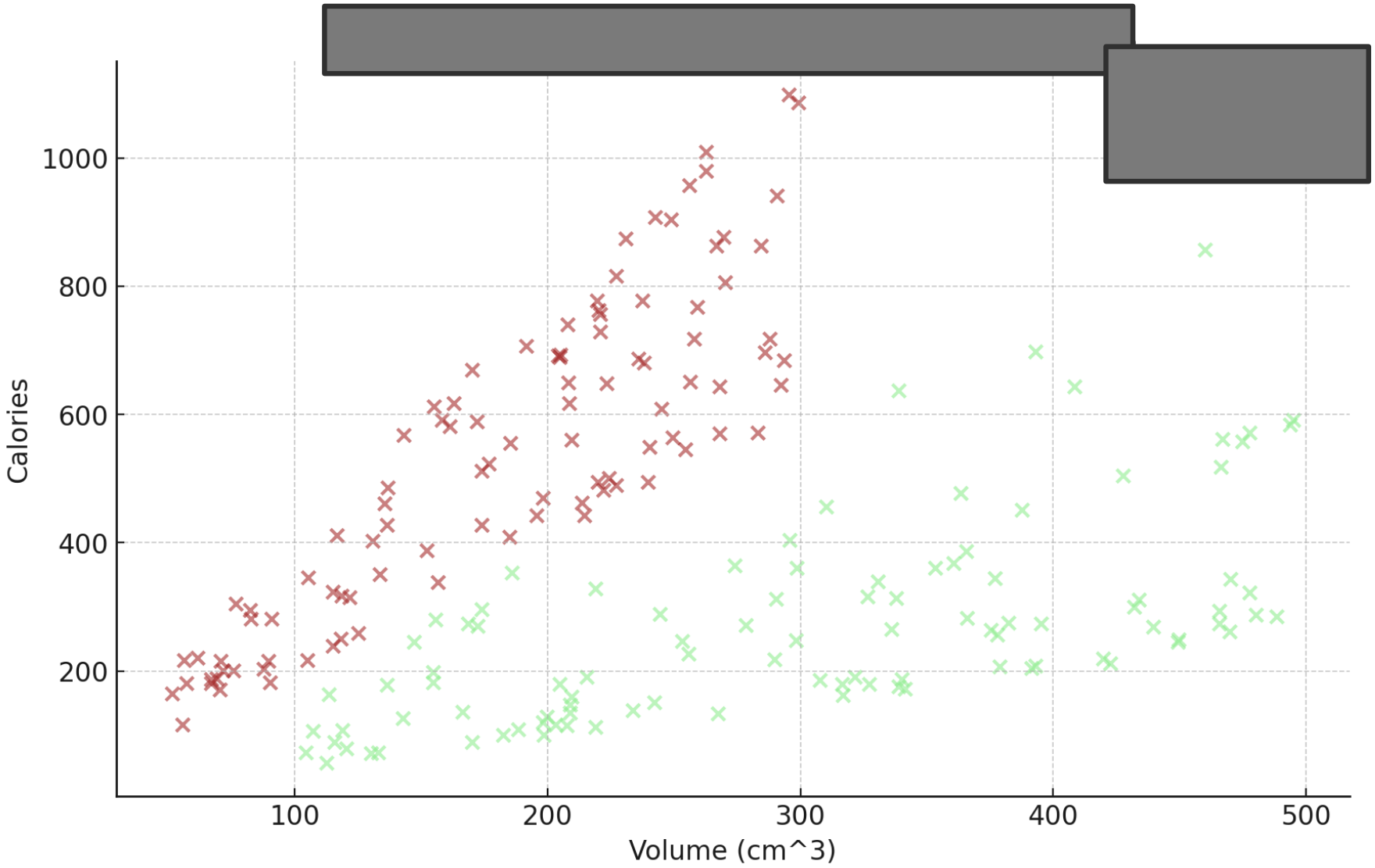
EXAMPLES


Distance vs. Time for Car and Plane Trips
Car Trips
Plane Trips
Time (hours)
Distance (km)
0
2
4
6
8
10
12
14
16
0
250
500
750
1000
1250
1500
1750
2000

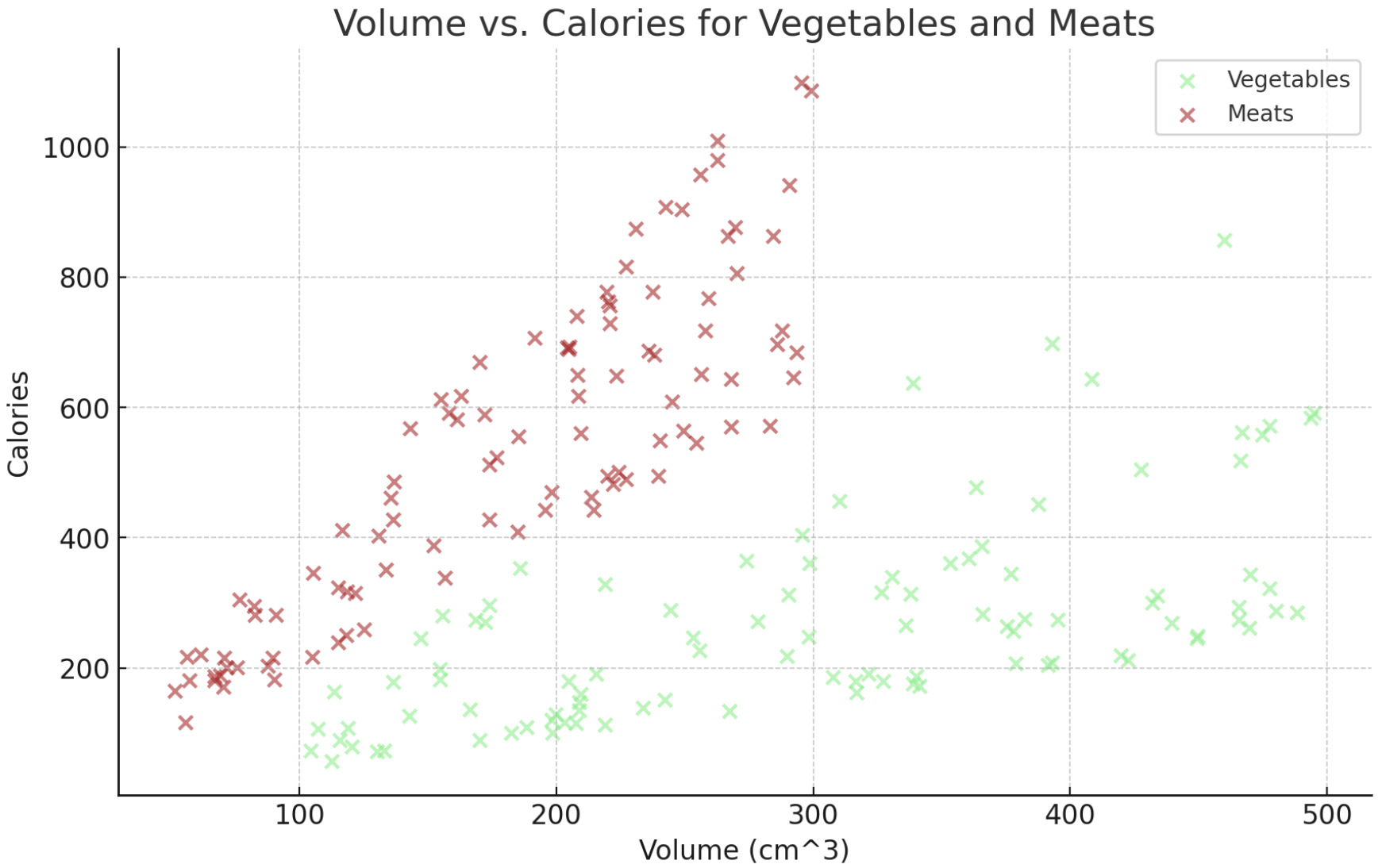
EXAMPLES



EXAMPLES

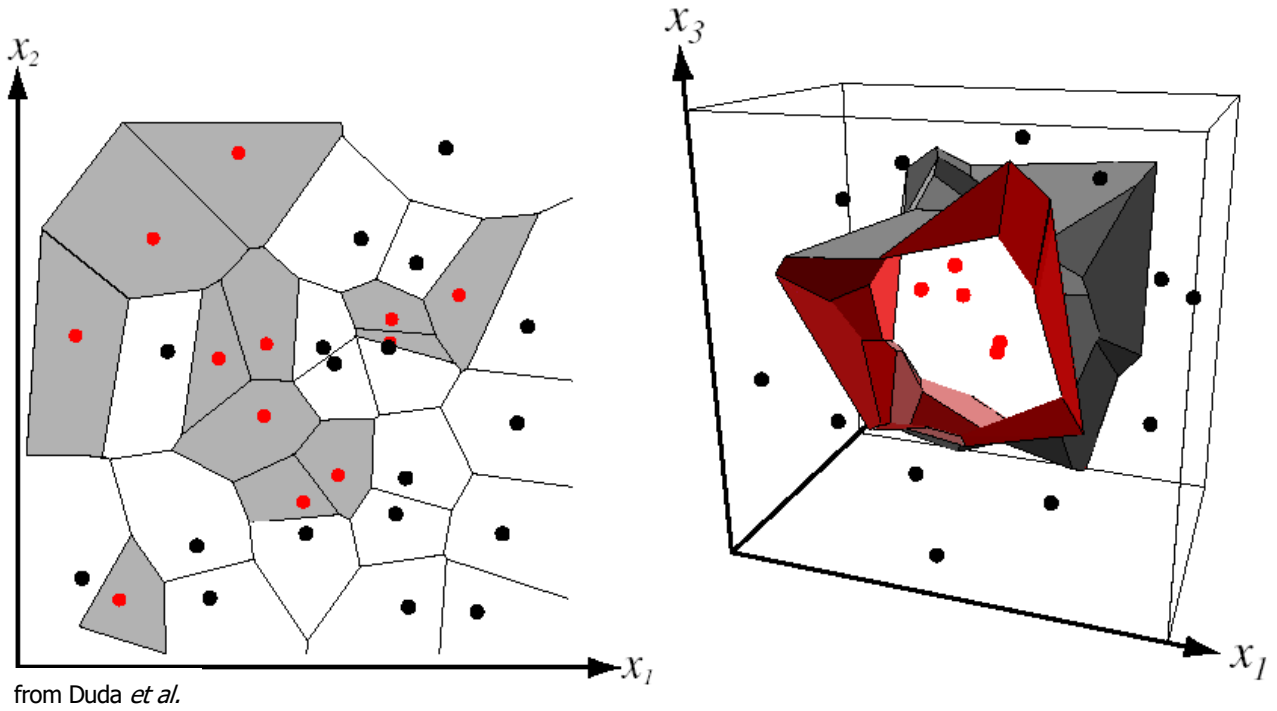


EXAMPLES


Volume vs. Calories for Vegetables and Meats
Vegetables
Meats
Calories
1000
800
600
400
200
100
200
300
400
500
Volume (cm^3)

DECISION BOUNDARIES KNN

Assign label of nearest training data point to each test data point



from Duda *et al.*

Voronoi partitioning of feature space
for two-category 2D and 3D data

Source: D. Lowe

MANY CLASSIFIERS TO CHOOSE FROM

K-nearest neighbor

Support Vector Machines

Decision Trees

Random Forest

(Gradient) Boosted Decision Trees

Logistic Regression

Naïve Bayes

Bayesian network

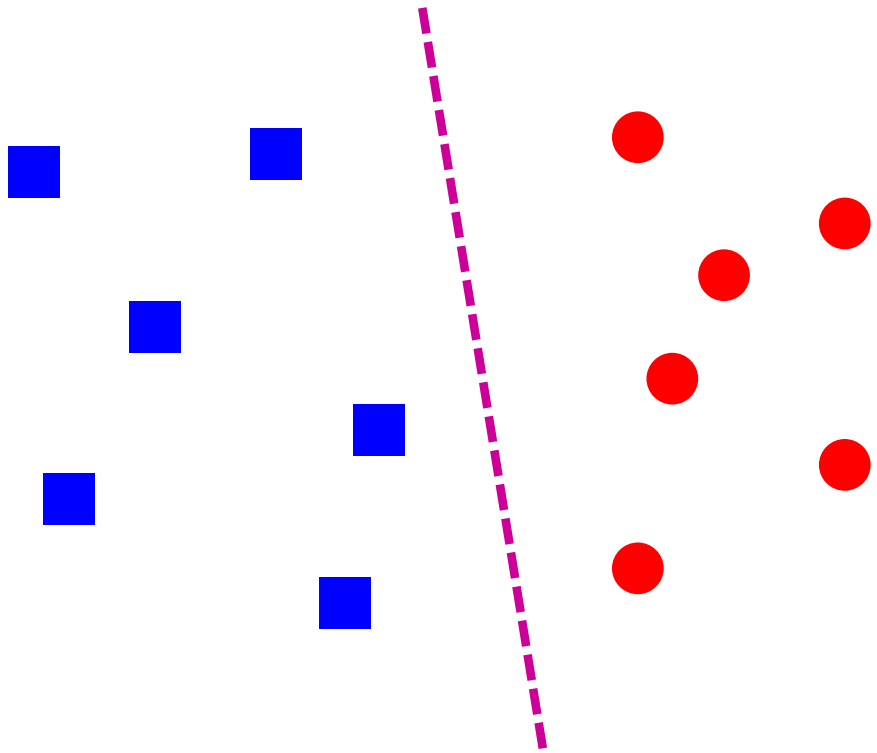
RBMs

….

Which is the best one?

Slide credit: D. Hoiem

CLASSIFIERS: LINEAR

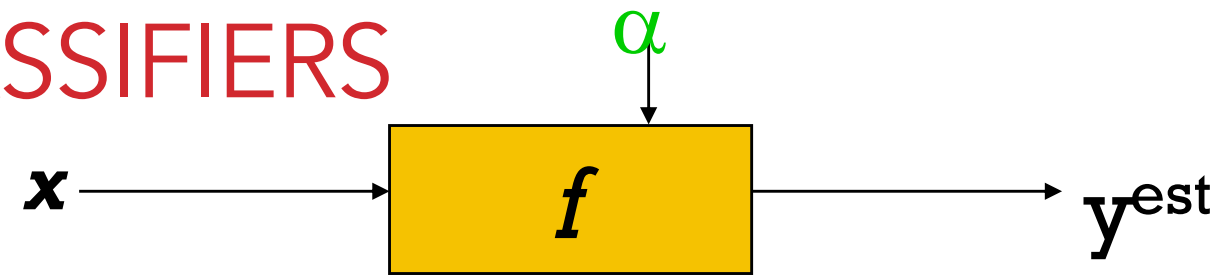


Find a *linear function* to separate the classes:

$$f(x) = \text{sgn}(w \cdot x + b)$$

Slide credit: L. Lazebnik

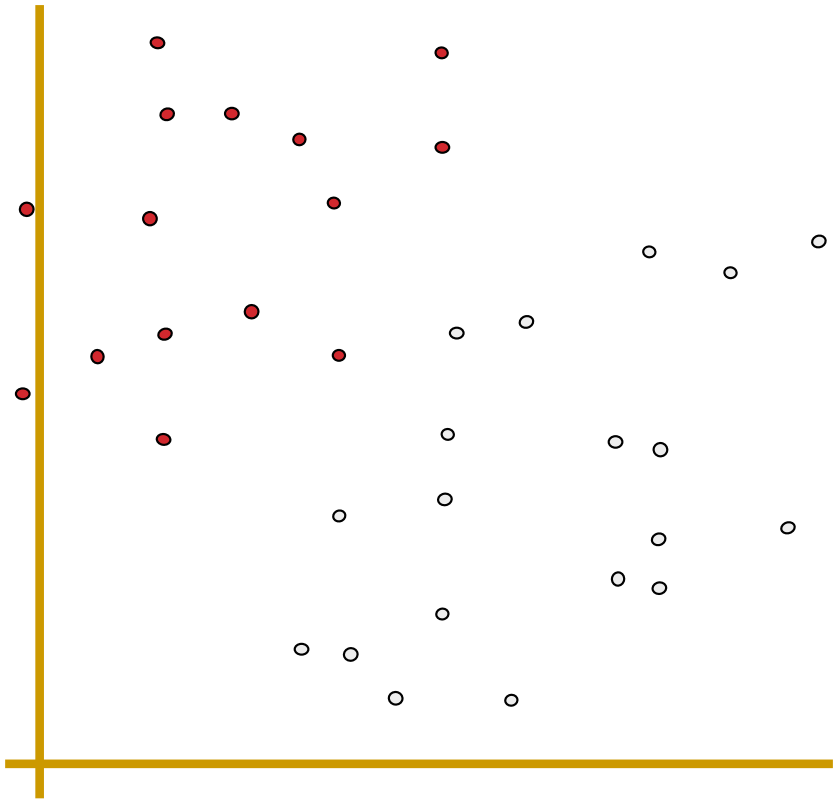
LINEAR CLASSIFIERS



$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\mathbf{x} - b)$

denotes +1

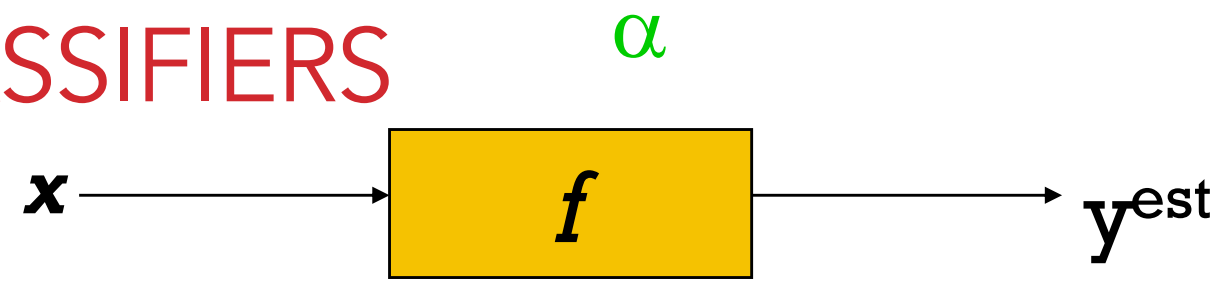
denotes -1



How would you classify this data?

Slides from Andrew W. Moore

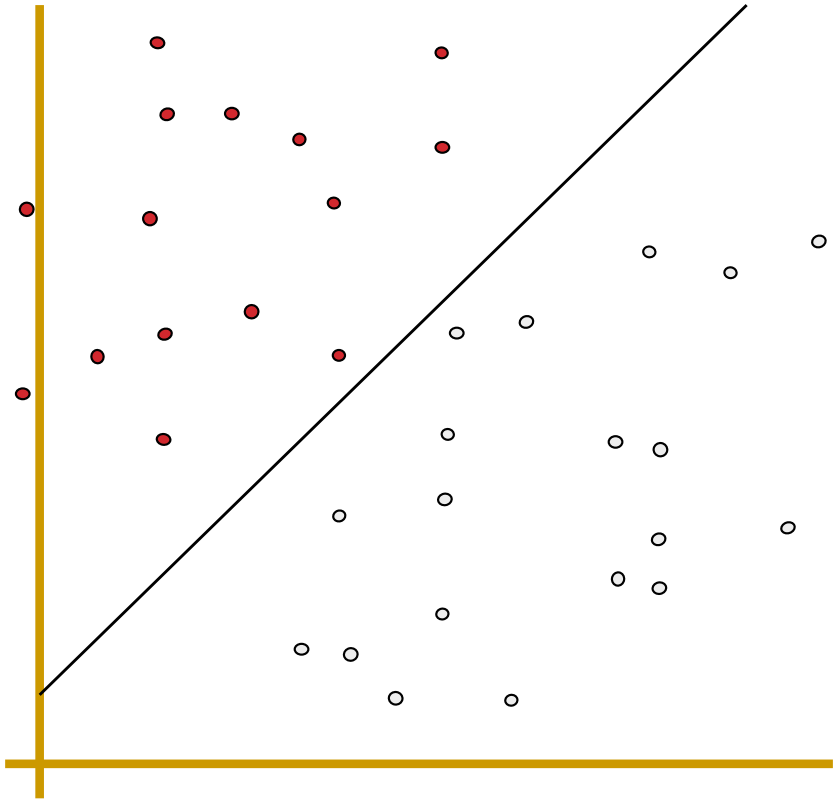
LINEAR CLASSIFIERS



$f(\mathbf{x}, \mathbf{w}, b) = sign(\mathbf{w}.\mathbf{x} - b)$

denotes +1

denotes -1

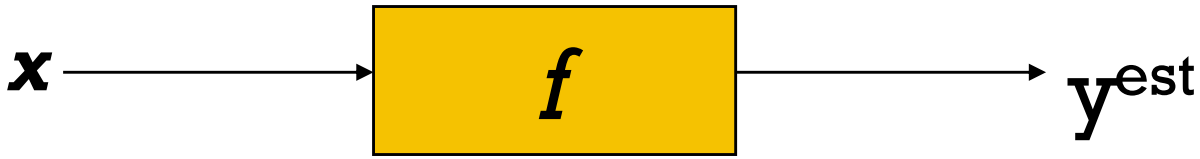


How would you classify this data?

Slides from Andrew W. Moore

LINEAR CLASSIFIERS

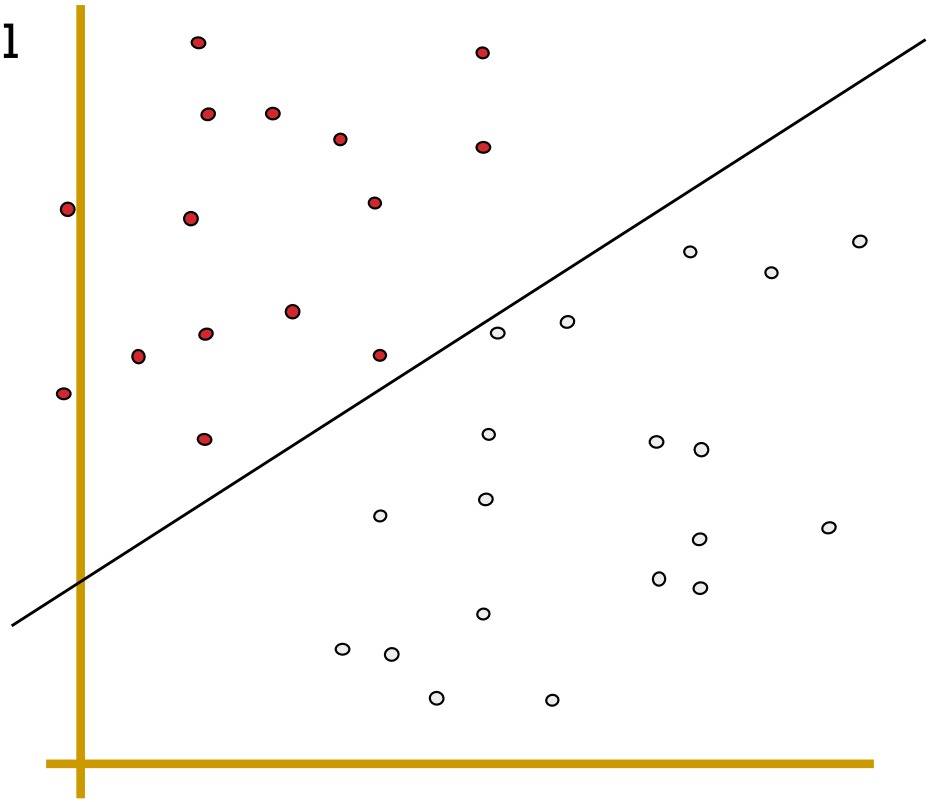
α



$f(x,w,b) = sign(w.x - b)$

denotes +1

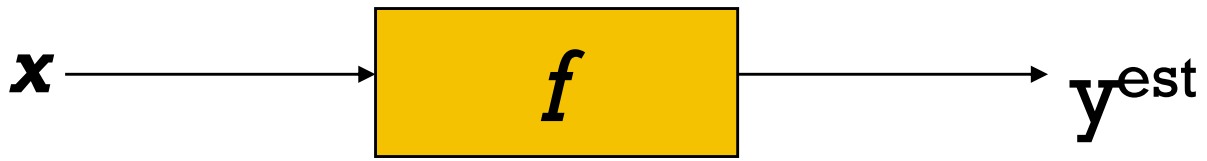
denotes -1



How would you classify this data?

Slides from Andrew W. Moore

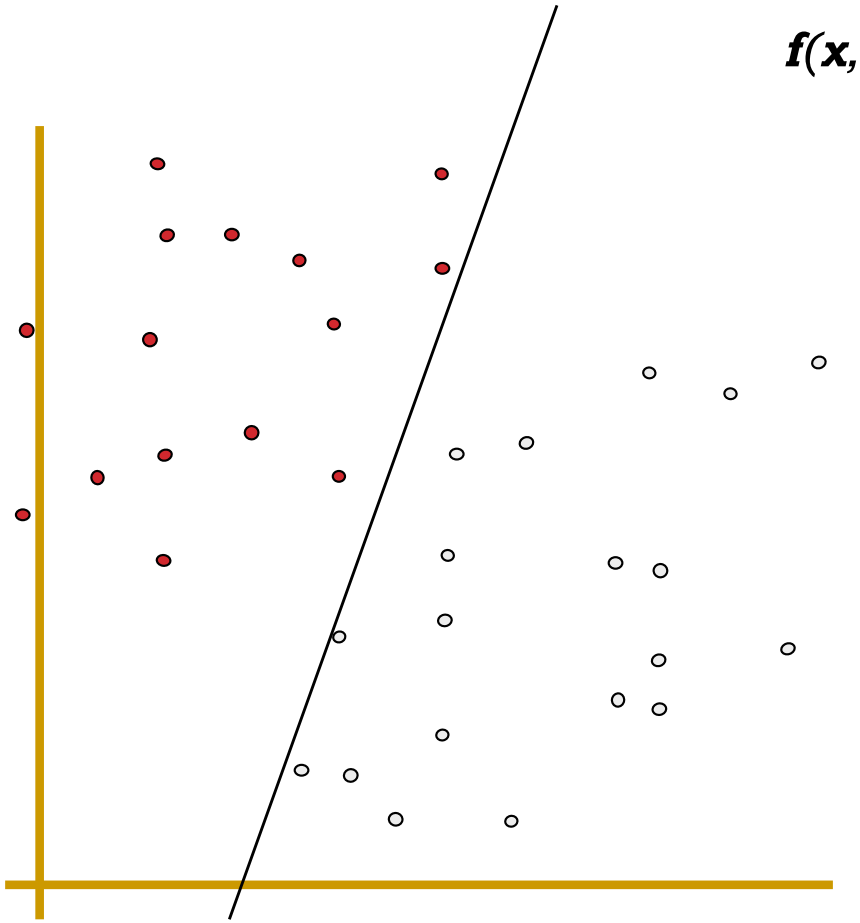
LINEAR CLASSIFIERS



$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\mathbf{x} - b)$

denotes +1

denotes -1

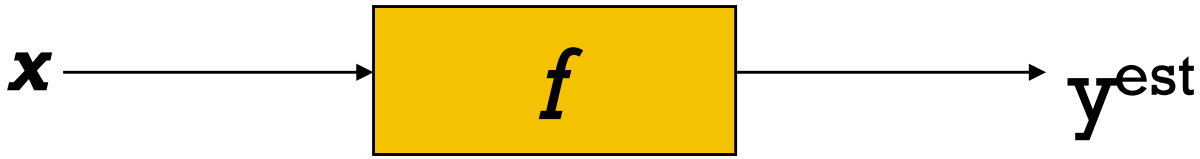


How would you classify this data?

Slides from Andrew W. Moore

LINEAR CLASSIFIERS

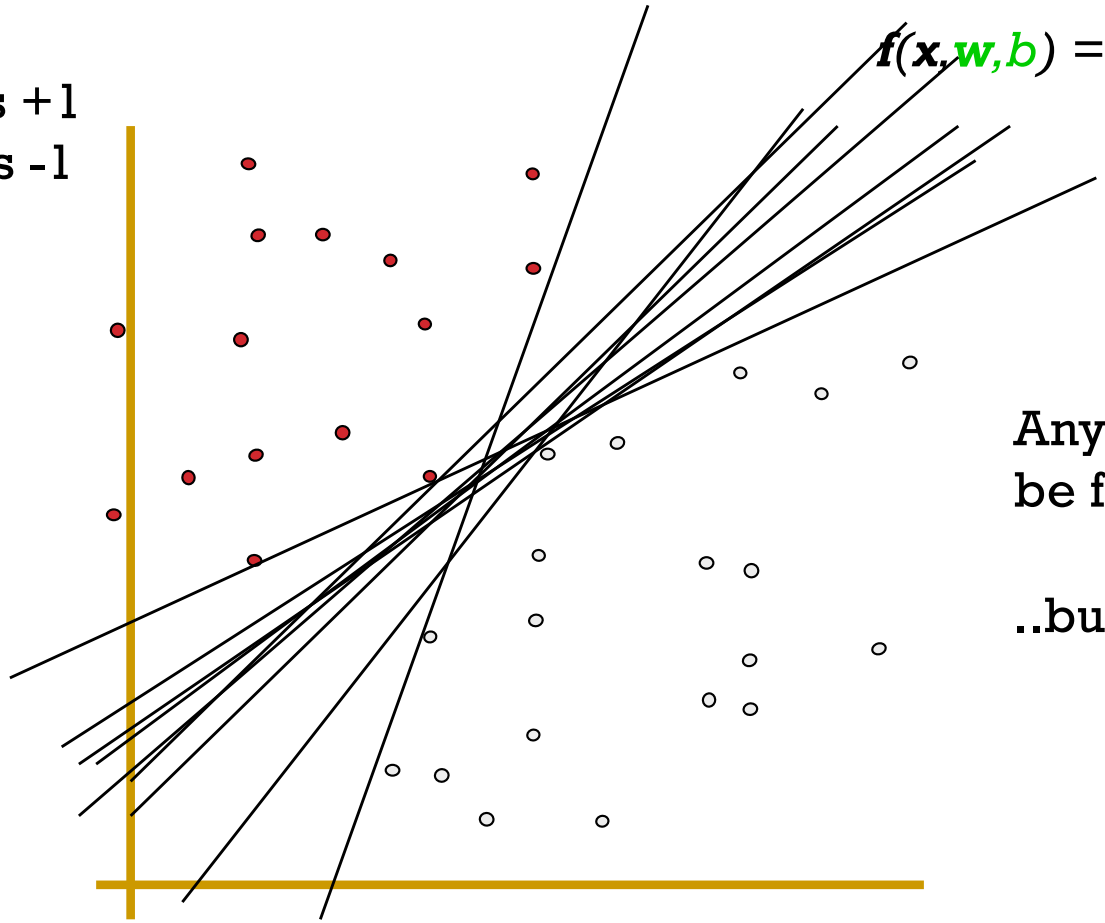
α



$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\mathbf{x} - b)$

denotes +1

denotes -1

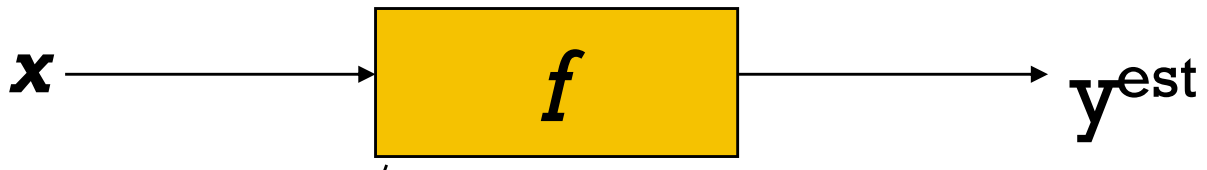


Any of these would be fine..

..but which is best?

Slides from Andrew W. Moore

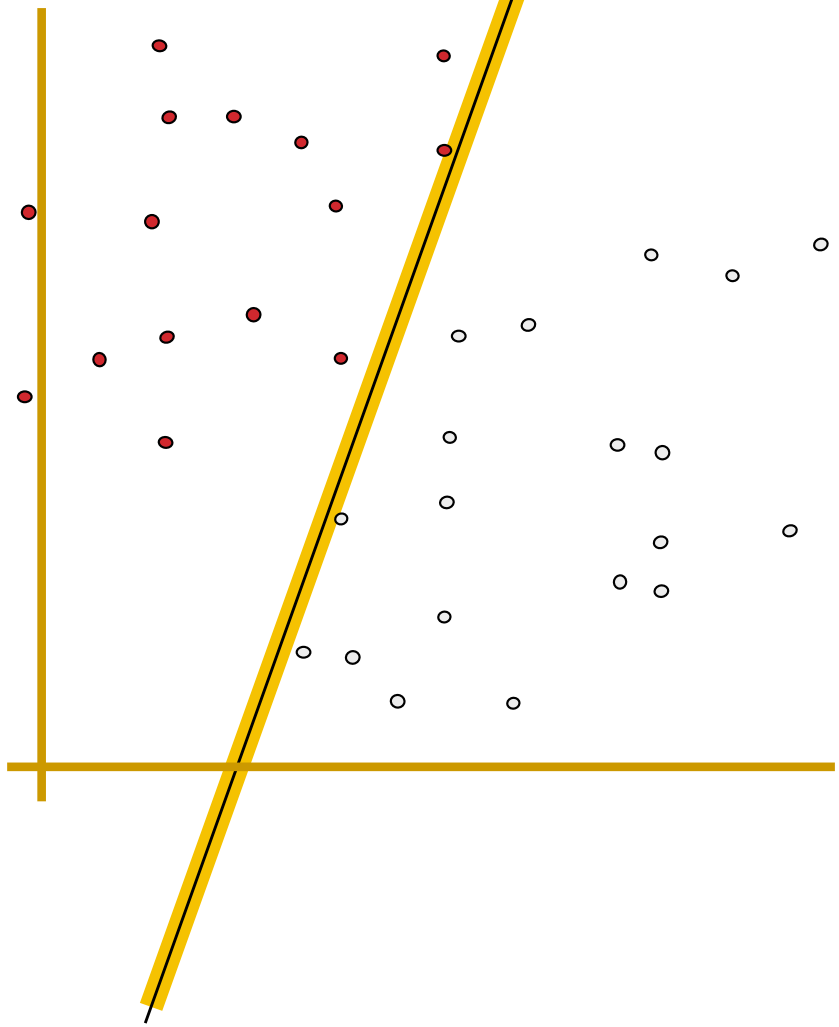
CLASSIFIER MARGIN



$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\mathbf{x} - b)$

denotes +1

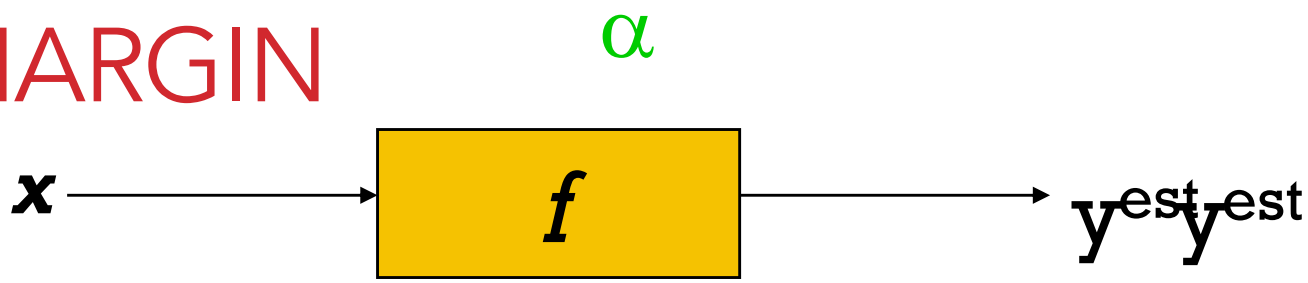
denotes -1

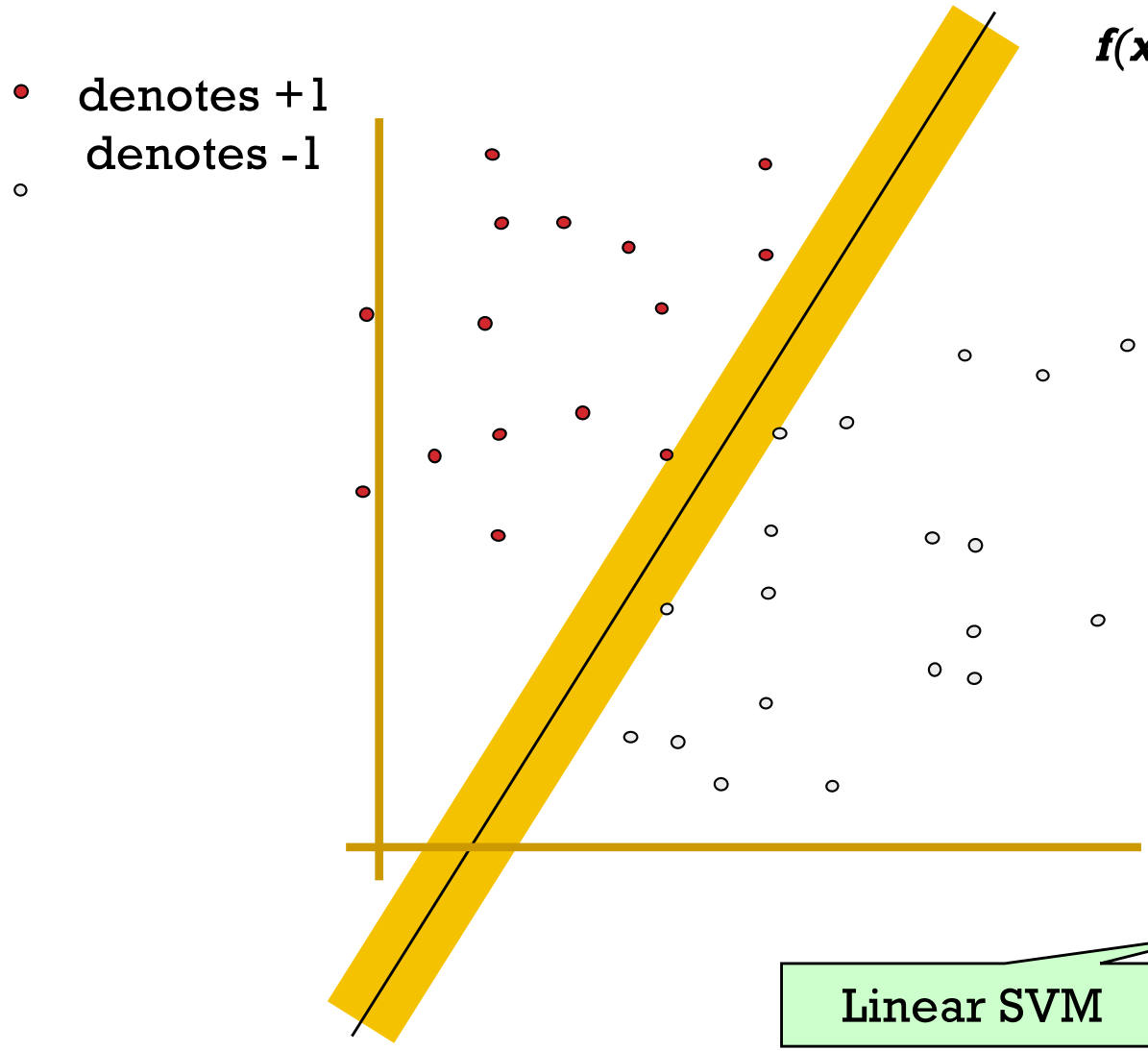


Define the margin of a linear classifier as the width that the boundary could be increased by before hitting a datapoint.

Slides from Andrew W. Moore

MAXIMUM MARGIN

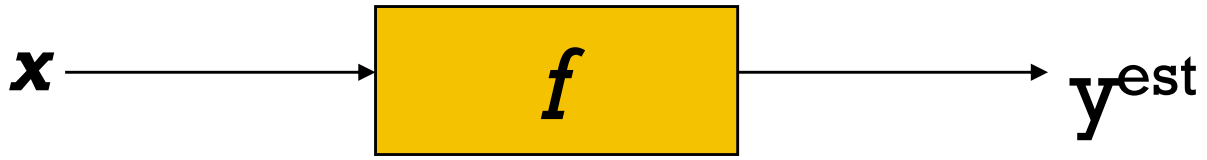




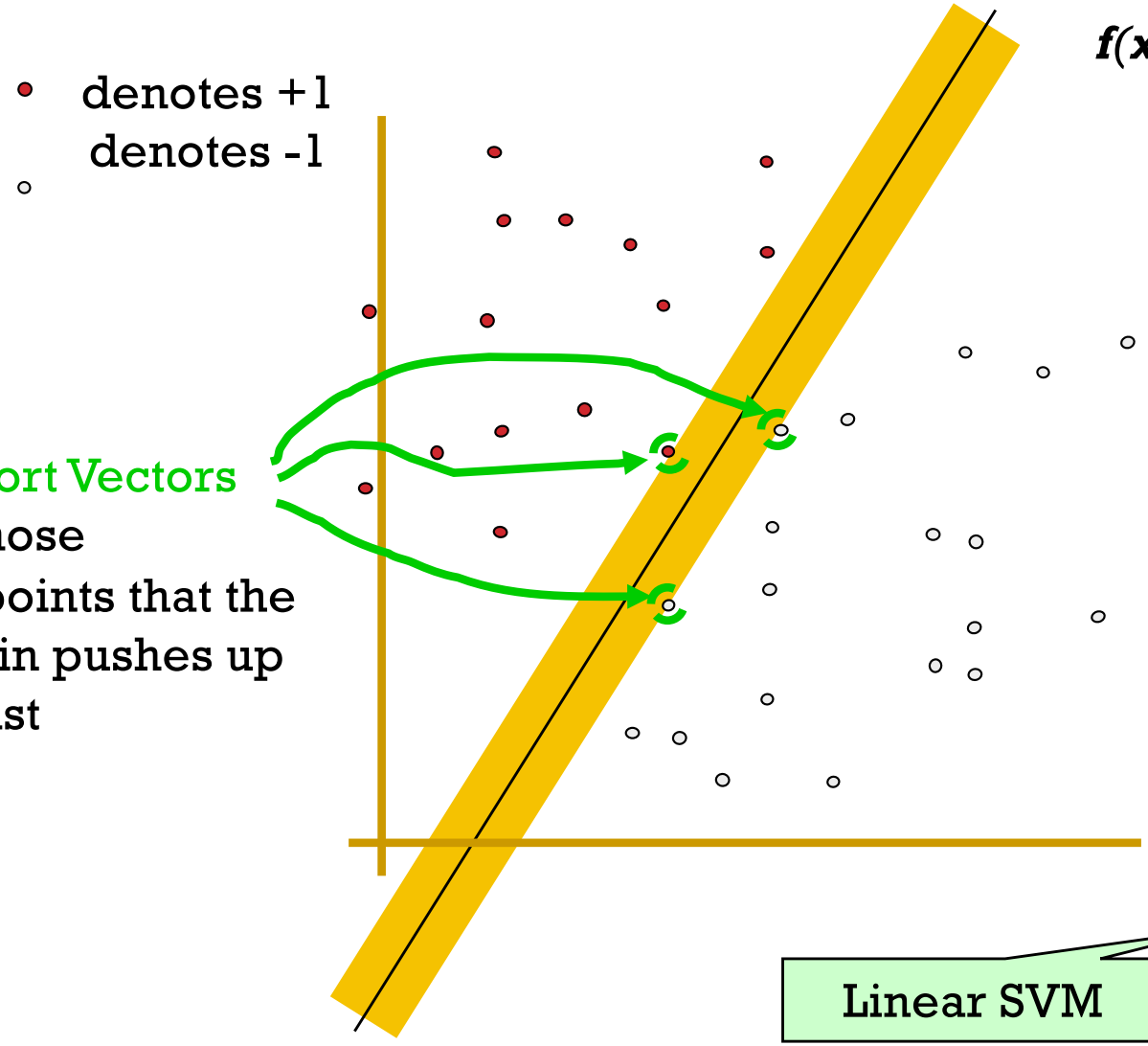
The maximum margin linear classifier is the linear classifier with the, um, maximum margin. This is the simplest kind of SVM (Called an LSVM)

Slides from Andrew W. Moore

MAXIMUM MARGIN



$f(x,w,b) = sign(w.x - b)$



Support Vectors are those datapoints that the margin pushes up against

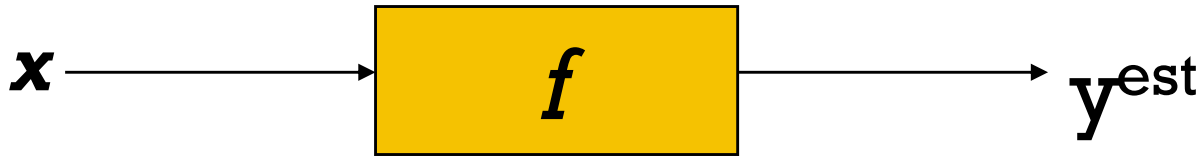
The maximum margin linear classifier is the linear classifier with the, um, maximum margin. This is the simplest kind of SVM (Called an LSVM)

Linear SVM

Slides from Andrew W. Moore

MAXIMUM MARGIN

α



$f(x,w,b) = sign(w.x - b)$

denotes +1

denotes -1

The maximum

ar

the

ifier

n,

margin.

This is the simplest kind of SVM (Called an LSVM)

Linear SVM

Support

are thos

datapoi

margin pushes up against

`sklearn.linear_model.SGDClassifier`

Default loss: "hinge" → linear SVM.

Slides from Andrew W. Moore

MANY CLASSIFIERS TO CHOOSE FROM

K-nearest neighbor

Support Vector Machines

Decision Trees

Random Forest

(Gradient) Boosted Decision Trees

Logistic Regression

Naïve Bayes

Bayesian network

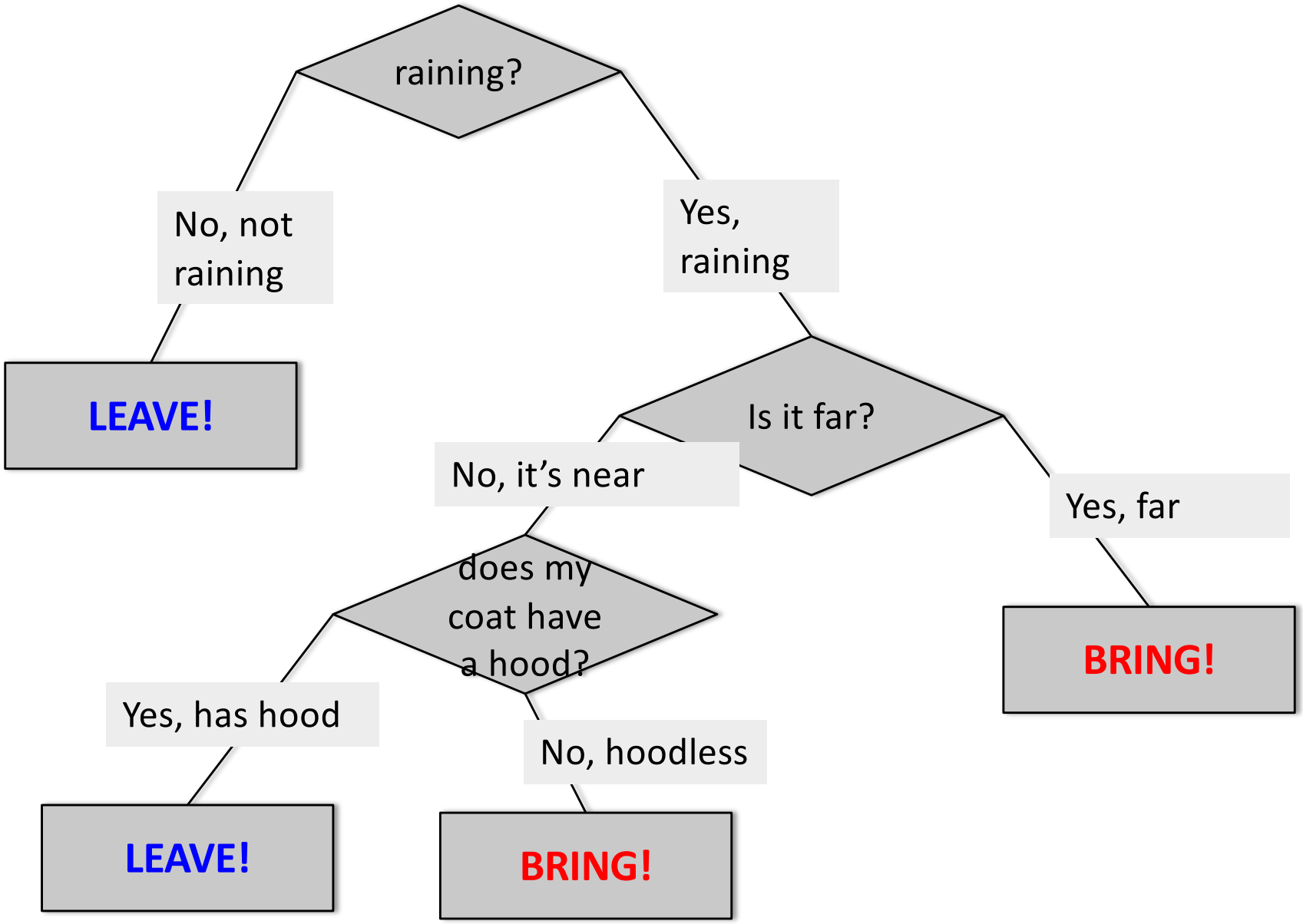
RBMs

….

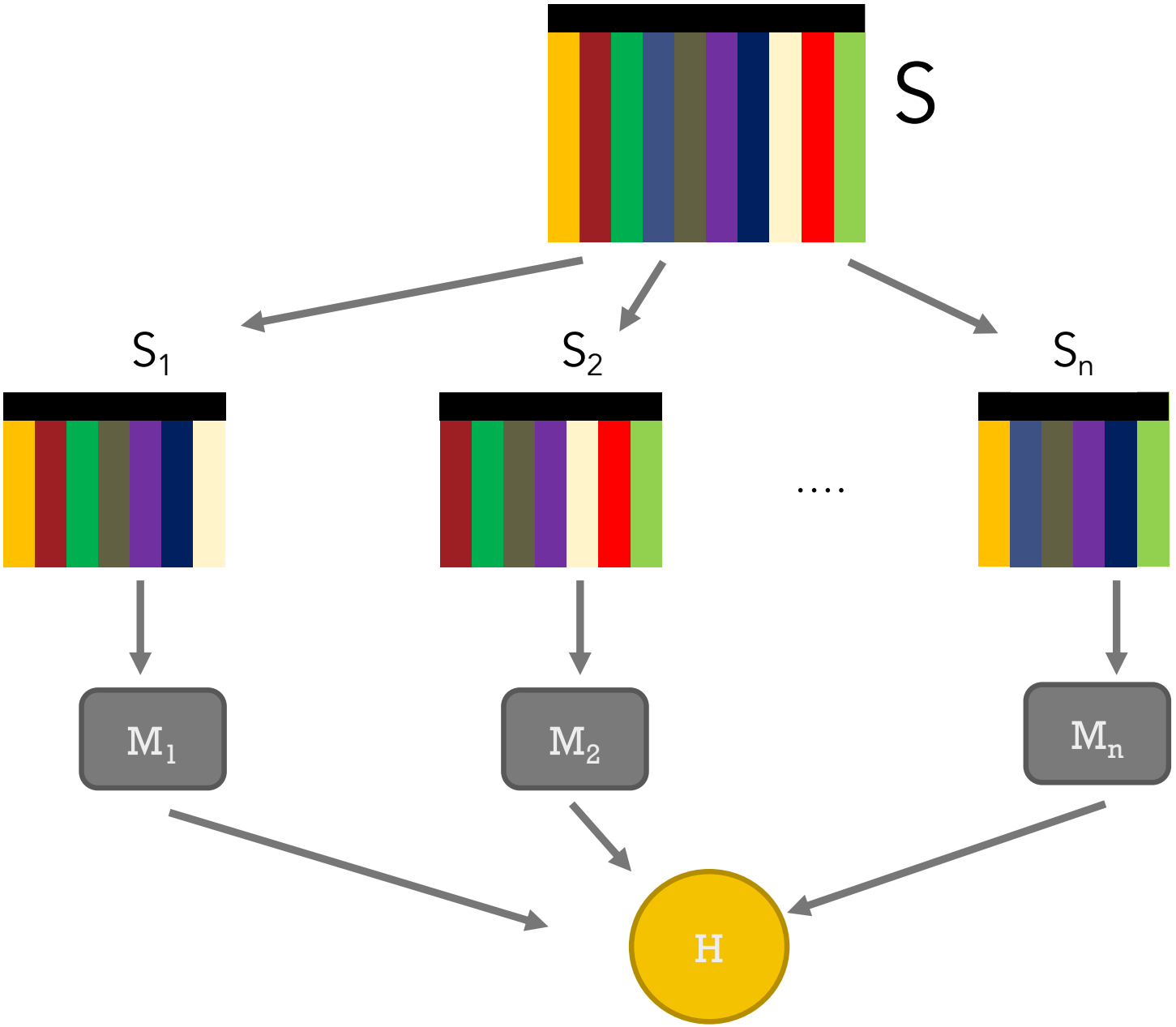
Which is the best one?

Slide credit: D. Hoiem

DECISION TREES



RANDOM FOREST



Sample with Replacement & select random subset of features*

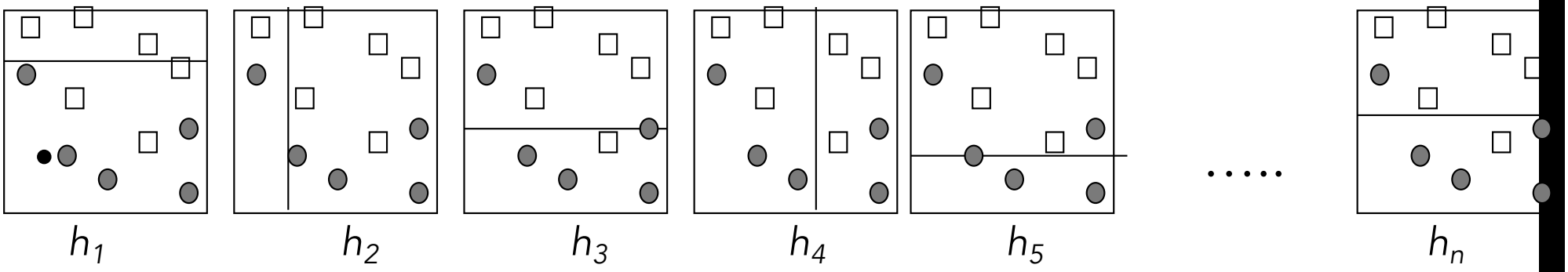
Build classifier over sample

Use majority Vote for classification (or avg. for regression)

* Normally done for each node of the decision tree – not once

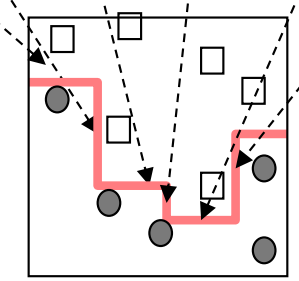
ADABOOST - CORE IDEA

Take a set of weak classifiers (normally they should do better than guessing)



Classification Result

θ_1 θ_2 θ_3 θ_4 θ_5 θ_n



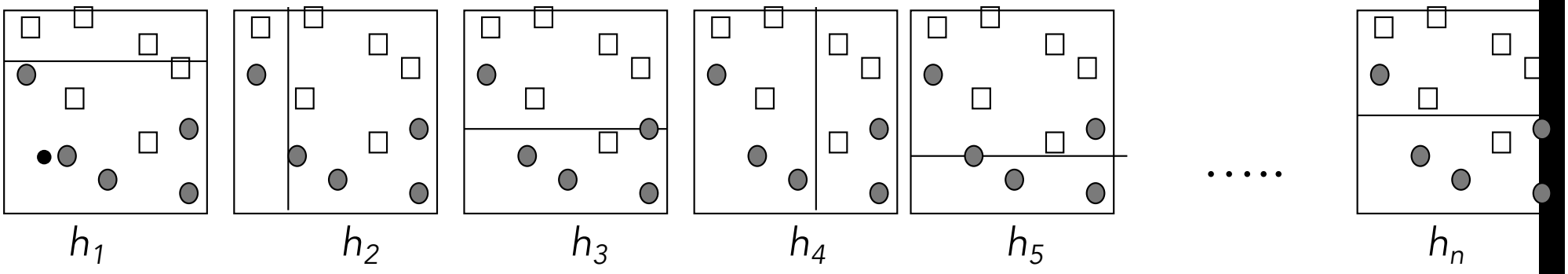
Weight the result of each classify with θ

Combine to form the Final strong classifier

$$H(x) = sign\left(\sum_{i=1}^{n} \theta_i h_i(x)\right)$$

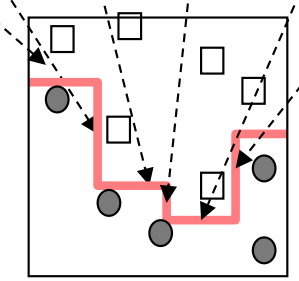
ADABOOST - CORE IDEA

Take a set of weak classifiers (normally they should do better than guessing)



Classification Result

θ_1 θ_2 θ_3 θ_4 θ_5 θ_n



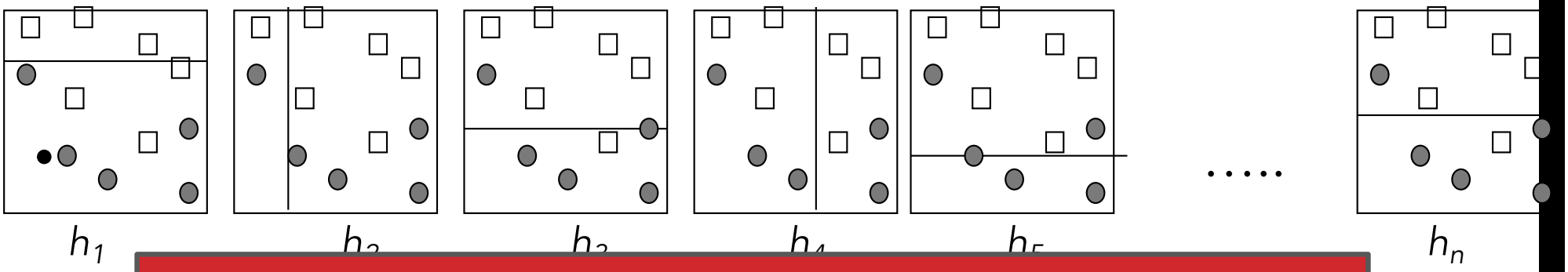
Weight the result of each classify with θ

Combine to form the Final strong classifier

$$H(x) = sign\left(\sum_{i=1}^{n} \theta_i h_i(x)\right)$$

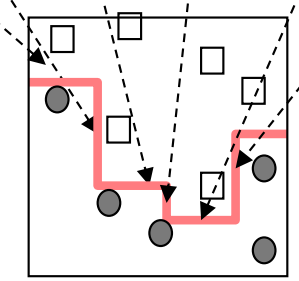
ADABOOST - CORE IDEA

Take a set of weak classifiers (normally they should do better than guessing)



Classificat
Result

XGBoost follows the same idea

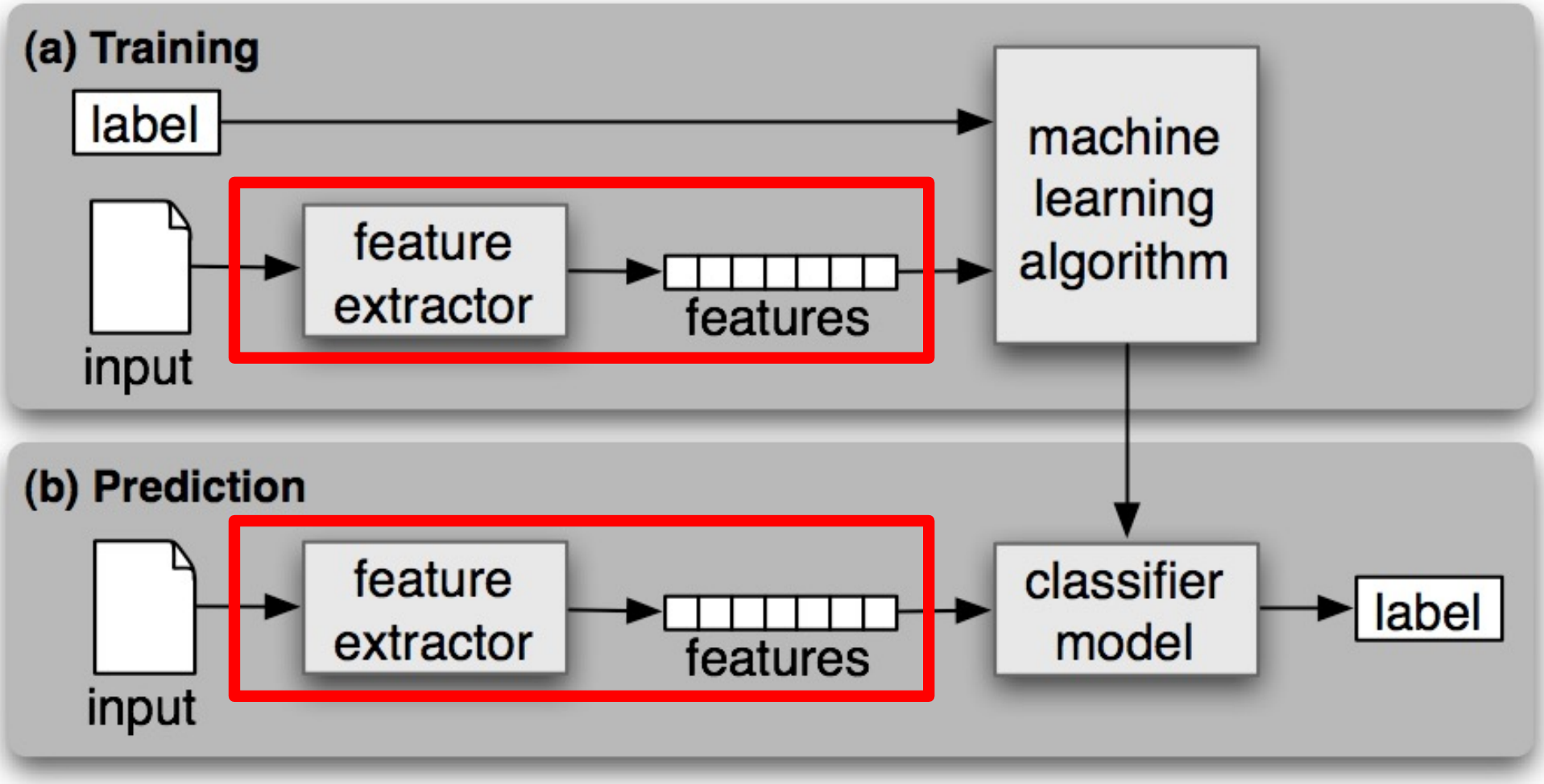


Weight the result of each classify with θ

Combine to form the Final strong classifier

$$H(x) = sign\left(\sum_{i=1}^{n} \theta_i h_i(x)\right)$$

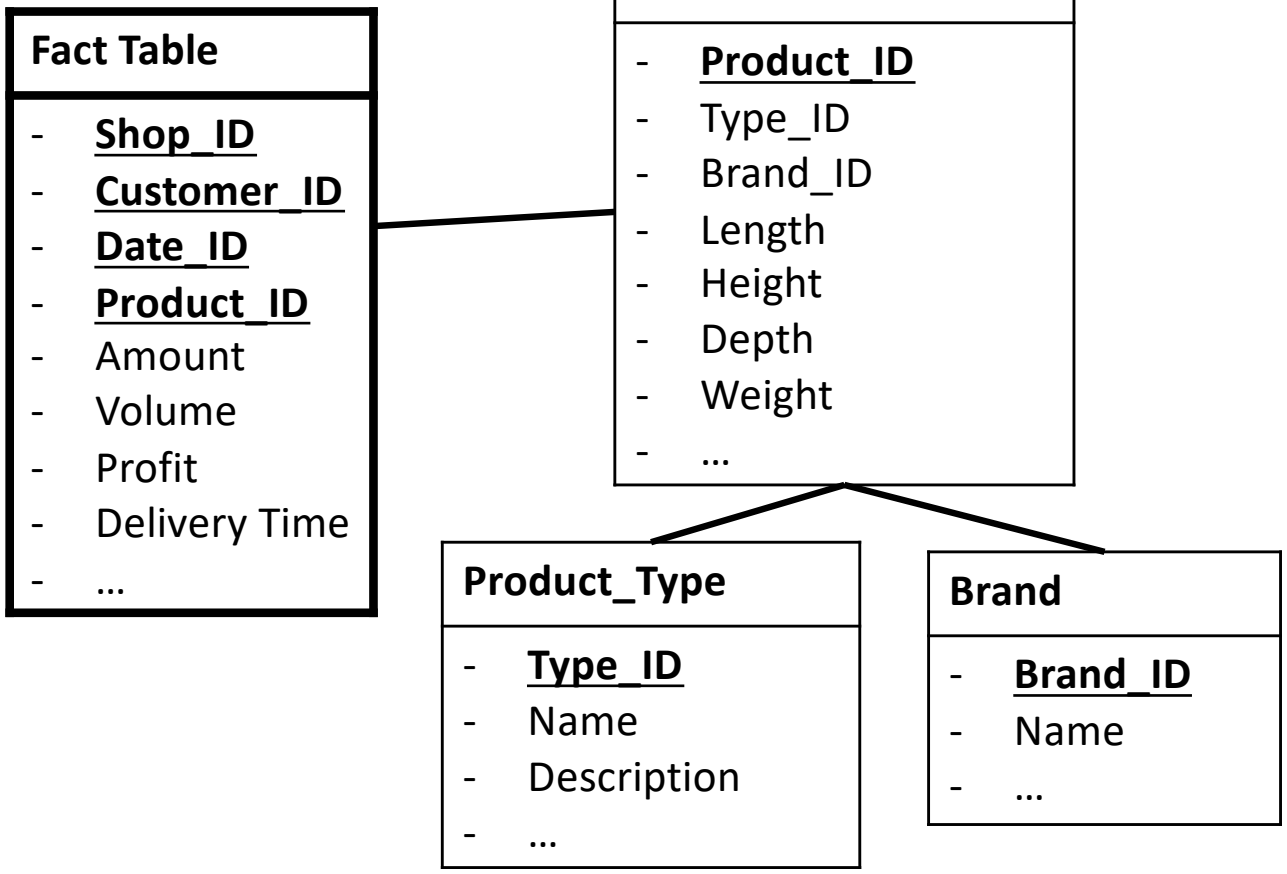
ML PIPELINE (SUPERVISED)



FEATURES

Fact Table

- **Shop_ID**
- **Customer_ID**
- **Date_ID**
- **Product_ID**
- Amount
- Volume
- Profit
- …

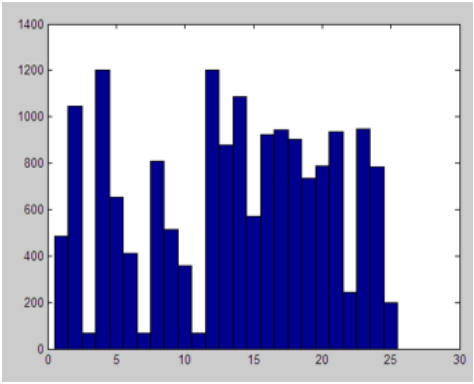


Customer State	Product Type	Product Weight	Volume (L*H*D)	Month	Delivery Time

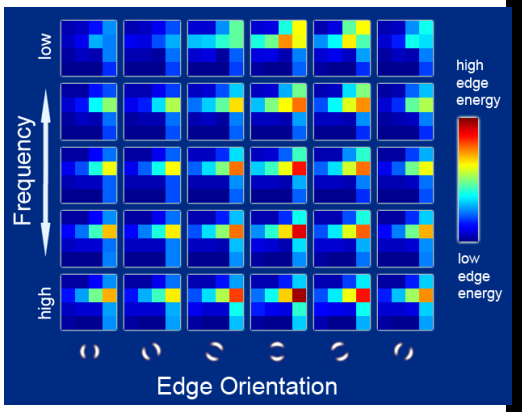
IMAGE FEATURES

Raw pixels

Histograms



GIST descriptors



…

Slide credit: L. Lazebnik

TEXT FEATURES

Dear Home Owner,

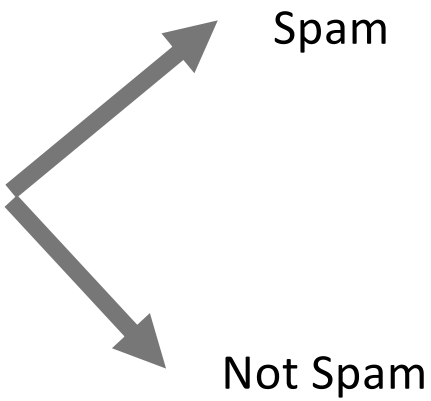
Your credit doesn't matter to us! If you own real estate and want IMMEDIATE cash to spend ANY way you like, or simply wish to LOWER your monthly payments by one third or more, here are the deals we have today:

$488.000,00 at 3.67% fixed rate
$372.000,00 at 3.90% variable-rate
$492.000,00 at 3.21% interest-only
$248.000,00 at 3.36% fixed rate
$198.000,00 at 3.55% variable rate

Hurry, when these deals are gone, they're gone!
Simple fill out the 1 minute form.

Don't worry about approval, credit is not a matter!

CLICK HERE AND FILL THE 60 SECS FORM!



Bag of Words

$$\begin{pmatrix} Urgent: 1 \\ money: 1 \\ Herbel: 2 \\ Pills: 2 \\ Are: 1 \\ \dots \end{pmatrix}$$

N-Grams

$$\begin{pmatrix} herbel\ pills: 1 \\ pills\ for: 1 \\ for\ Hair: 2 \\ Hair\ growth: 1 \\ surgeries: 2 \\ \dots \end{pmatrix}$$

ONE-HOT ENCODING

Bag of Words

$$\begin{pmatrix} Urgent \\ Money \\ Herbel \\ Pills \\ Are \\ \dots \end{pmatrix}$$

ID	Urgent	Money	Herbel	Pills	Are	….
Mail1	0	1	1	0	1	…
Mail2	1	0	0	1	1	…
…	…	…	…	…	…	…

PREDICTOR FOR GRAD-SCHOOL APPLICATIONS

Name	ZipCode	Age	Sex	Area	Avg Grade	Statement	Early admit	Accepted
Mike	02474	23	M	DB	B-	Since I was born, I knew I wanted to code. My first program I wrote in binary code literally in the sandbox, though I am not sure it was correct…	No	NO
Sam	02456	21	M	Sensor	A	Celine Dion's song "A New Day Has Come" taught me that CS is the best subject in the world. I never felt…	Yes	Yes
Amadou	15106	22	M	DB	A+	I want to get out of Pittsburgh.	No	Yes
Anna	02319	22	F	ML	A-	I already wrote 10 papers and I think I am ready to graduate now.	Yes	Yes
…	…	..	…	…	…	…	…	…

HOW WOULD YOU ENCODE THE TABLE?

PREDICTOR FOR GRAD-SCHOOL APPLICATIONS

Encode as numbers (0-1)

Encode as numbers (0-1)

Bag of words 1-Hot Encoding

Remove (information leakage)

Name	ZipCode	Age	Sex	Area	Avg Grade	Statement	Early admit	Accepted
Mike	02474	23	M	DB	B-	Since I was born, I knew I wanted to code. My first program I wrote in binary code literally in the sandbox, though I am not sure it was correct…	No	NO
Sam	02456	21	null	Sensor	A	Celine Dion's song "A New Day Has Come" taught me that CS is the best subject in the world. I never felt…	Yes	Yes
Amadou	15106	22	M	DB	A+	I want to get out of Pittsburgh.	No	Yes
Anna	null	22	F	ML	A-	I already wrote 10 papers and I think I am ready to graduate now.	Yes	Yes
…	…	..	…	…	…	…	…	…

Remove identifiers

Encode as (1) Lat/Lon and scale to 0-1, or remove

Scale to 0-1

1-Hot Encode or remove

PREDICTOR FOR GRAD-SCHOOL APPLICATIONS

Name	ZipCode	Age	Sex	Area	Avg Grade	Statement	Early admit	Accepted
Mike	02474	23	M	DB	B-	Since I was born, I knew I wanted to code. My first program I wrote in binary code literally in the sandbox, though I am not sure it was correct…	No	NO
Sam	02456	21	null	Sensor	A	Celine Dion's song "A New Day Has Come" taught me that CS is the best subject in the world. I never felt…	Yes	Yes
Amadou	15106	22	M	DB	A+	I want to get out of Pittsburgh.	No	Yes
Anna	null	22	F	ML	A-	I already wrote 10 papers and I think I am ready to graduate now.	Yes	Yes
…	…	..	…	…	…	…	…	…

No zip	Lon	Lat	Age	M	F	Other	ML	DB	Sens	Avg Grade	Born	Celine	Pittsburg	…	Accepted
0	.123	.0222	0.8	1	0	1	0	1	0	0.5	1	0	0	…	NO
0	.123	.0123	0.5	1	0	M	0	0	1	0	0	1	0	…	Yes
0	.031	.0322	0.5	1	0	M	0	1	0	0	0	0	1	…	Yes
1			0.5	0	1	F	1	0	0	0.1	0	0	0	…	Yes
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…