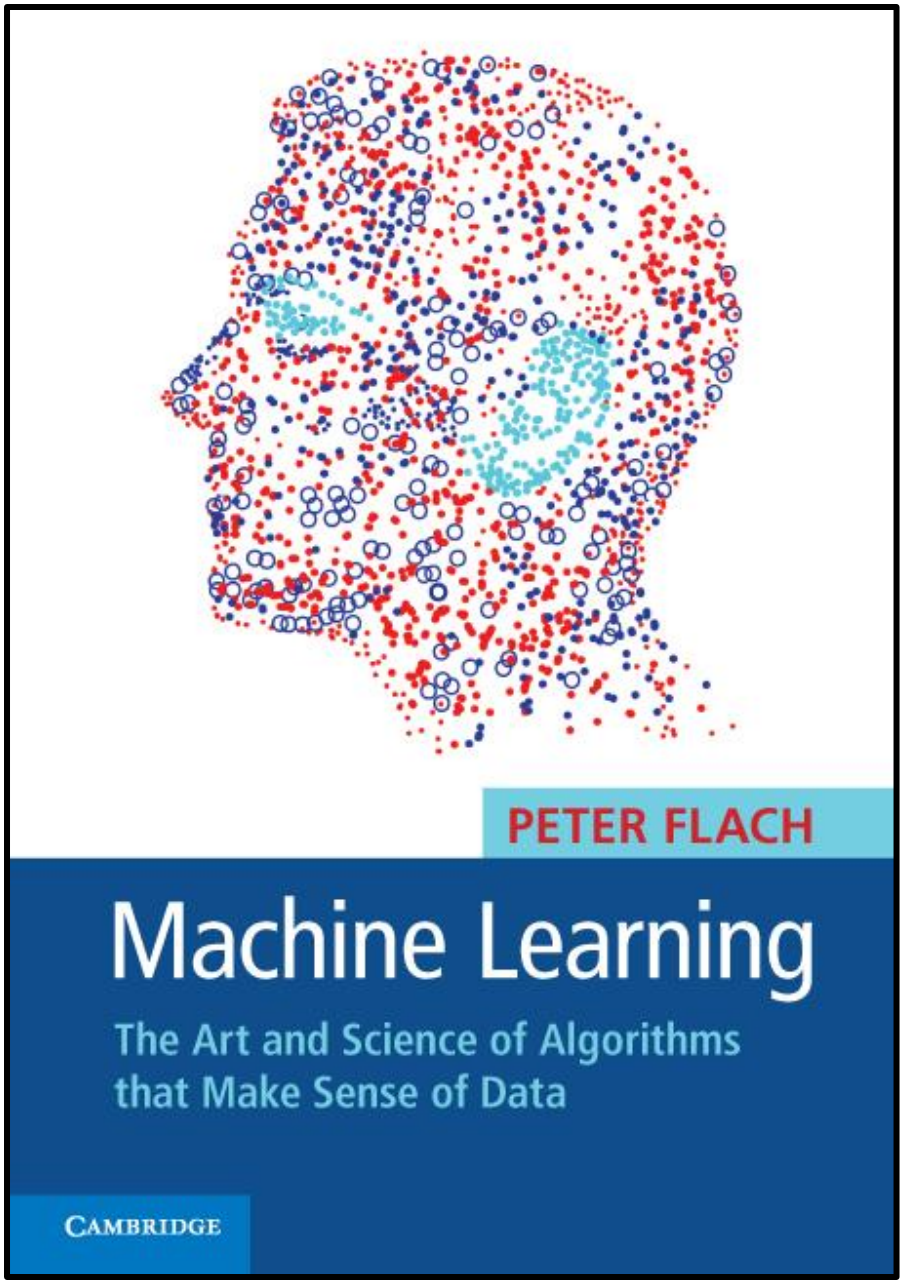
A PRACTICAL GUIDE TO MACHINE LEARNING


PETER FLACH
Machine Learning
The Art and Science of Algorithms
that Make Sense of Data
CAMBRIDGE

TODAY:

Part 1: ML

Part 2: In-class lab

WEDNESDAY:

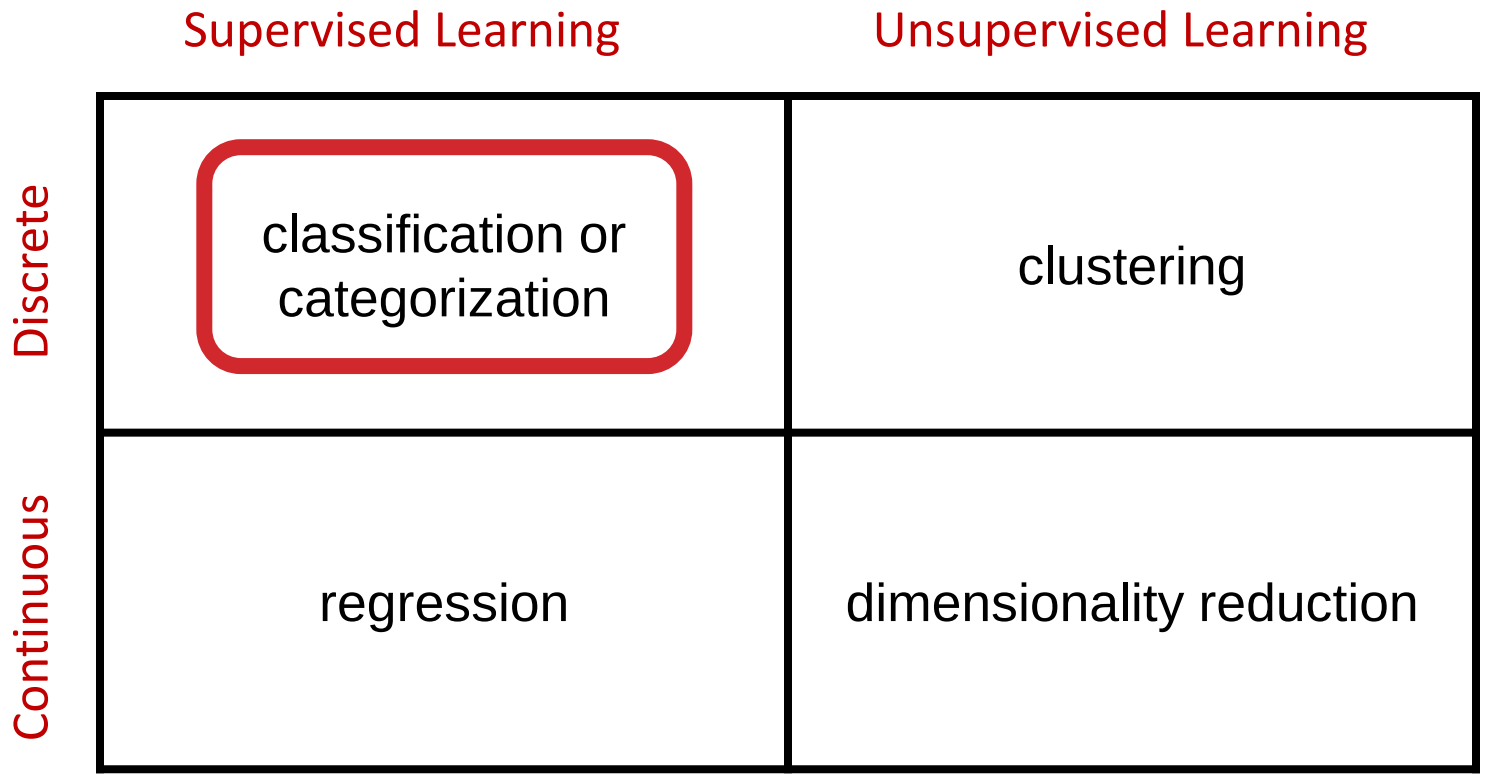
Part 1: More on embeddings and classification/regressions

Part 2: Overview of potential quiz questions

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”

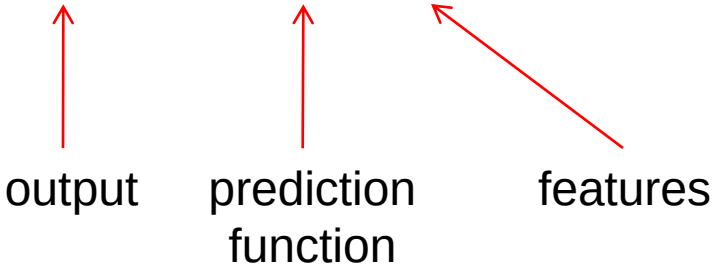
f() = “tomato”

f() = “cow”

Slide credit: L. Lazebnik

THE MACHINE LEARNING FRAMEWORK

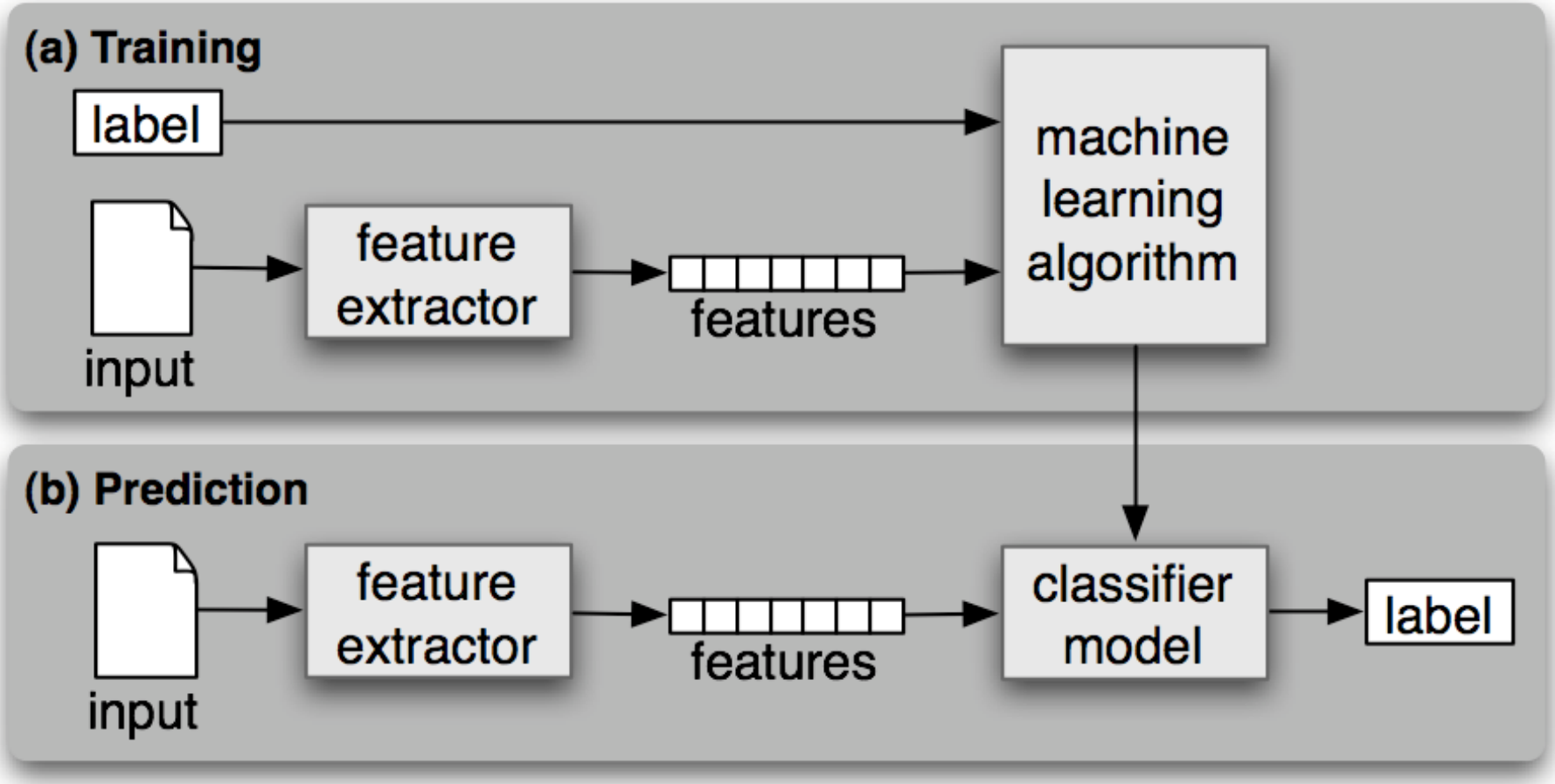
$$y = f(x)$$



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 $y = f(x)$

ML PIPELINE (SUPERVISED)



MANY CLASSIFIERS TO CHOOSE FROM

K-nearest neighbor

Support Vector Machines

Decision Trees

Random Forrest

(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 Forrest

(Gradient) Boosted Decision Trees

Logistic Regression

Naïve Bayes

Bayesian network

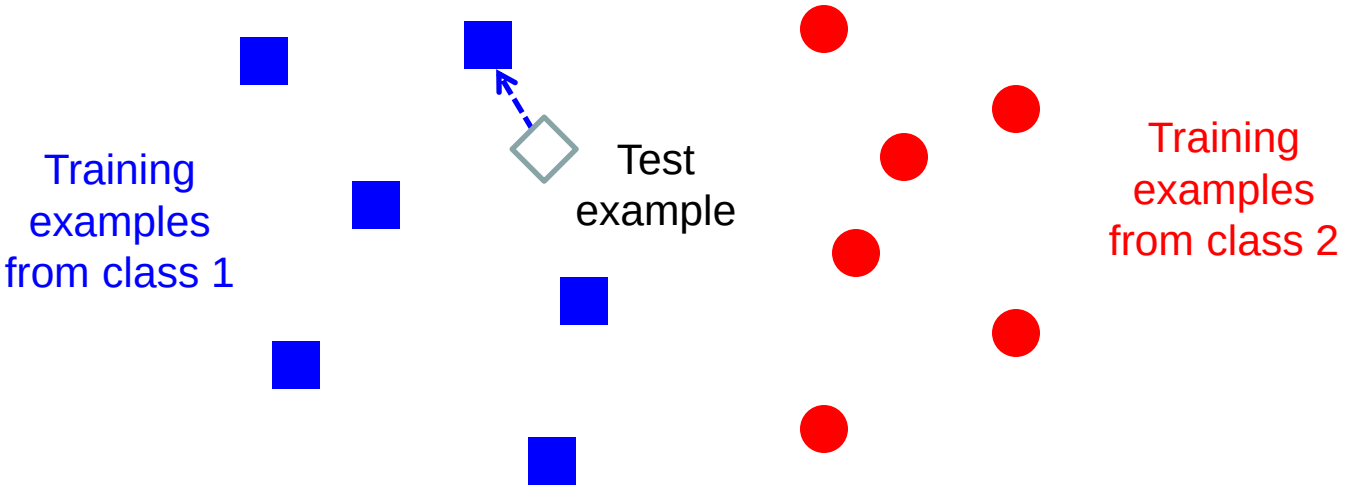
RBMs

….

Which is the best one?

Slide credit: D. Hoiem

CLASSIFIERS: NEAREST NEIGHBOR

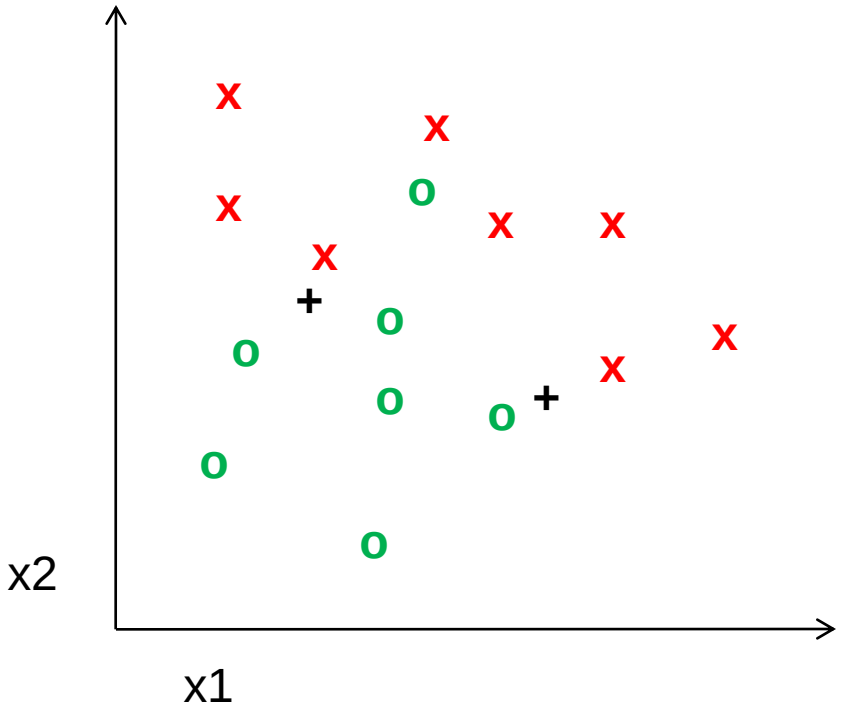


f(**x**) = label of the training example nearest to **x**

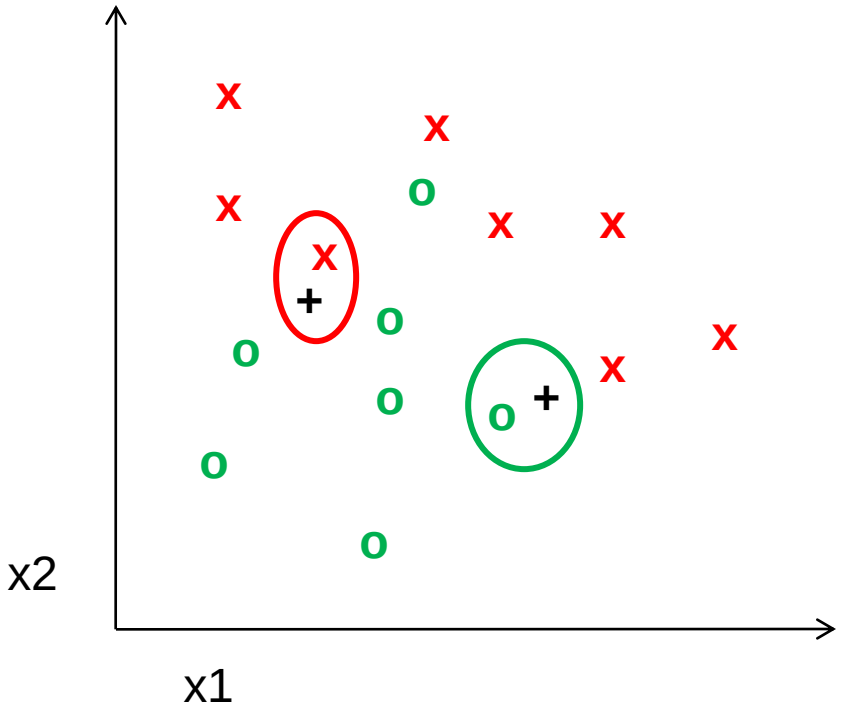
- All we need is a distance function for our inputs
- No training required!

Slide credit: L. Lazebnik

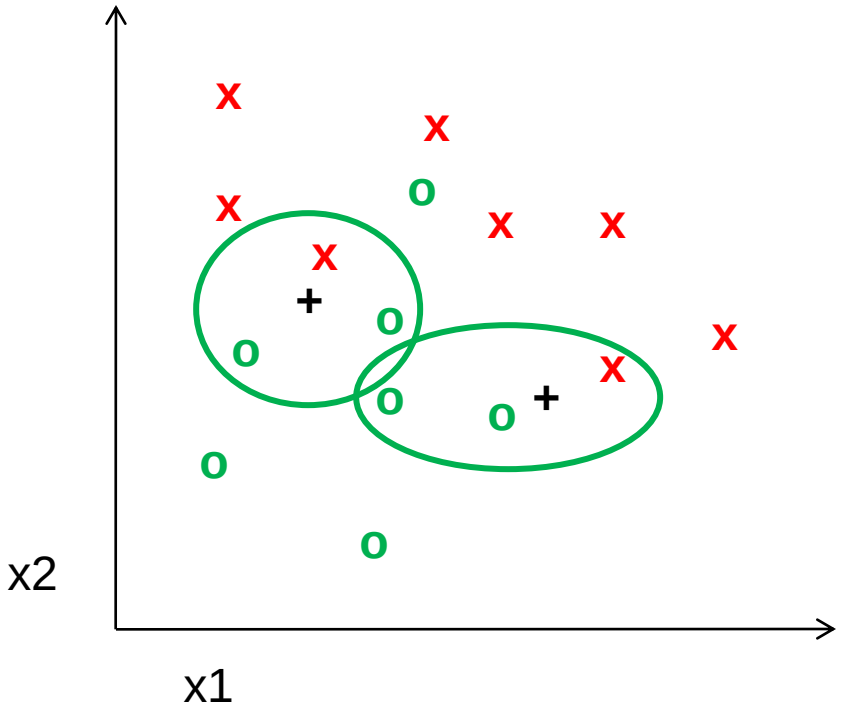
K-NEAREST NEIGHBOR



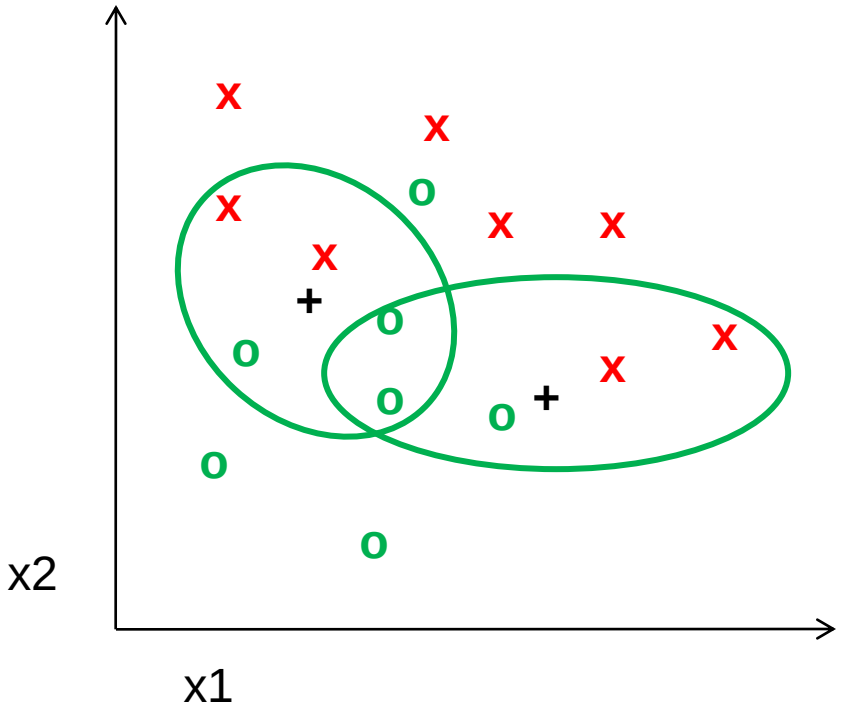
1-NEAREST NEIGHBOR



3-NEAREST NEIGHBOR

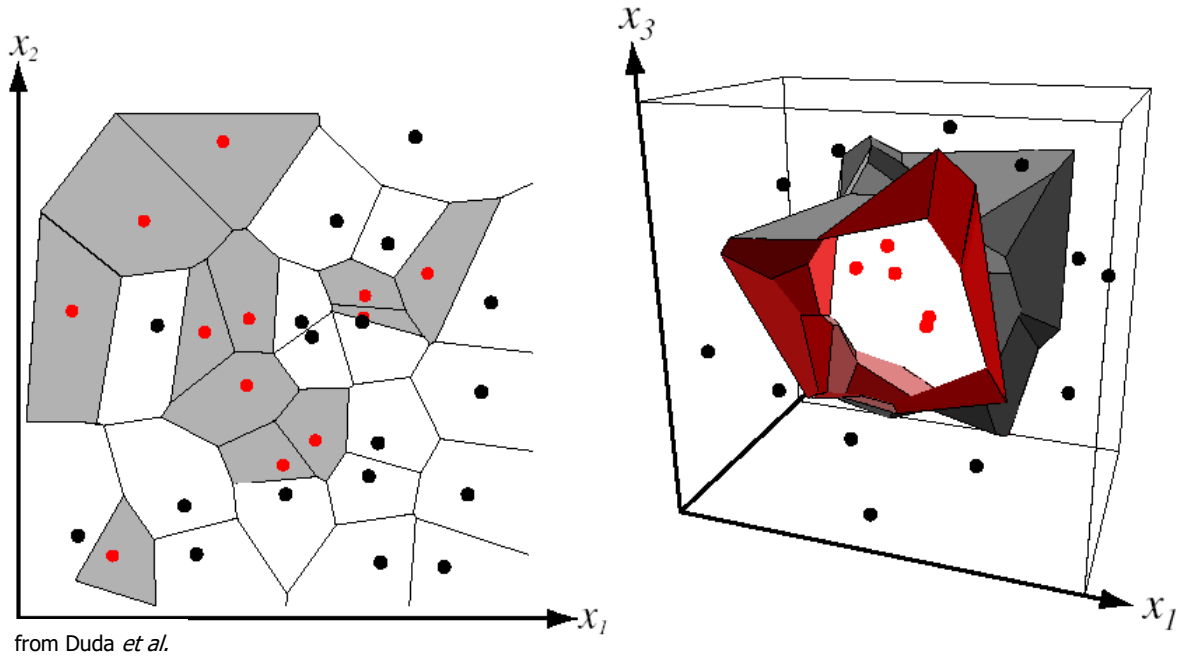


5-NEAREST NEIGHBOR



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 Forrest

(Gradient) Boosted Decision Trees

Logistic Regression

Naïve Bayes

Bayesian network

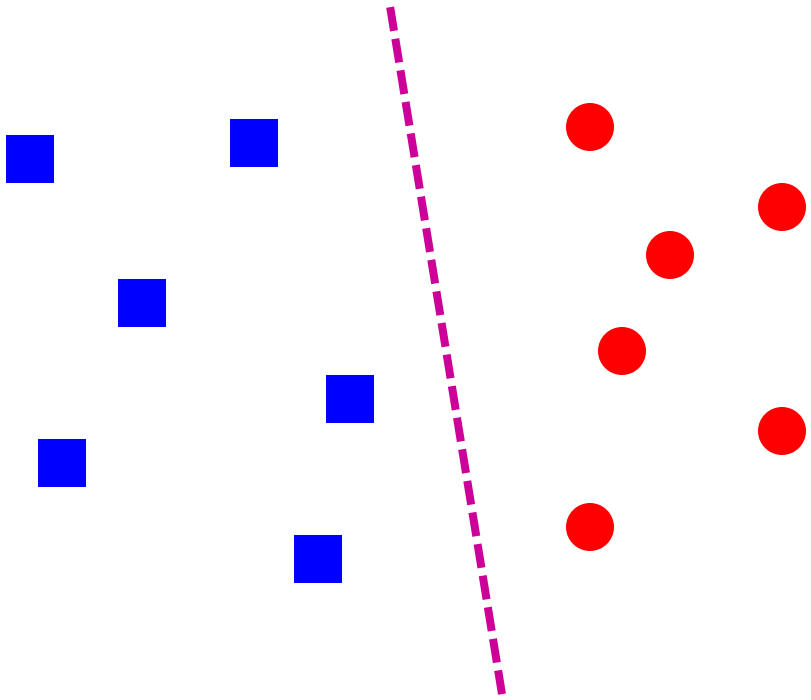
RBMs

….

Which is the best one?

Slide credit: D. Hoiem

CLASSIFIERS: LINEAR

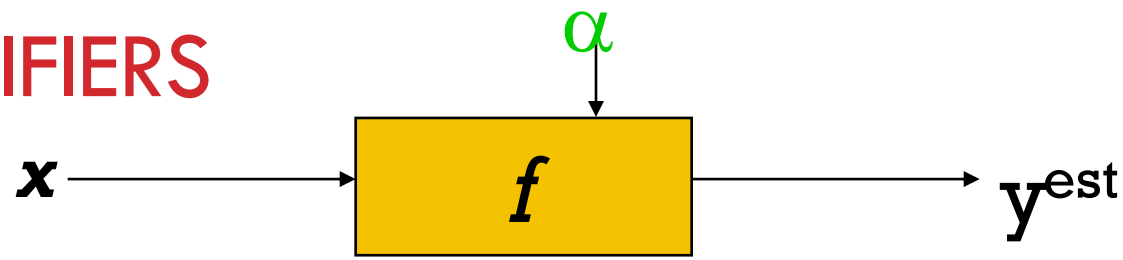


Find a *linear function* to separate the classes:

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

Slide credit: L. Lazebnik

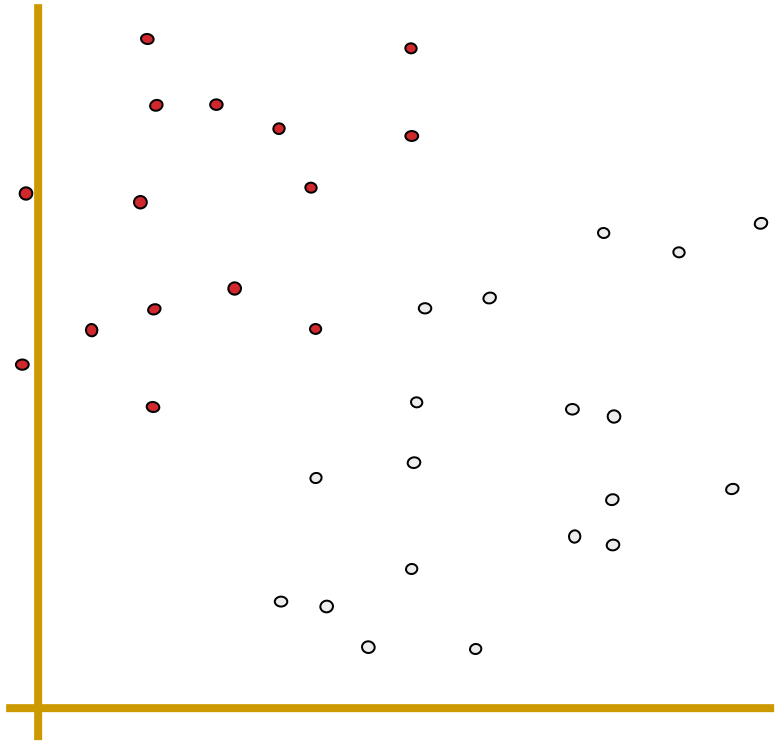
LINEAR CLASSIFIERS



$f(x, w, b) = sign(w \cdot x - b)$

denotes +1

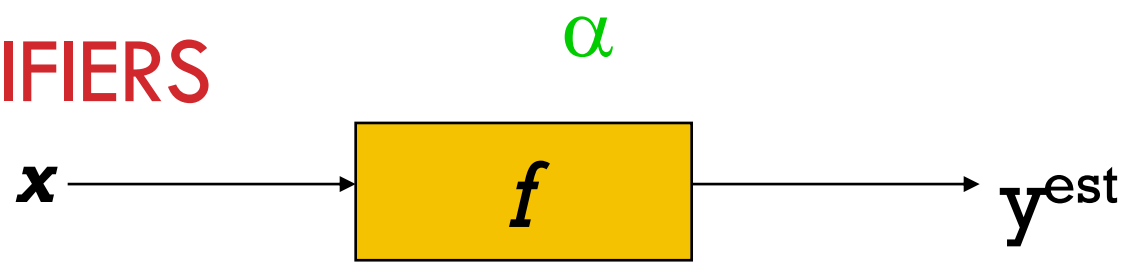
denotes -1



How would you classify this data?

Slides from Andrew W. Moore

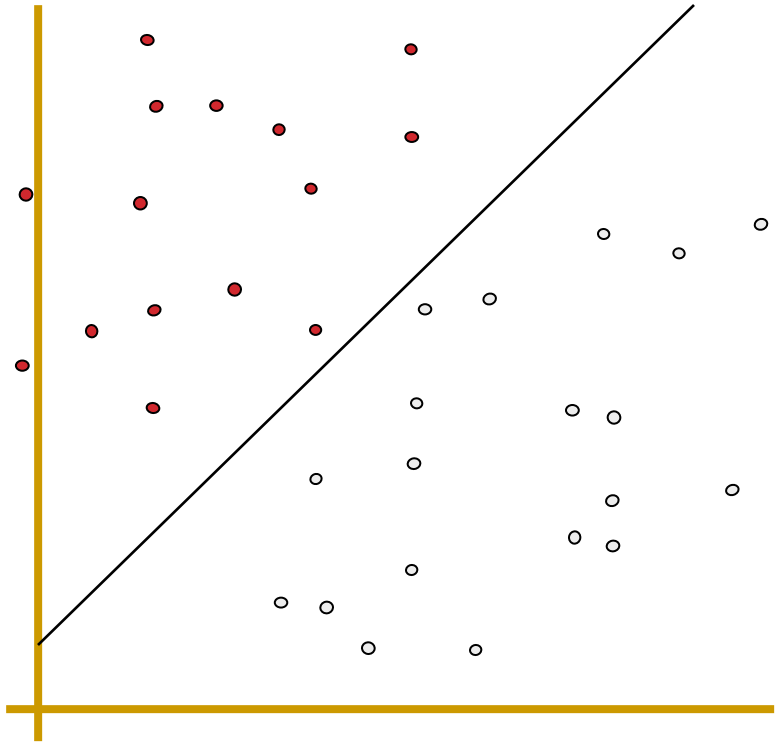
LINEAR CLASSIFIERS



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

denotes +1

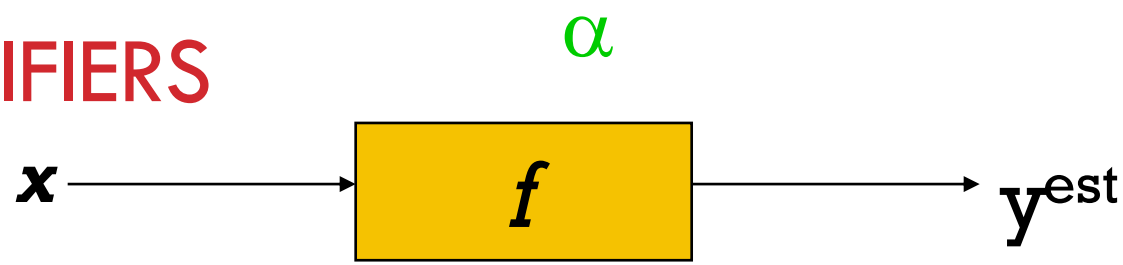
denotes -1



How would you classify this data?

Slides from Andrew W. Moore

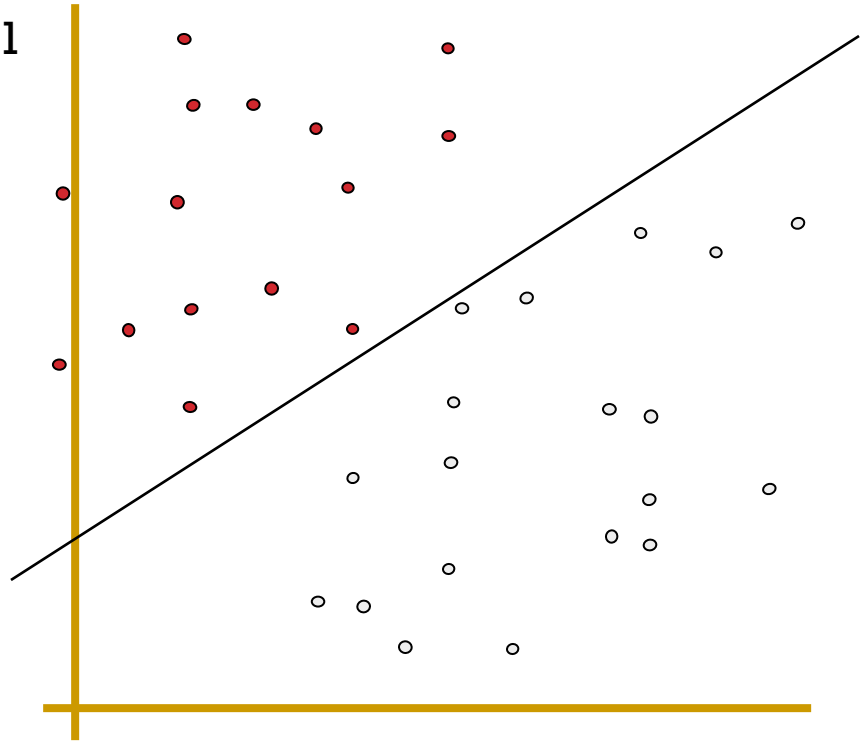
LINEAR CLASSIFIERS



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

denotes +1

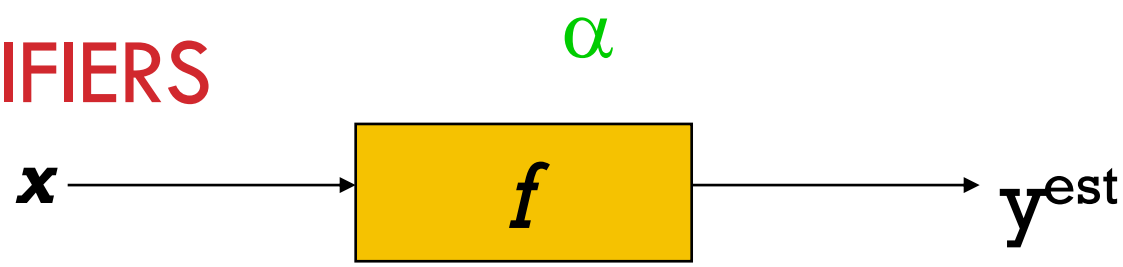
denotes -1



How would you classify this data?

Slides from Andrew W. Moore

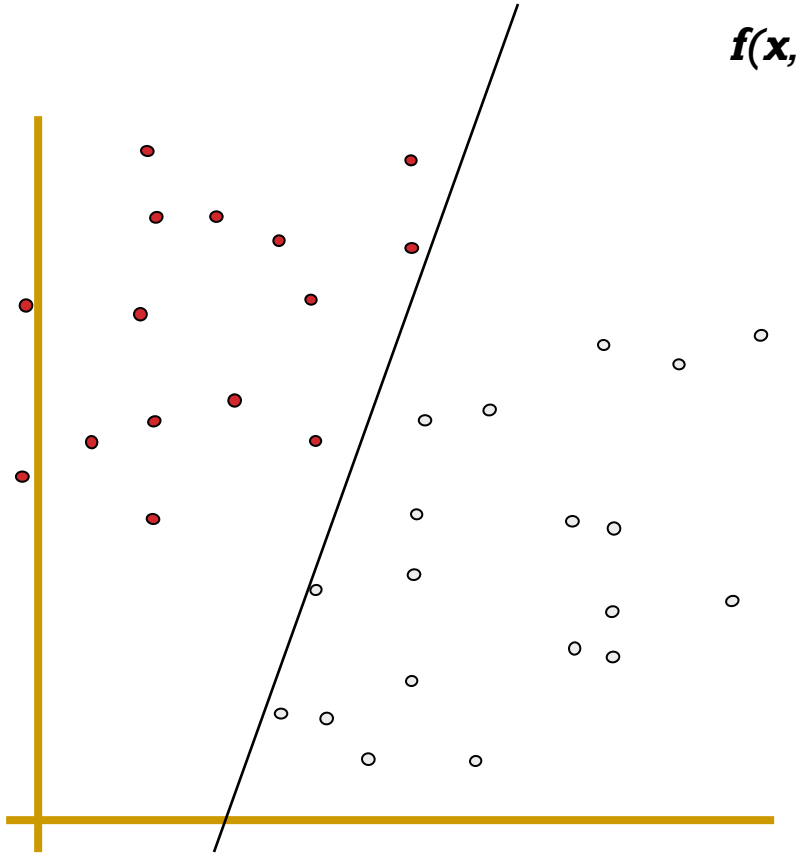
LINEAR CLASSIFIERS



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

denotes +1

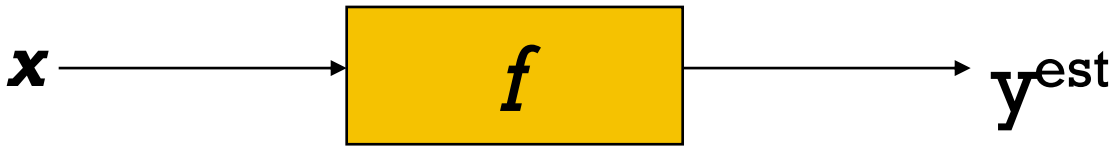
denotes -1

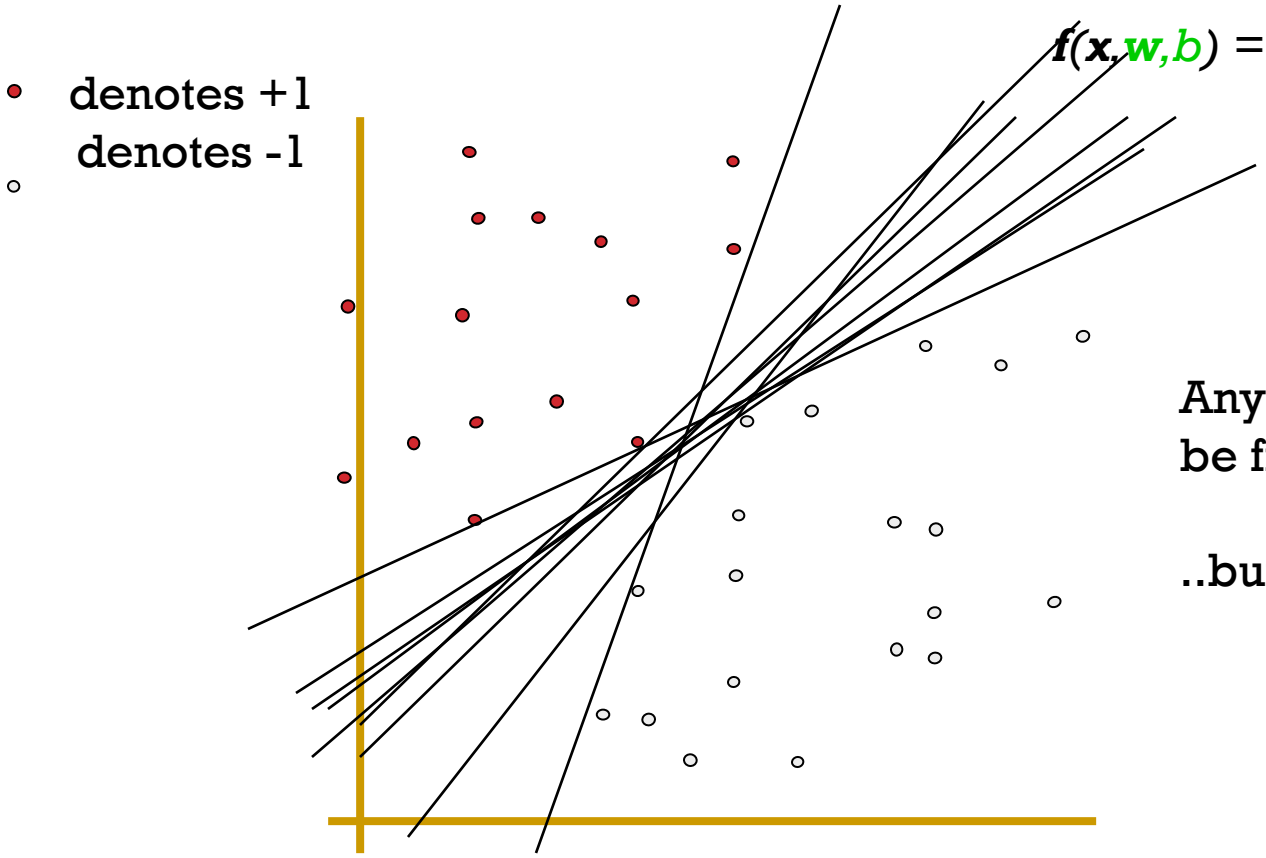


How would you classify this data?

Slides from Andrew W. Moore

LINEAR CLASSIFIERS



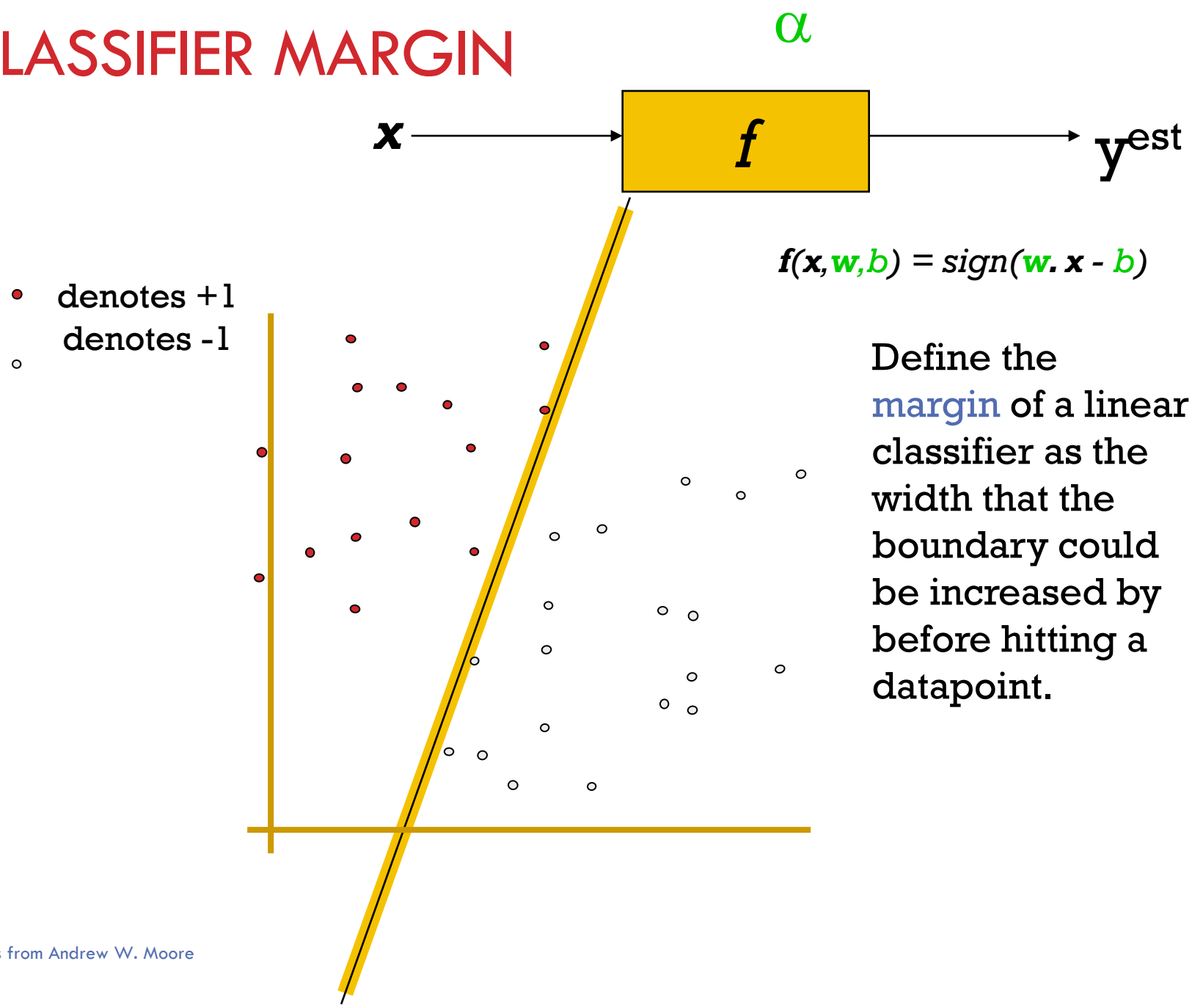


Any of these would be fine..

..but which is best?

Slides from Andrew W. Moore

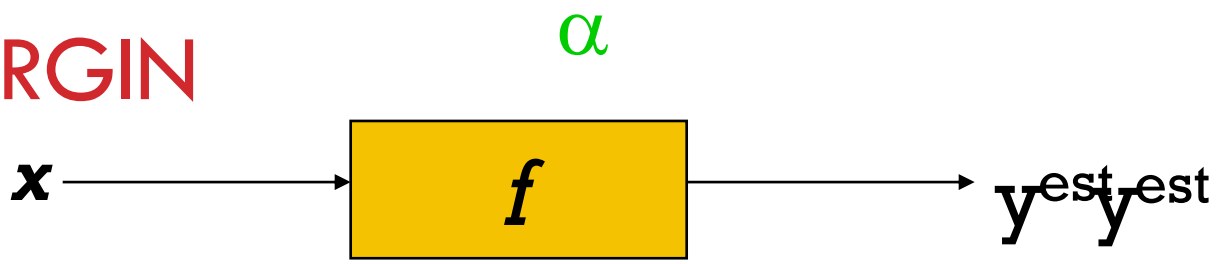
CLASSIFIER MARGIN

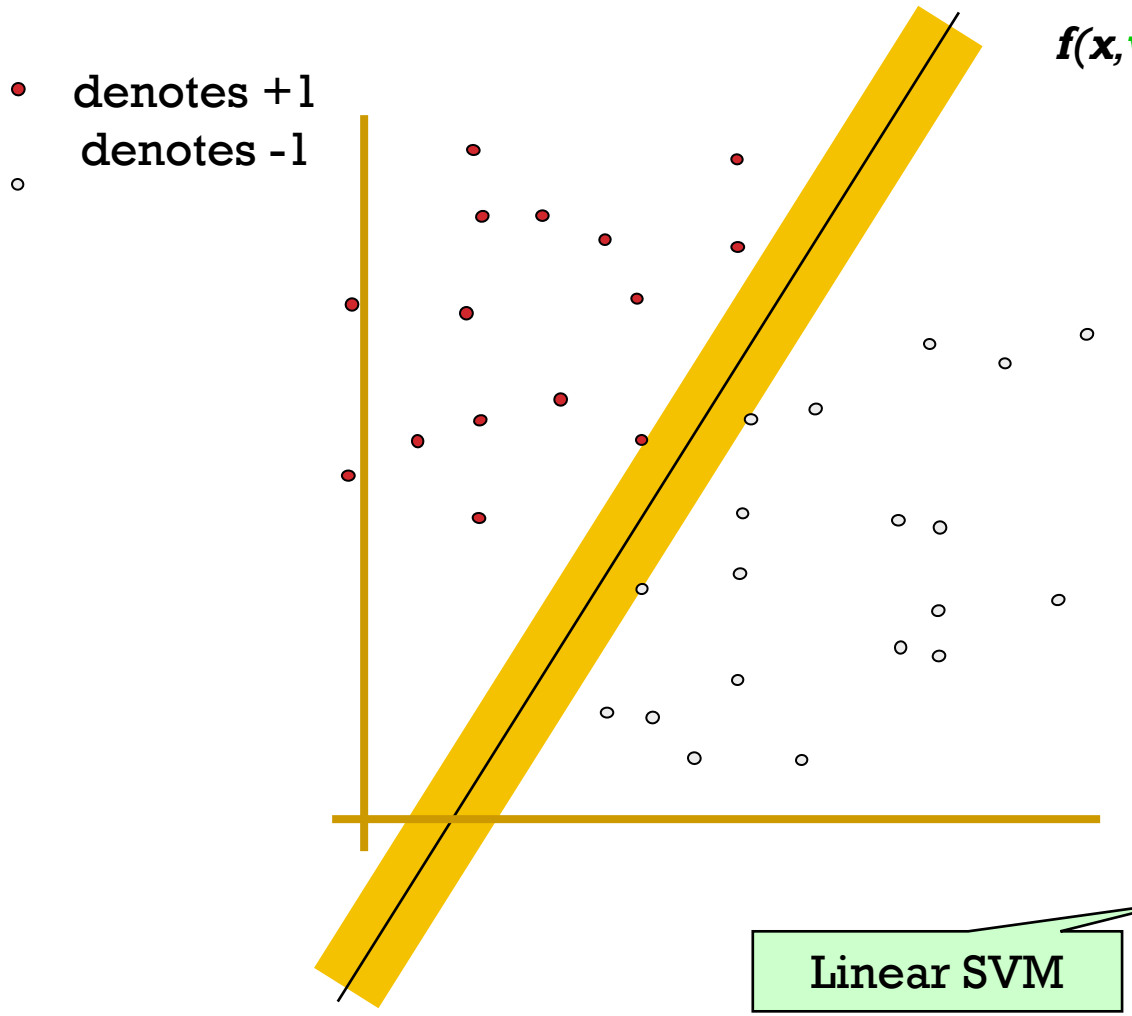


Define the margin of a linear classifier as the width that the boundary could be increased by before hitting a datapoint.

Slides from Andrew W. Moore

MAXIMUM MARGIN





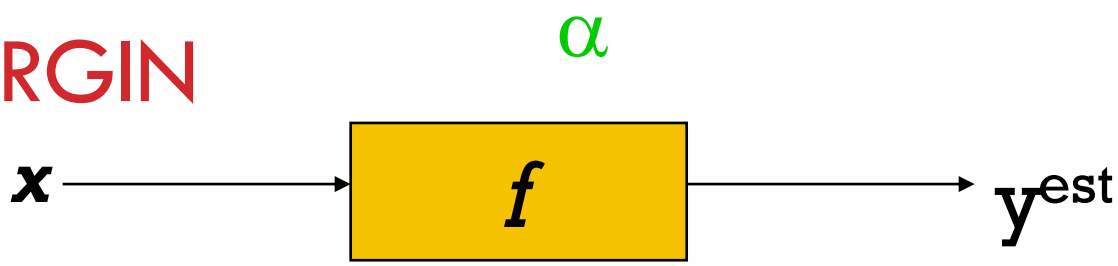
$f(x,w,b) = sign(w.x - b)$

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



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

denotes +1
denotes -1

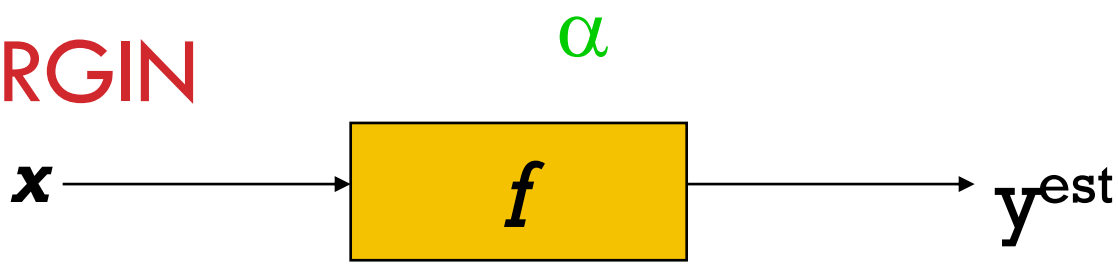
Support Vectors are those datapoints that the margin pushes up against

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



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

denotes +1

denotes -1

The maximum

ear

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 Forrest

(Gradient) Boosted Decision Trees

Logistic Regression

Naïve Bayes

Bayesian network

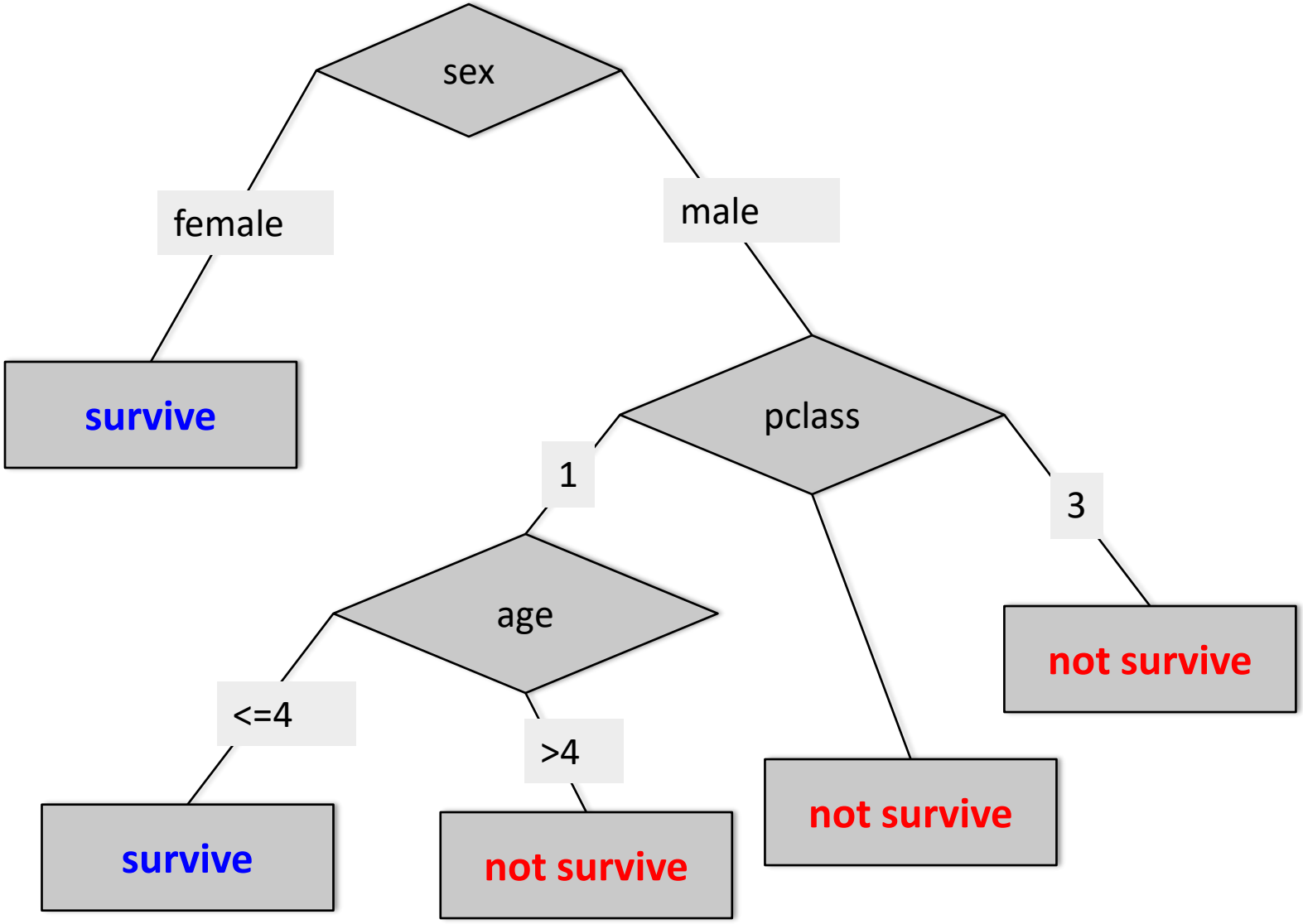
RBMs

….

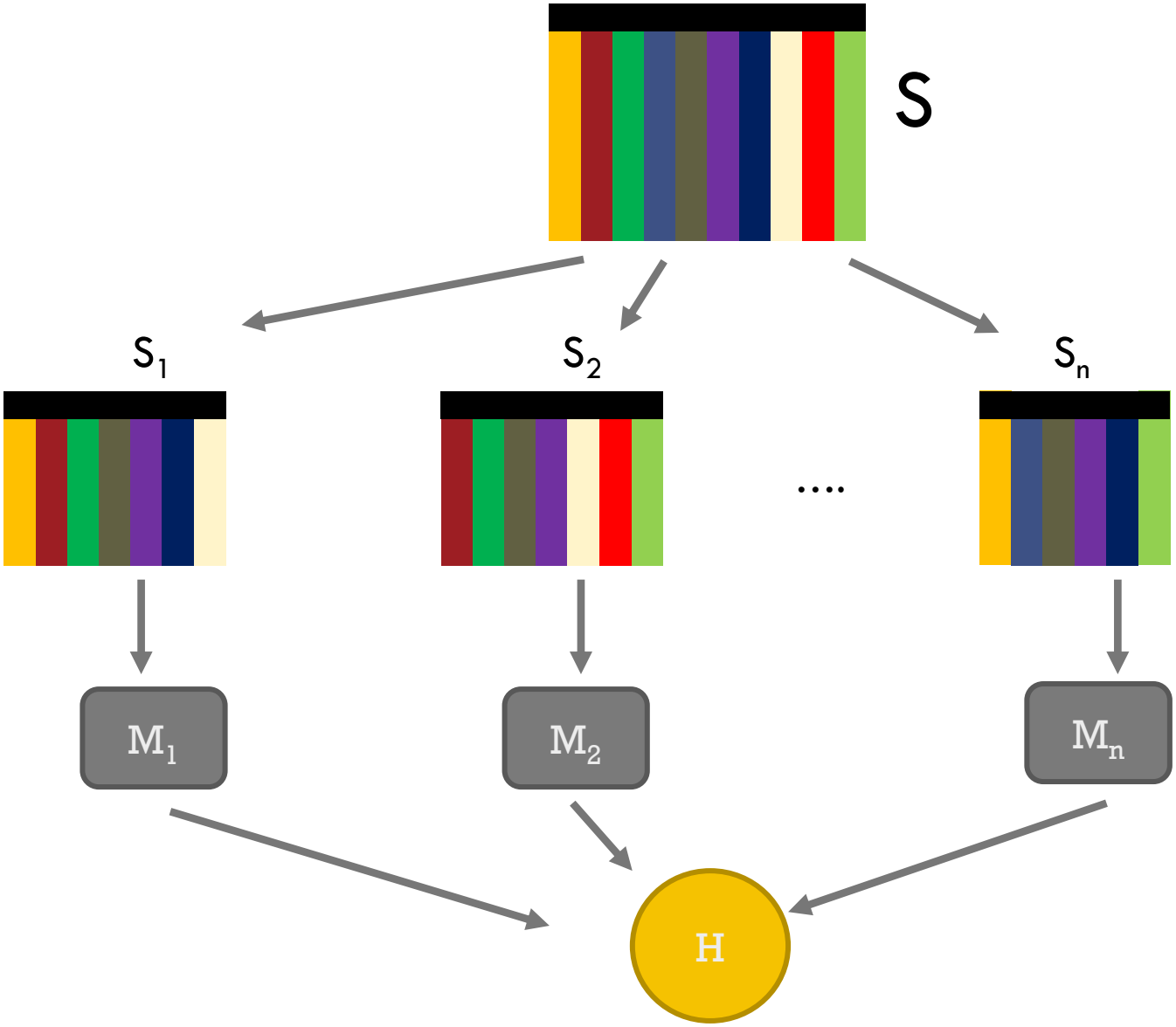
Which is the best one?

Slide credit: D. Hoiem

DECISION TREES



RANDOM FORREST



Sample with Replacement & select random subset of features*

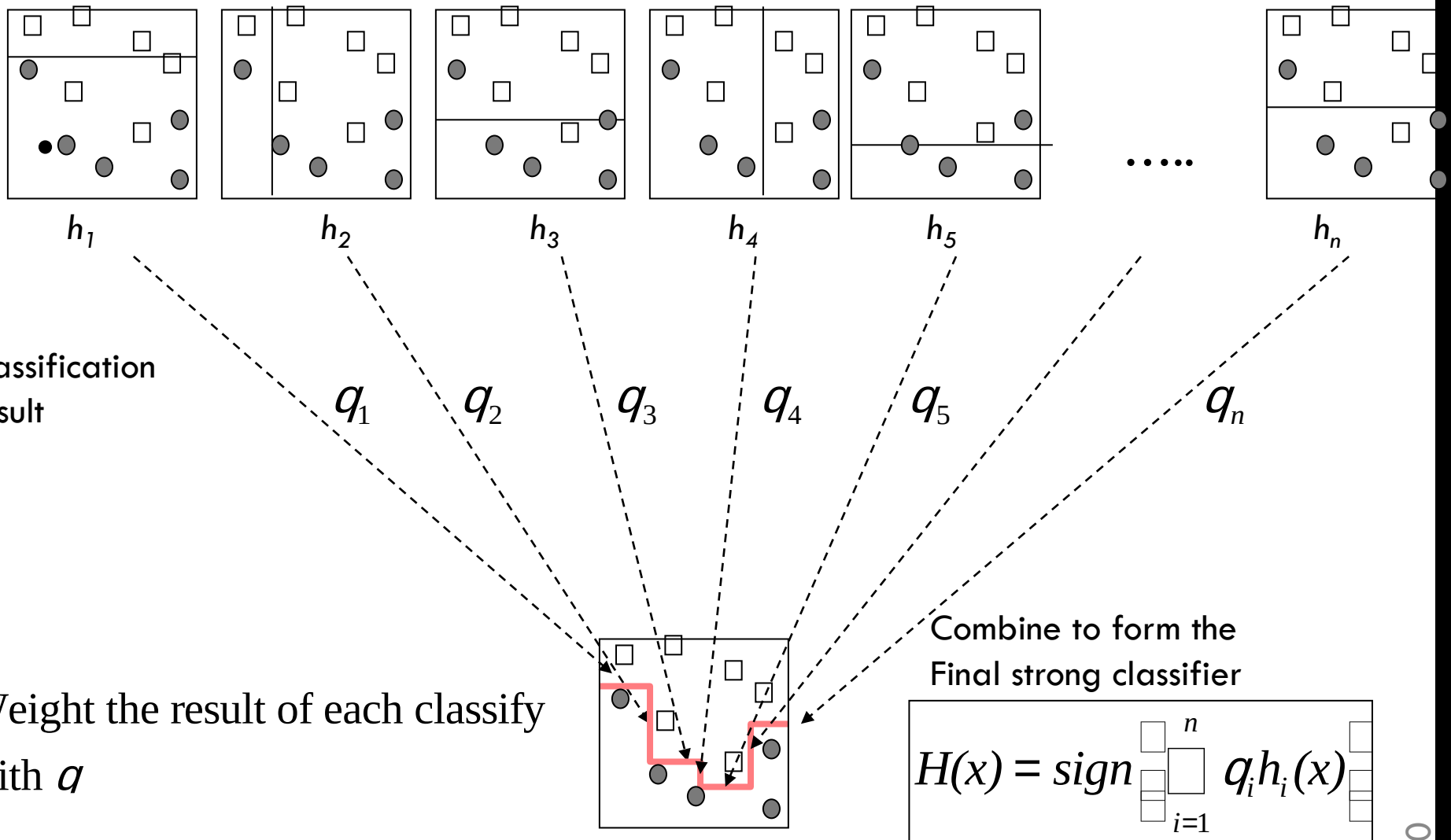
Build classifier over sample

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

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

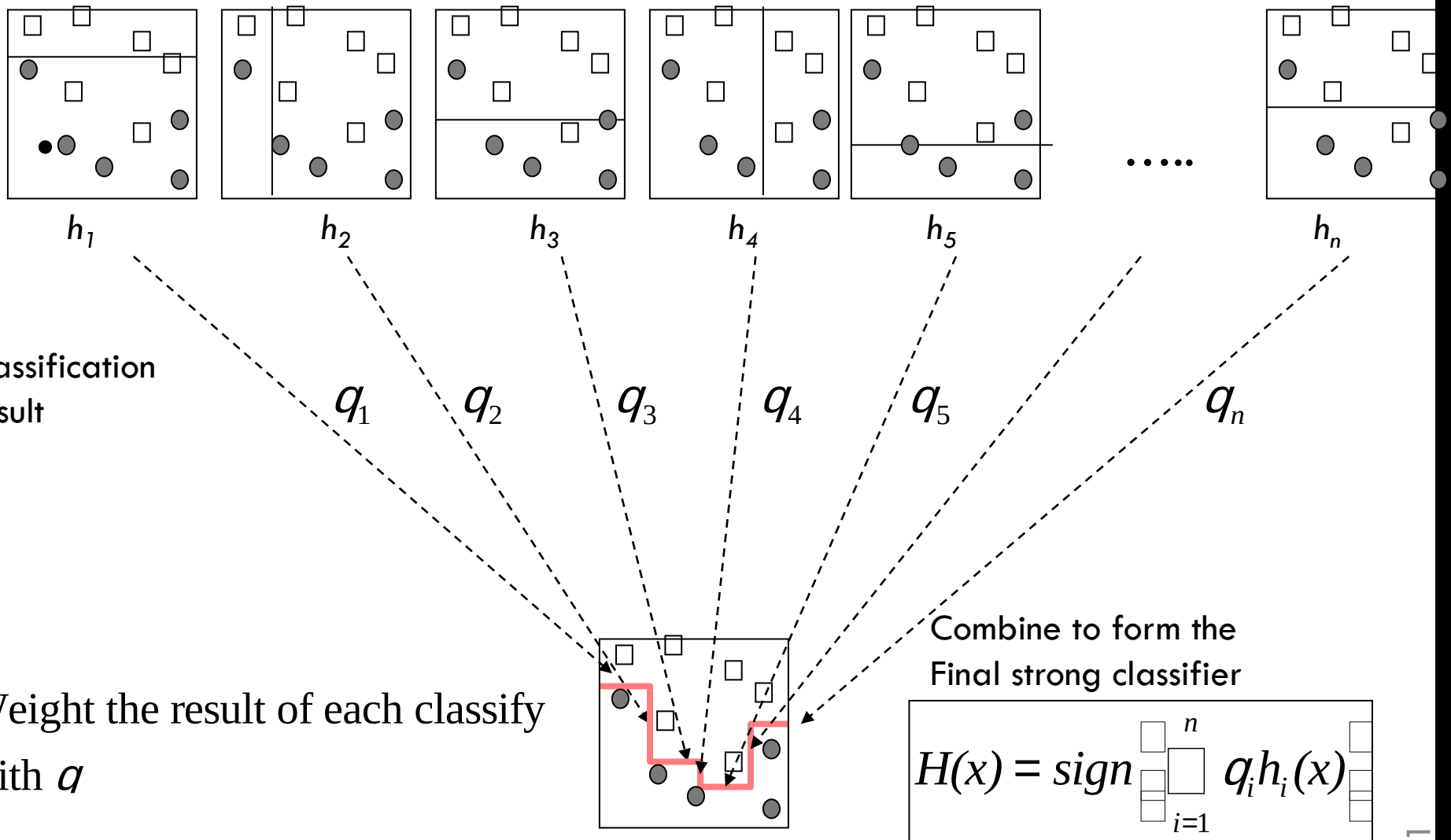
ADABOOST - CORE IDEA

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



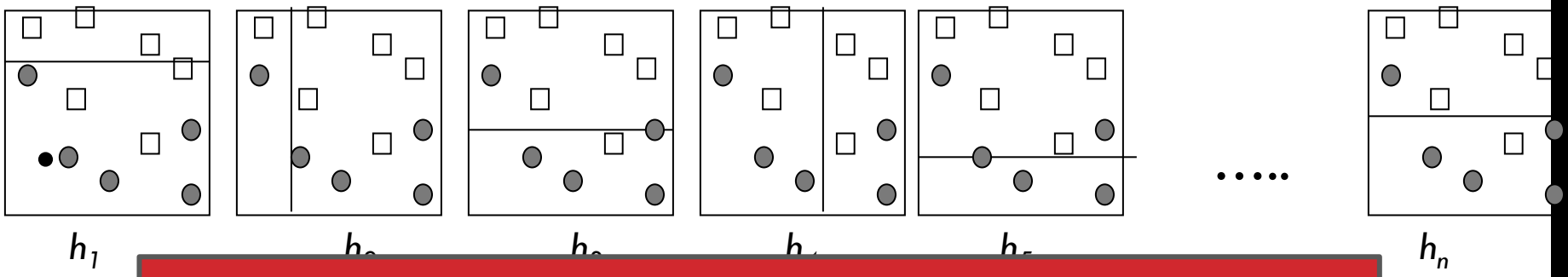
ADABOOST - CORE IDEA

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



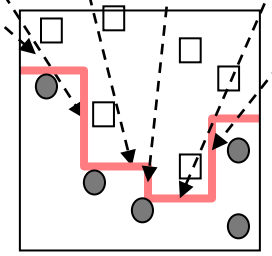
ADABOOST - CORE IDEA

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



Classificatio
Result

XGBoost follows the same idea

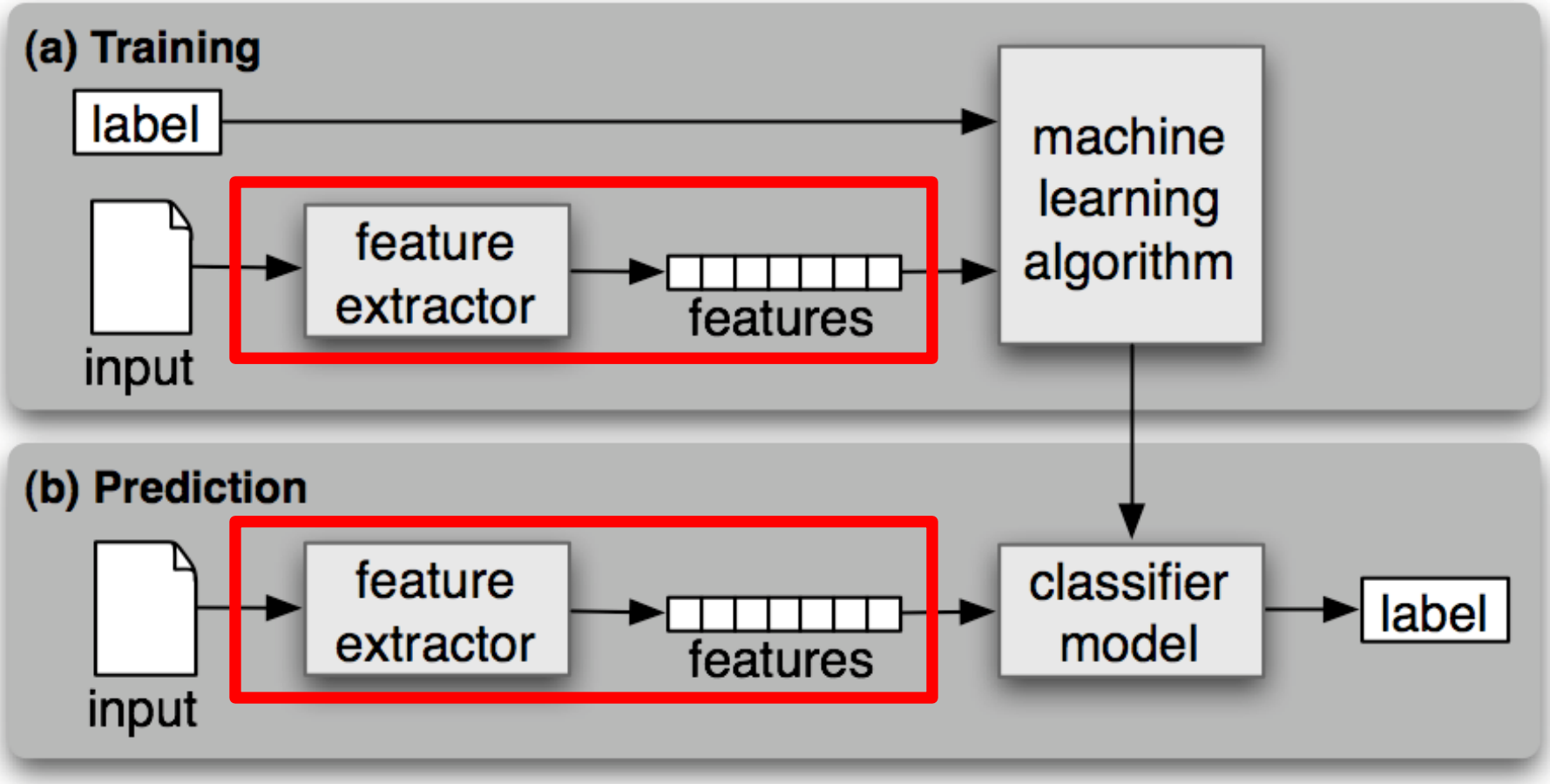


Weight the result of each classify with α

Combine to form the Final strong classifier

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

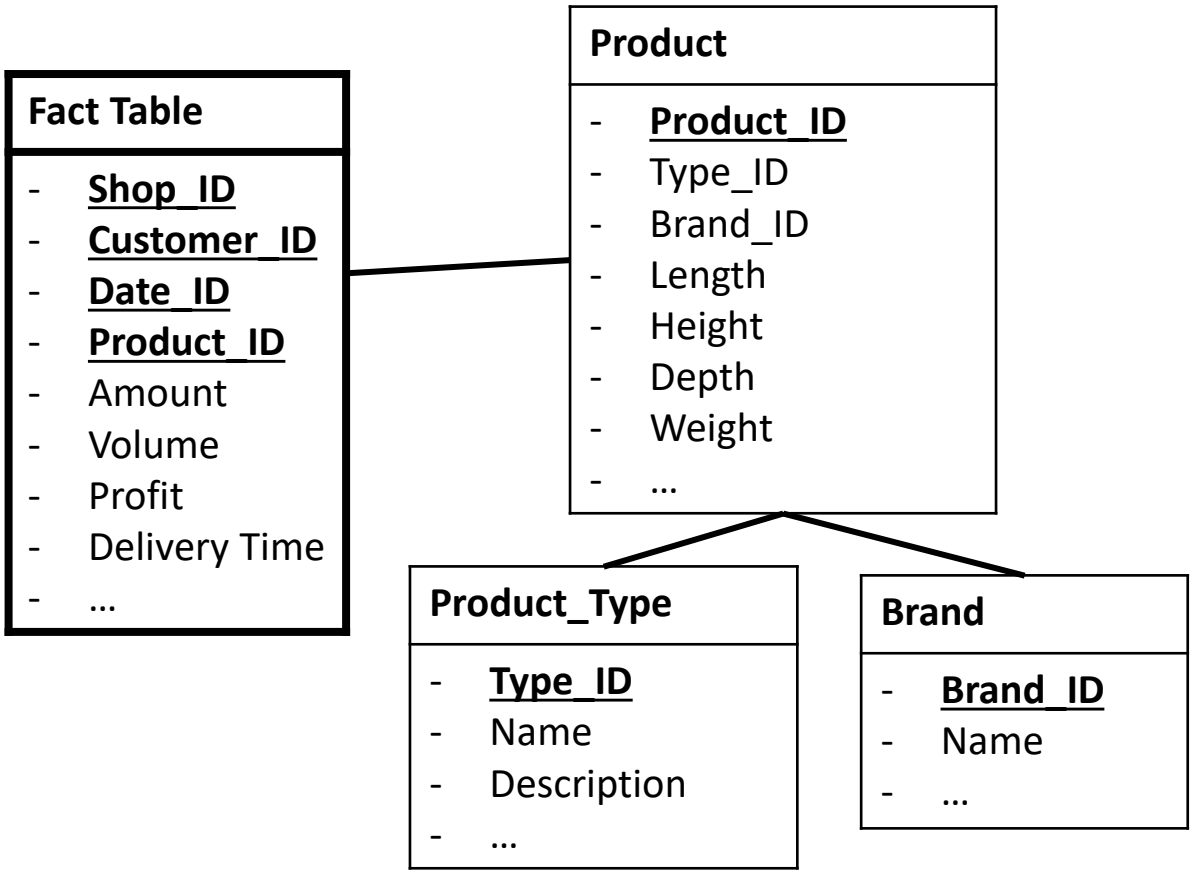
ML PIPELINE (SUPERVISED)



FEATURES

Fact Table

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

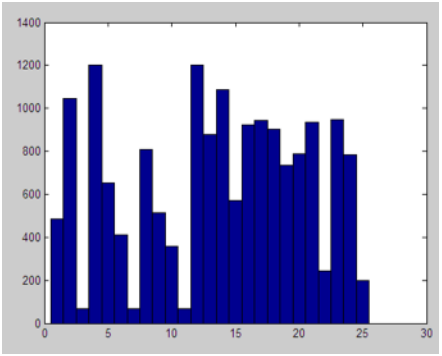


Custermer State	Product Type	Product Weight	Volume (L*H*D)	Month	Delivery Time

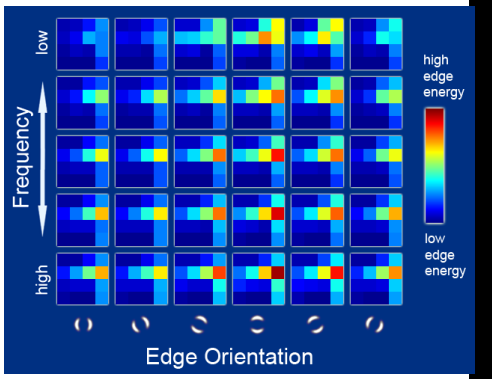
IMAGE FEATURES

Raw pixels

Histograms



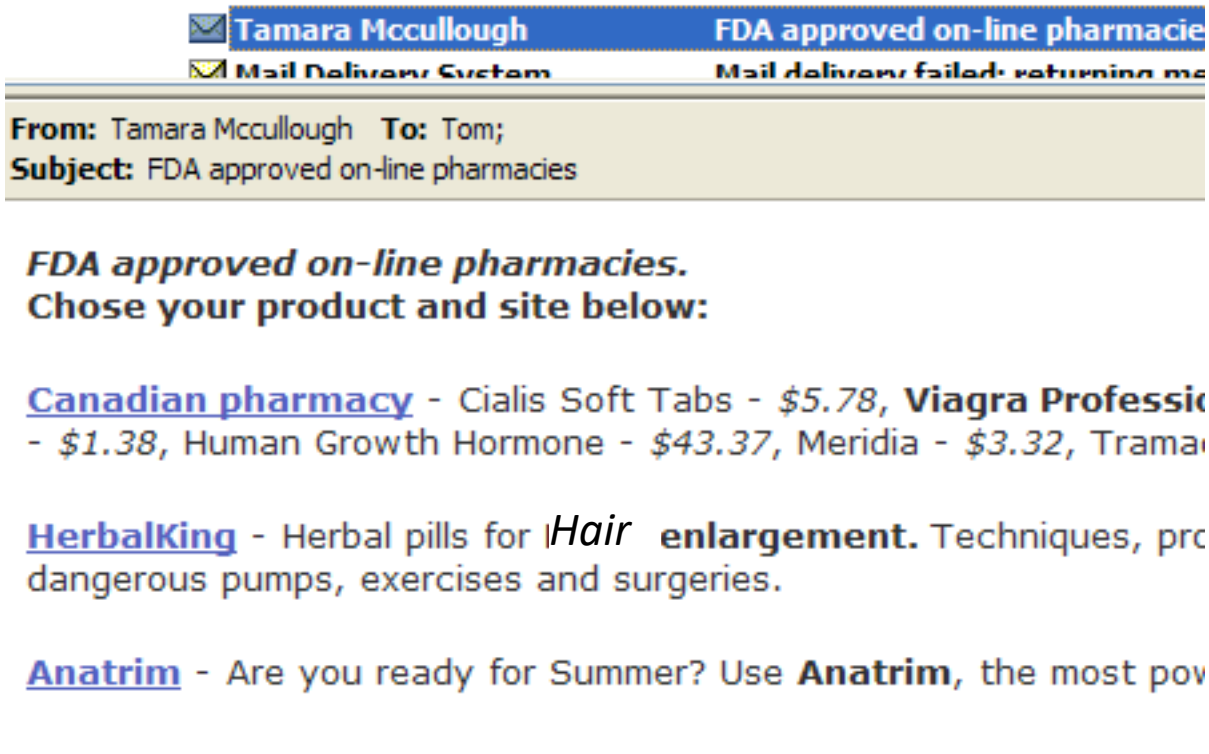
GIST descriptors

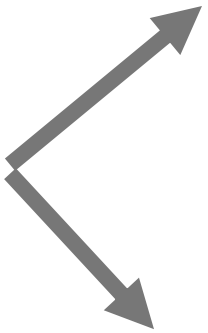


…

Slide credit: L. Lazebnik

TEXT FEATURES





Spam

Not Spam

Bag of Words

$$\begin{pmatrix} Viagra: 1 \\ Soft: 1 \\ Herbel: 2 \\ Pills: 2 \\ Are: 1 \\ \dots \end{pmatrix}$$

N-Grams

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

ONE-HOT ENCODING

Bag of Words

$$\begin{pmatrix} Viagra \\ Soft \\ Herbel \\ Pills \\ Are \\ \dots \end{pmatrix}$$

ID	Viagra	Soft	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 GR[illegible]OOL APPLIC[illegible]

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
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…