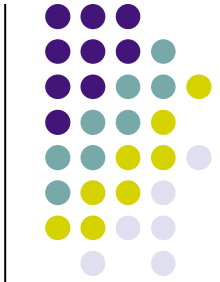
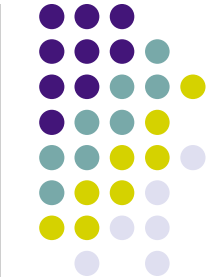


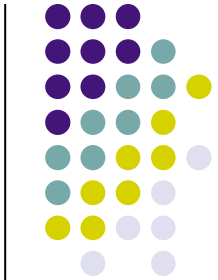
Supervised Learning Evaluation (via Sentiment Analysis)

Independence: Intuition



- Events are independent if one has nothing whatever to do with others. Therefore, for two independent events, knowing one happening does not change the probability of the other event happening.
  - one toss of coin is independent of another coin toss (assuming it is a regular coin).
  - price of tea in England is independent of the result of general election in Canada.

Independence: Definition



- Events A and B are independent iff:

$$P(A, B) = P(A) \times P(B)$$

which is equivalent to

$$P(A|B) = P(A) \text{ and}$$

$$P(B|A) = P(B)$$

when $P(A, B) > 0$.

T1: the first toss is a head.

T2: the second toss is a tail.

$$P(T2|T1) = P(T2)$$

Conditional Independence

- Dependent events can become independent given certain other events.
- Example,
  - Size of shoe
  - Size of vocabulary
  - ??
- Two events A, B are conditionally independent given a third event C iff

  $P(A|B, C) = P(A|C)$

Conditional Independence: Definition

- Let E1 and E2 be two events, they are conditionally independent given E iff

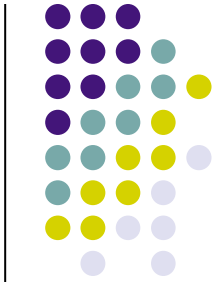
  $P(E1|E, E2)=P(E1|E)$,

  that is the probability of E1 is not changed after knowing E2, given E is true.
- Equivalent formulations:

  $P(E1, E2|E)=P(E1|E)\ P(E2|E)$

  $P(E2|E, E1)=P(E2|E)$

Bayes' Rule and conditional independence



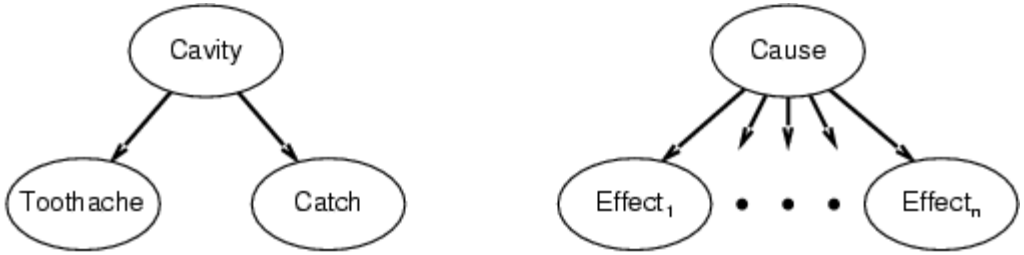
P(*Cavity* | *toothache* ∧ *catch*)

= α**P**(*toothache* ∧ *catch* | *Cavity*) **P**(*Cavity*)

= α**P**(*toothache* | *Cavity*) **P**(*catch* | *Cavity*) **P**(*Cavity*)

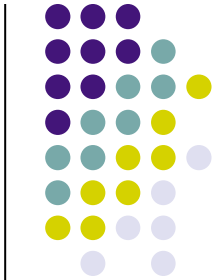
- This is an example of a naïve Bayes model:

  P(Cause,Effect$_1$, … ,Effect$_n$) = **P**(Cause) π_i**P**(Effect$_i$|Cause)



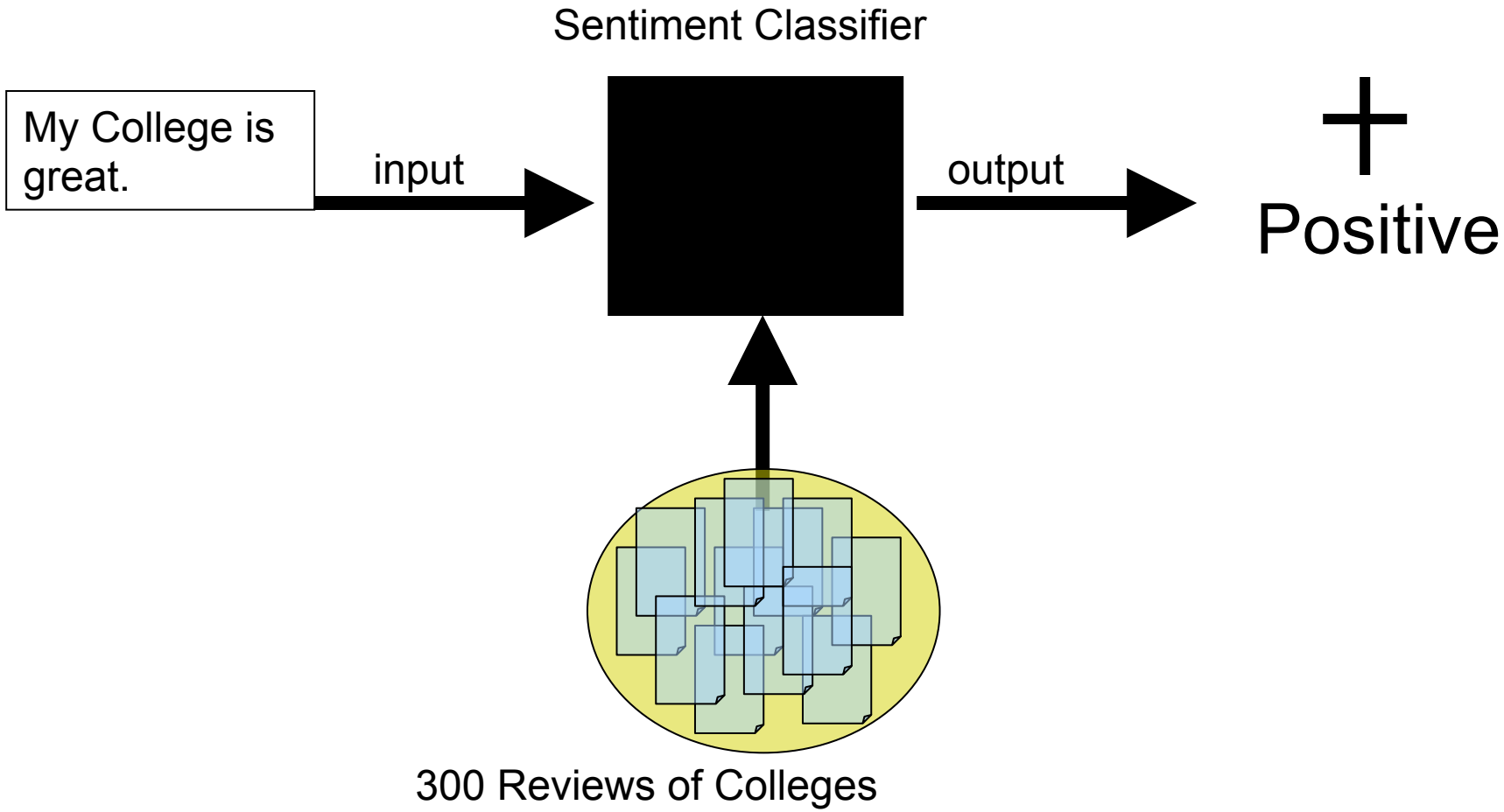
- Total number of parameters is linear in *n*

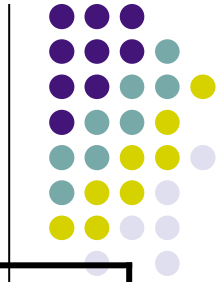
Summary



- Probability is a rigorous formalism for uncertain knowledge
- Joint probability distribution specifies probability of every atomic event
- Queries can be answered by summing over atomic events
- For nontrivial domains, we must find a way to reduce the joint size
- Independence and conditional independence provide the tools

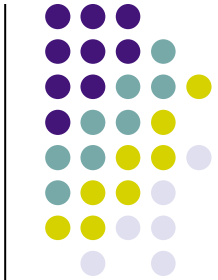
Naïve Bayes Classification





Words	**Positive Doc. Count**	**Negative Doc. Count**	**Neutral Doc. Count**
my	6	5	5
college	100	100	100
great	40	1	2
the	100	100	100
bad	2	30	2
is	98	99	98
Total count	5000	5000	5000

P(pos|features)



= P(pos)* product of probabilities P(feature|pos)

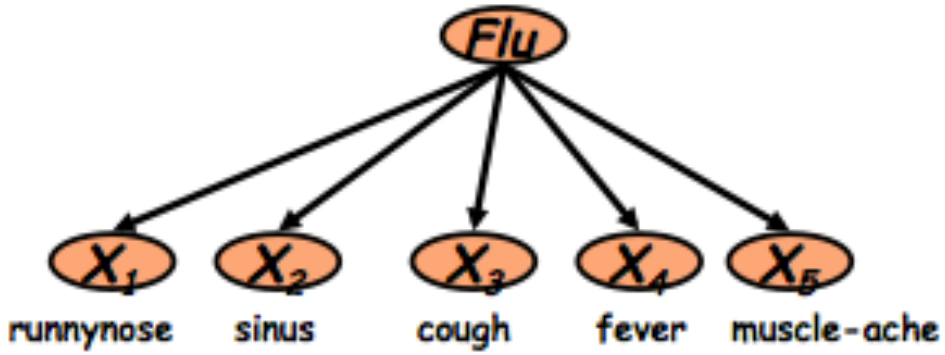
= P(pos) * P(“my”|pos) * P(“college”|pos)
* P(“is”|pos) * P(“great”|pos)

=0.333 * 6/5000 * 100/5000 * 98/5000 * 40/5000

Naïve Bayes Classifier: Naïve Bayes Assumption

- $P(c_j)$
  - Estimated from the frequency of classes in the training examples
- $P(x_1, x_2, \ldots, x_n|c_j)$
  - Could only be estimated from a very large number of training examples
- Naïve Bayes Conditional Independence Assumption:
  - Assume that the prob of observing the conjunction of attributes is equal to the product of the indiv probs $P(x_i|c_j)$

The Naive Bayes Classifier



- **Conditional Independence Assumption:** features are **independent** of each other **given the class**:

$$P(X_1,\ldots,X_5 \mid C) = P(X_1 \mid C) \bullet P(X_2 \mid C) \bullet \cdots \bullet P(X_5 \mid C)$$

Why Analyze Sentiment?

Customer Reviews

2,162 Reviews

5 star:		(1,458)
4 star:		(368)
3 star:		(122)
2 star:		(69)
1 star:		(145)

Average Customer Review

★★★★½ (2,162 customer reviews)

Most Helpful Customer Reviews

4,448 of 4,653 people found the following review helpful:

★★★★★ **The Lines Between iPod Touch and iPhone Have Started to Blur**, September 7, 2010

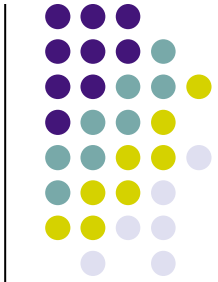
By **Scott Showalter "purefusion"** (Ohio, USA) - See all my reviews

TOP 1000 REVIEWER REAL NAME

This review is from: Apple iPod touch 32GB (4th Generation) - Black - Current Version (Electronics)

Having had a chance to spend a little time with a review model gives me a chance to share the experience with you a bit early (before my own arrives). I'll take you hands-on with the new model, plus I'll share from

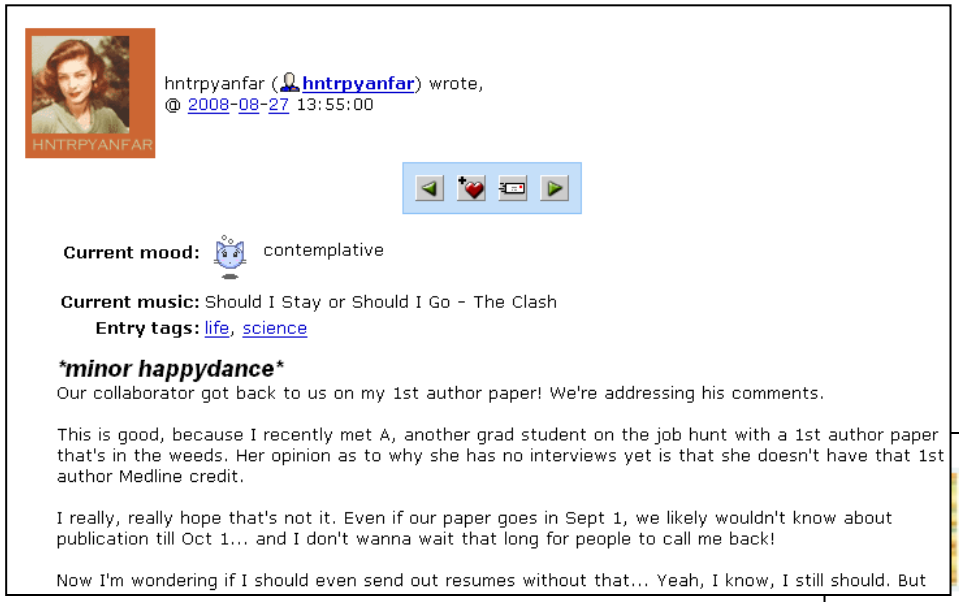
Sentiment Analysis (Opinion Mining)



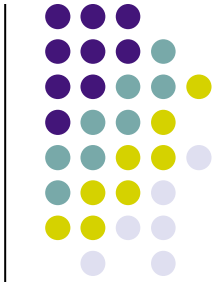
Automatically label documents with their ‘sentiment’

- Toward a topic
- Aggregated over documents
- More fine-grained analysis
- Within specific domains

Sentiment Analysis - Approaches



What are the challenges to Sentiment Analysis?



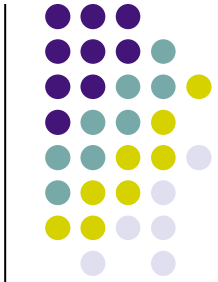
- **domain specificity**
- “thwarted expectations”
- sarcasm and subtle nature of sentiment
- sufficient, high quality training data

What are the challenges to Sentiment Analysis?

Cold

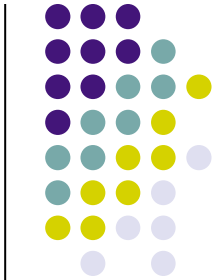
Small

What are the challenges to Sentiment Analysis?



- domain specificity
- **“thwarted expectations”**
- sarcasm and subtle nature of sentiment
- sufficient, high quality training data

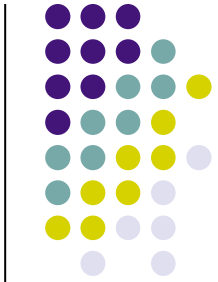
What are the challenges to Sentiment Analysis?



- domain specificity
- **“thw**
- sarc
- suffi

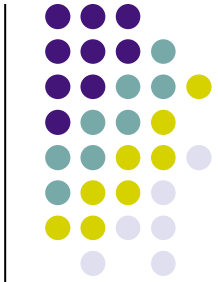
“This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can’t hold up.” (Pang et al, 2002)

What are the challenges to Sentiment Analysis?



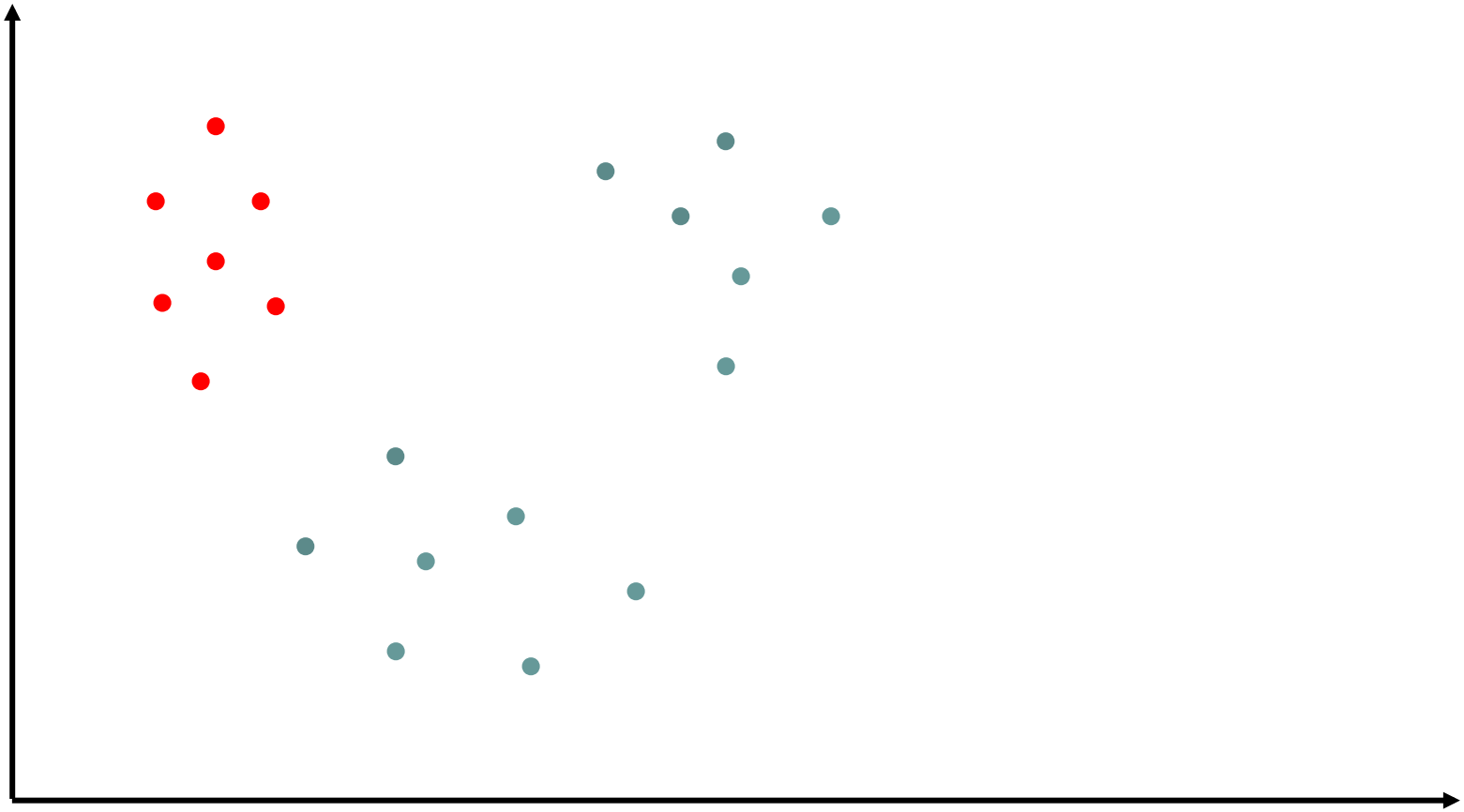
- domain specificity
- “thwarted expectations”
- **sarcasm and subtle nature of sentiment**
- sufficient, high quality training data

What are the challenges to Sentiment Analysis?

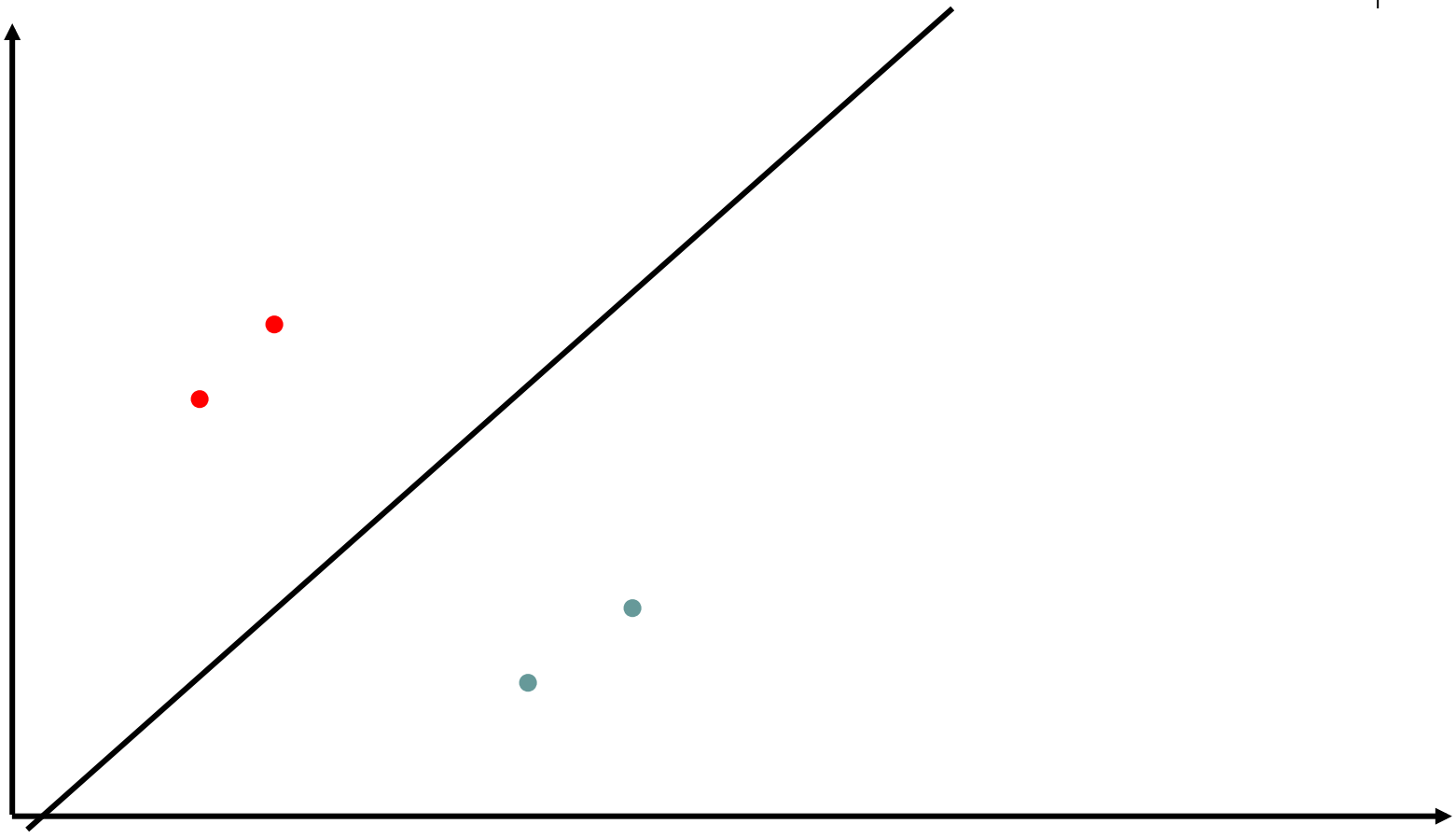


- domain specificity
- “thwarted expectations”
- sarcasm and subtle nature of sentiment
- **sufficient, high quality training data**

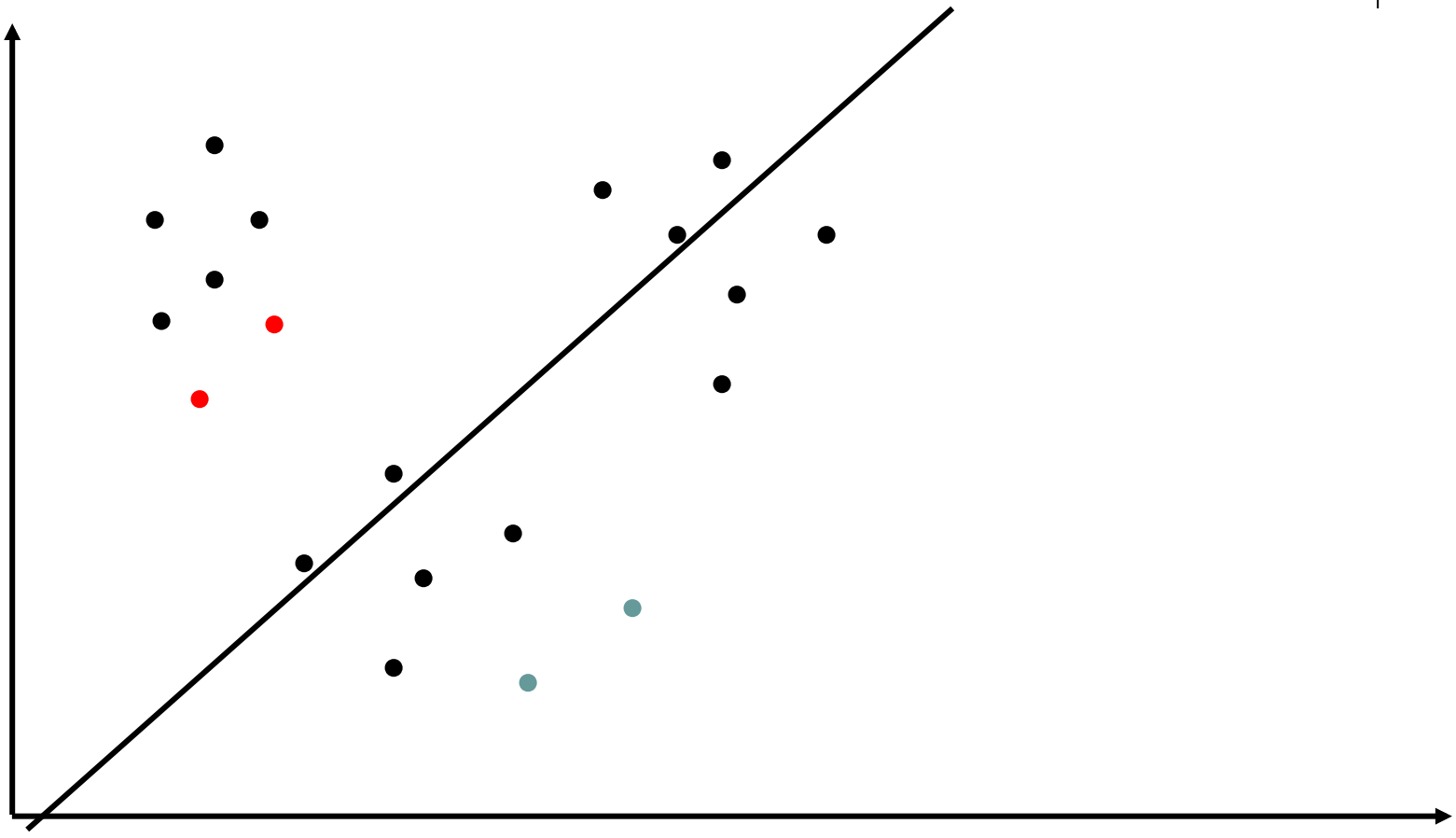
Supervised Classification Example



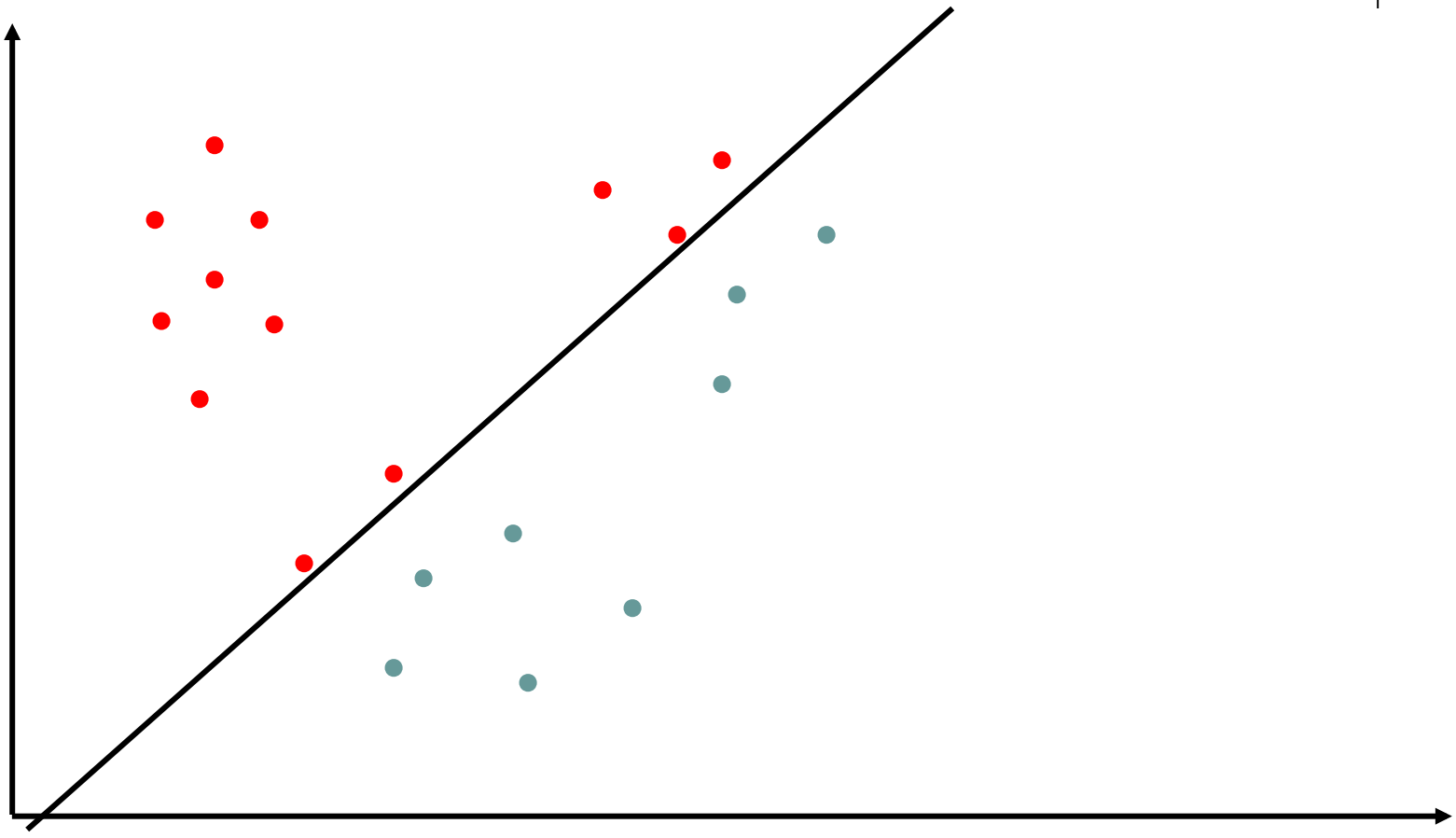
Supervised Classification Example



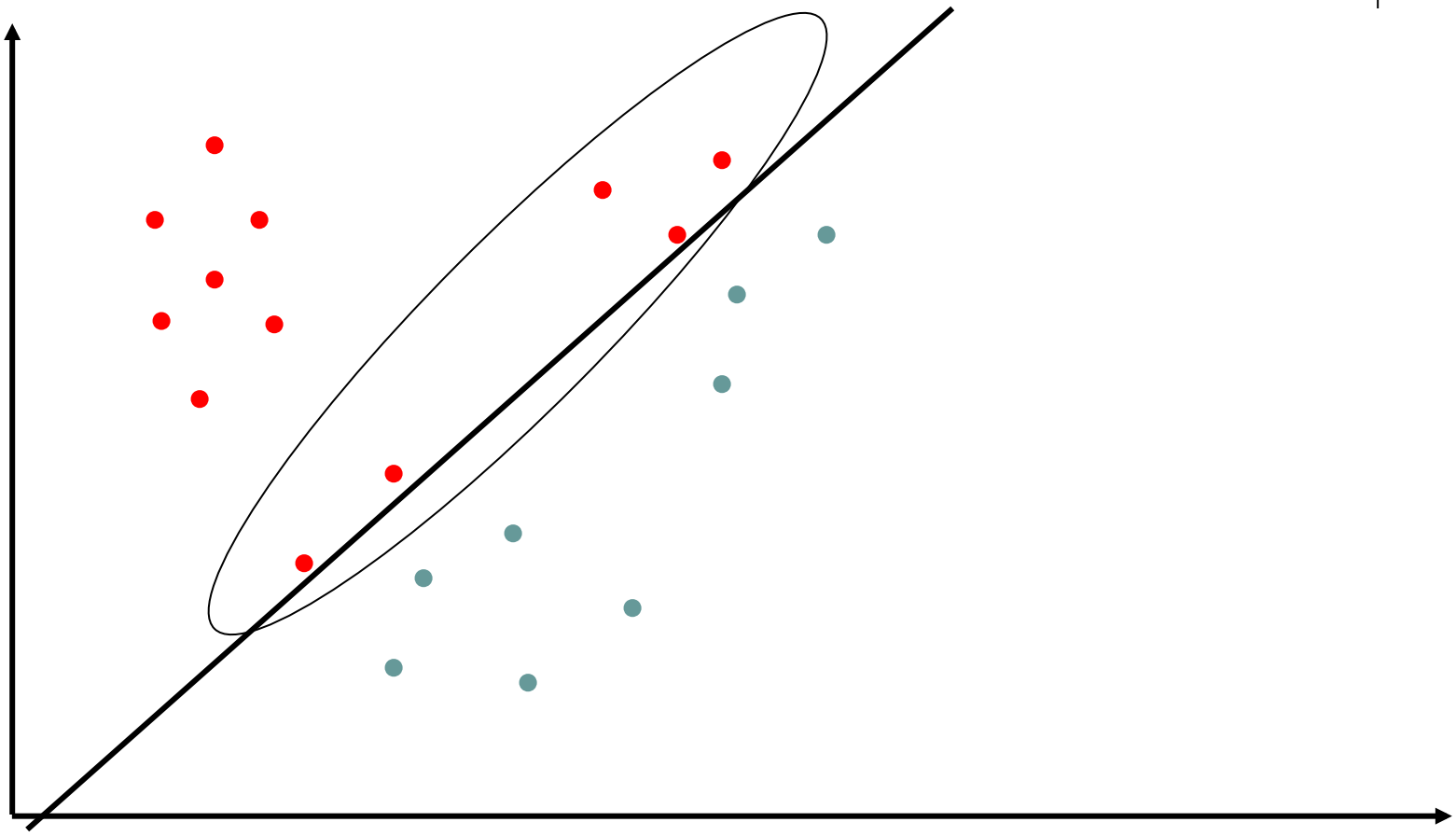
Supervised Classification Example



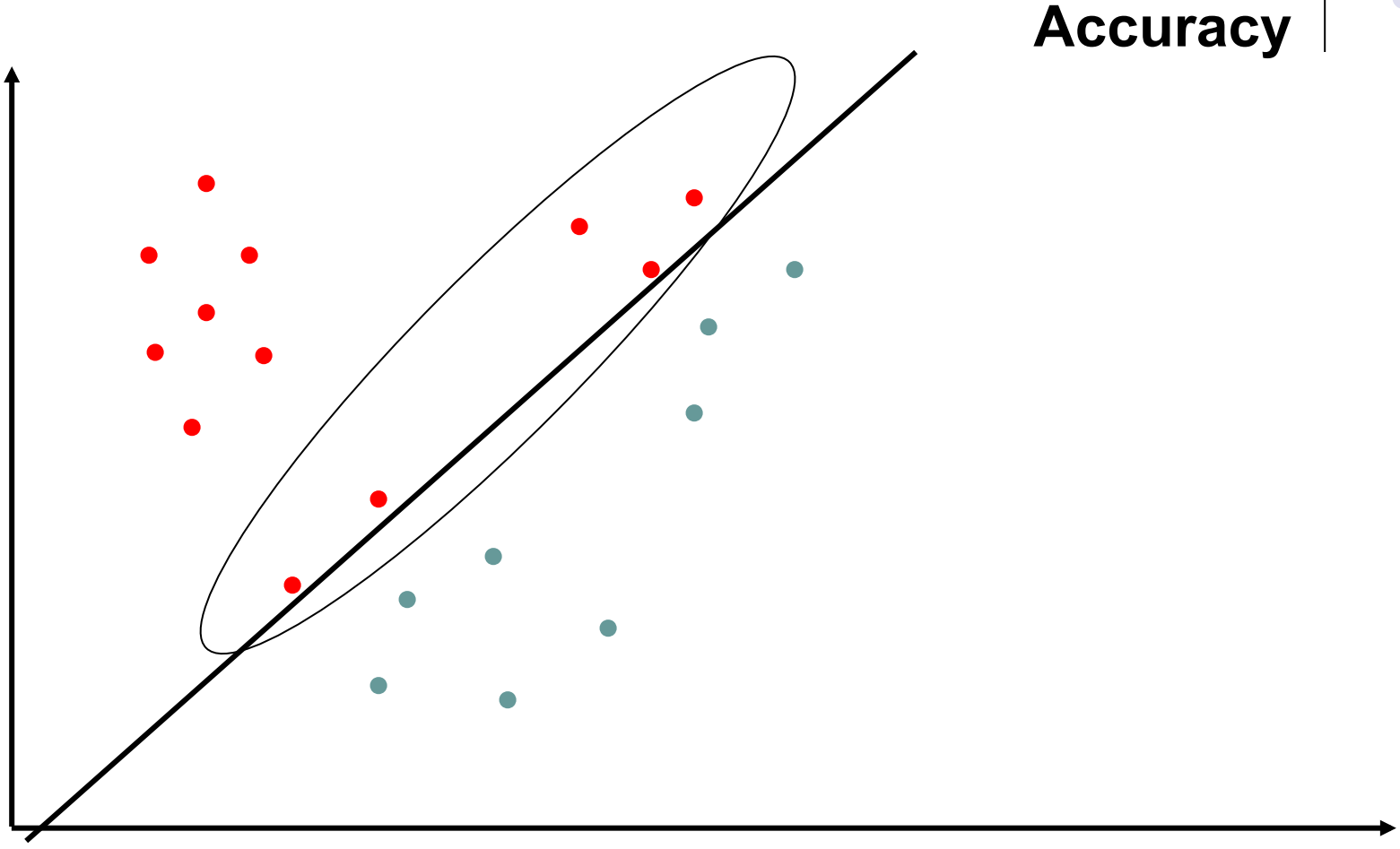
Supervised Classification Example



Supervised Classification Example



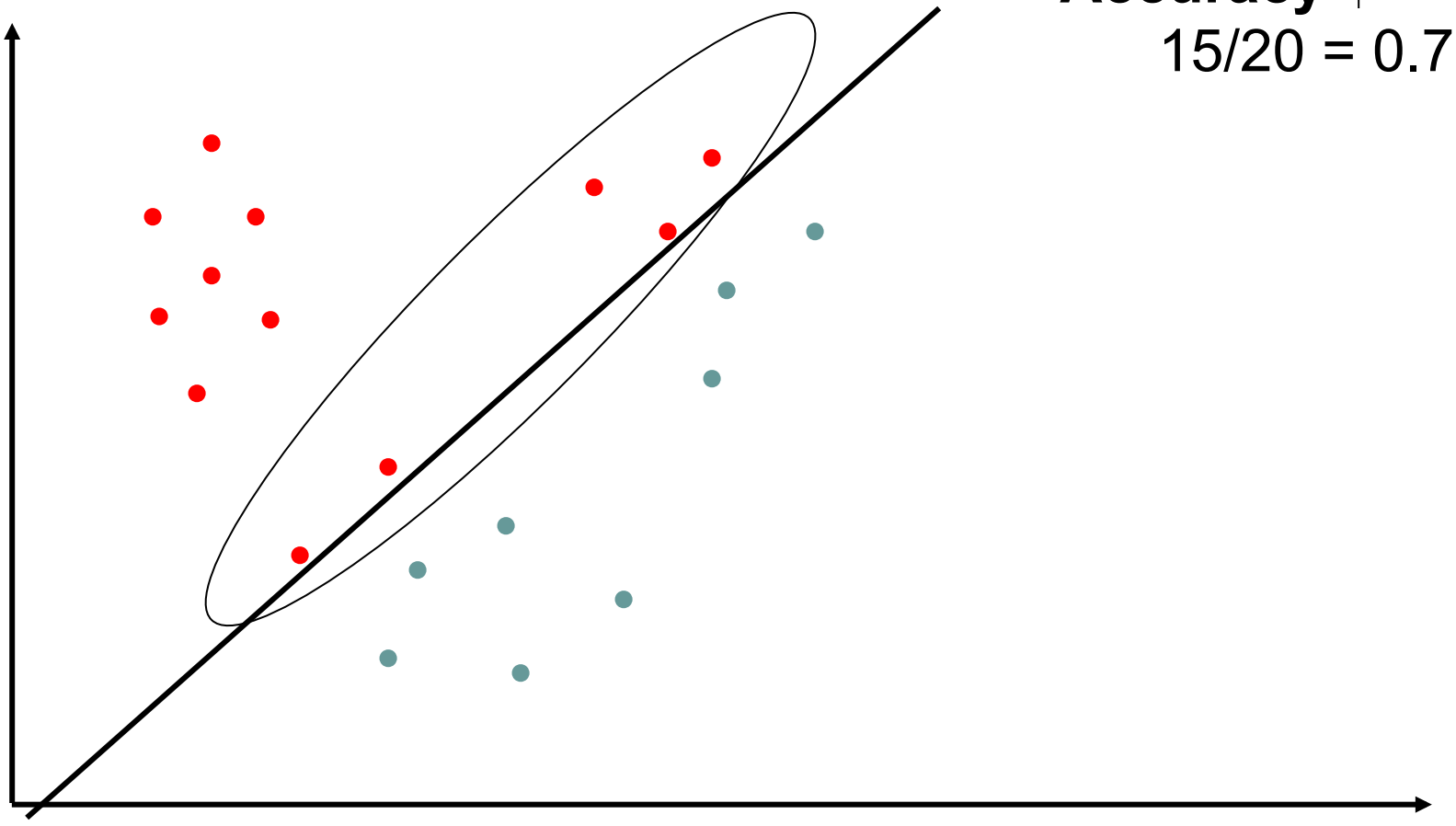
Supervised Classification Example



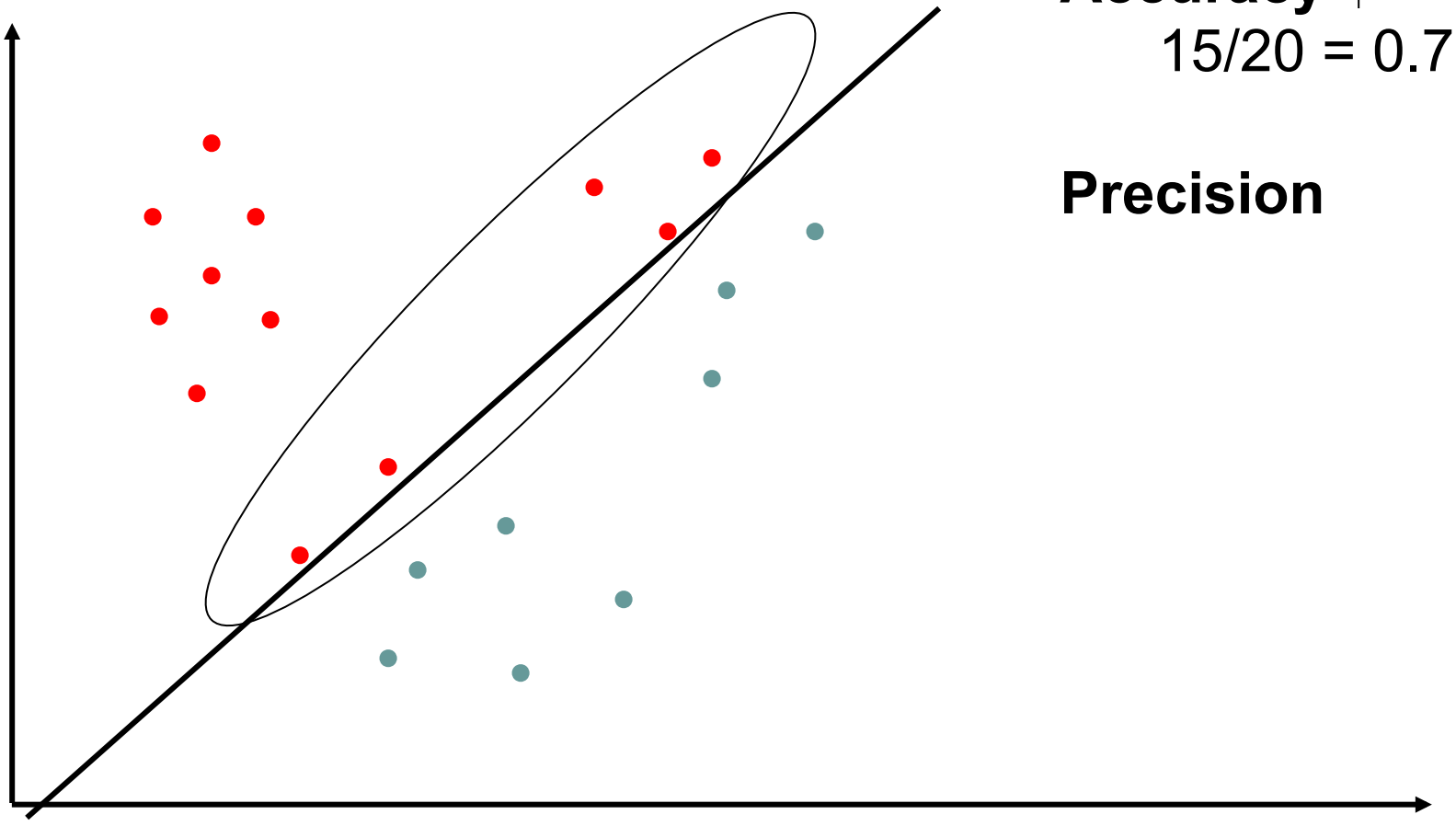
Supervised Classification Example

Accuracy

15/20 = 0.75



Supervised Classification Example

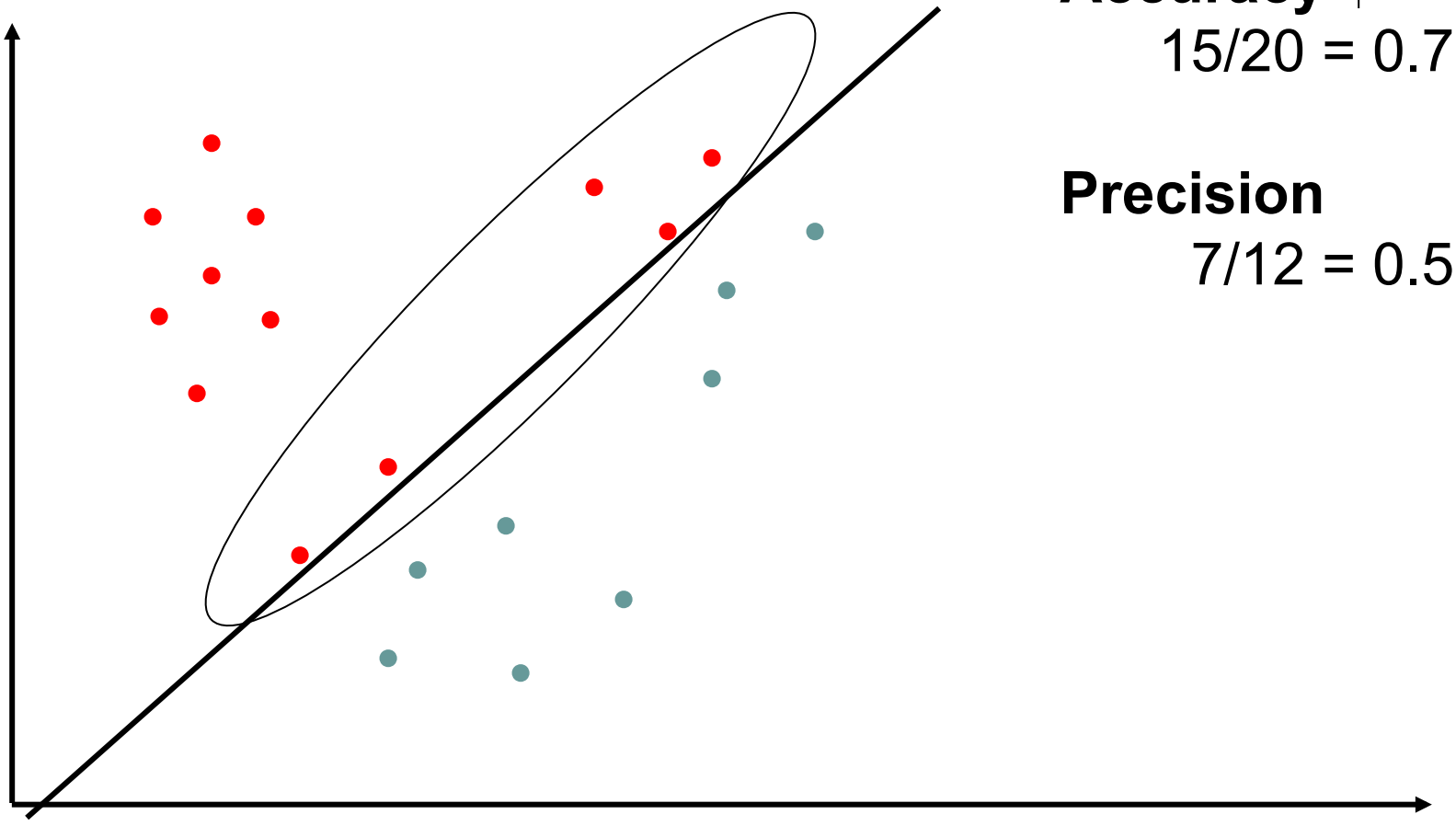


Accuracy

15/20 = 0.75

Precision

Supervised Classification Example



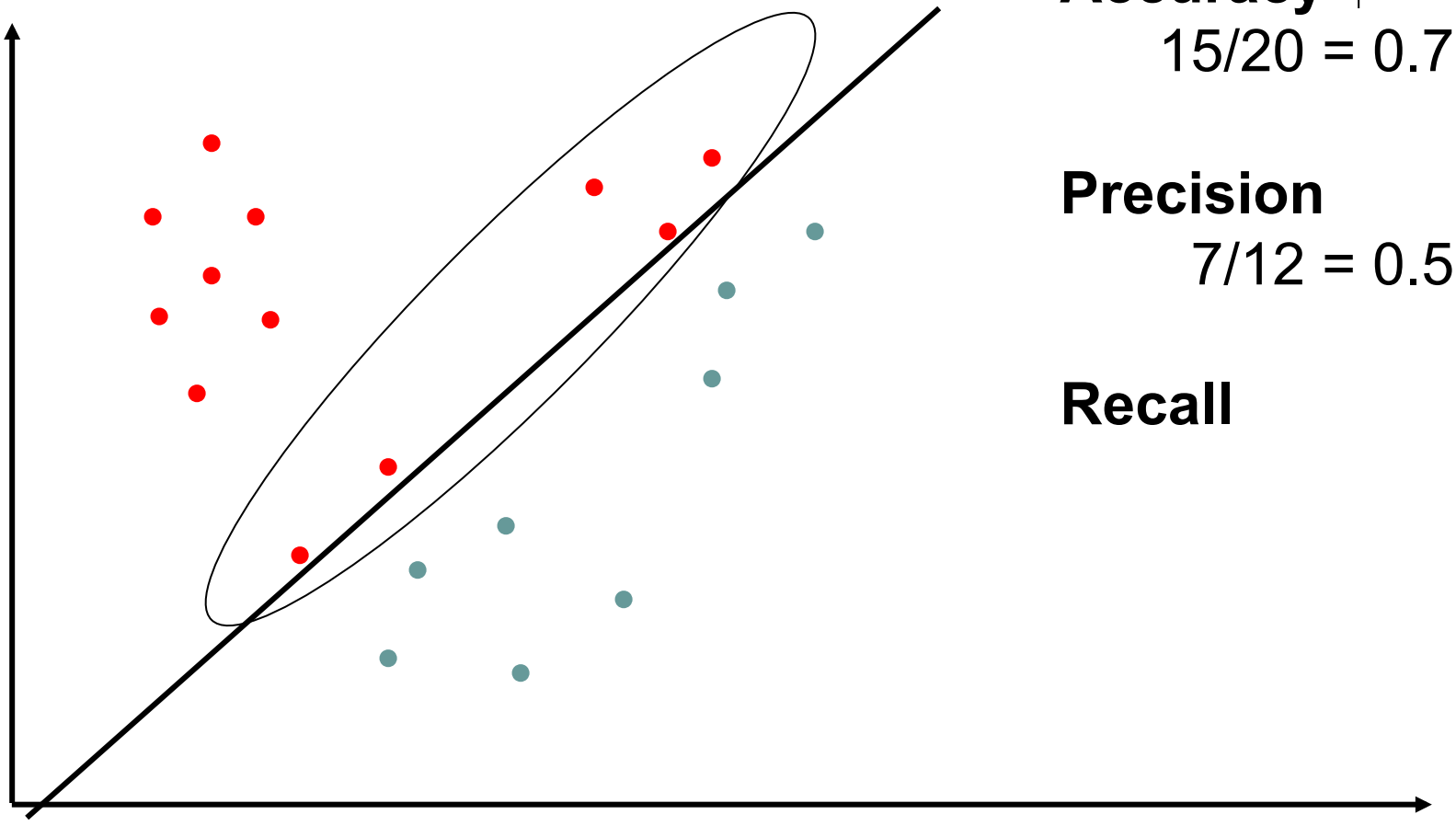
Accuracy

15/20 = 0.75

Precision

7/12 = 0.58

Supervised Classification Example



Accuracy

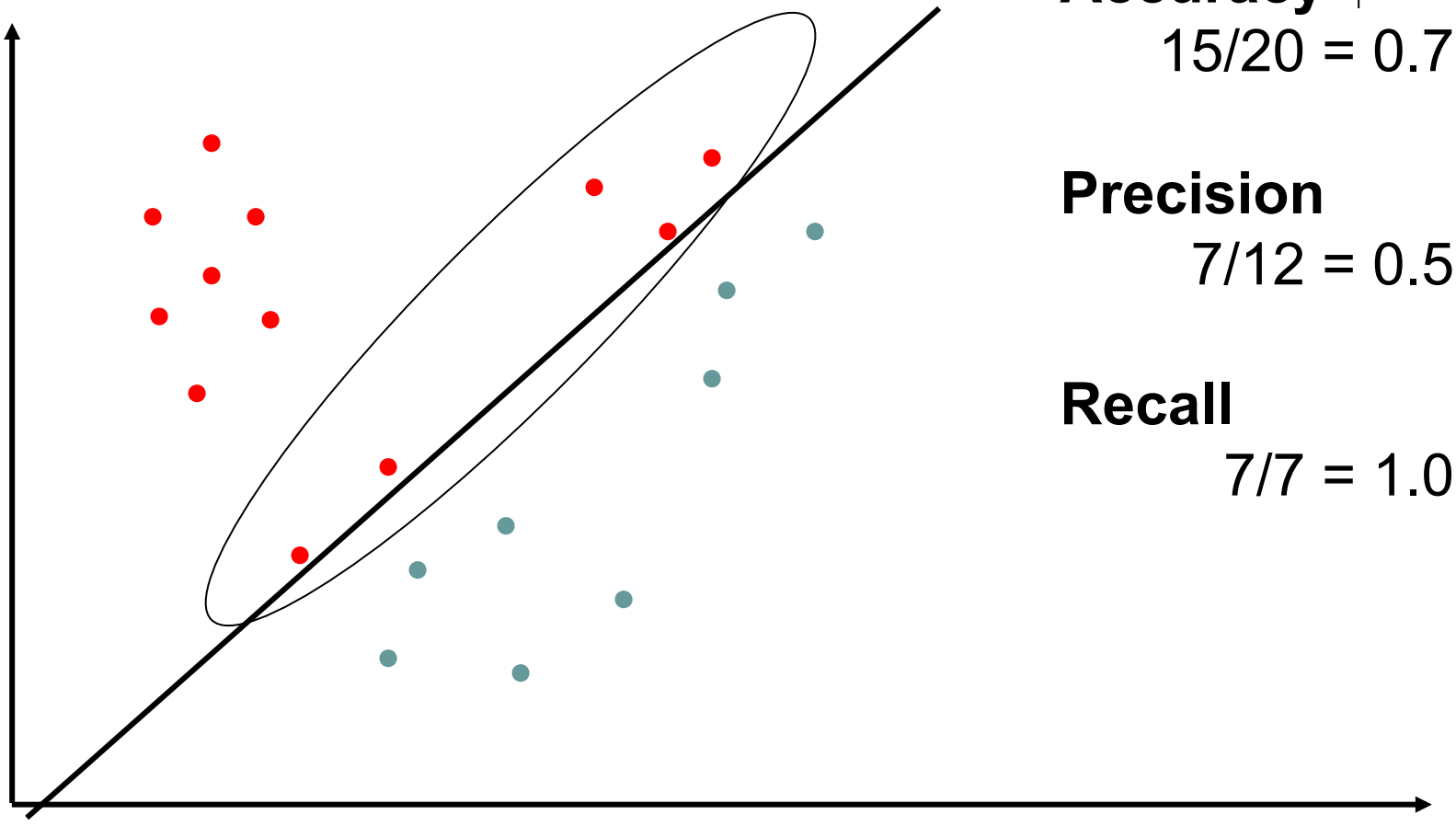
15/20 = 0.75

Precision

7/12 = 0.58

Recall

Supervised Classification Example



Accuracy

15/20 = 0.75

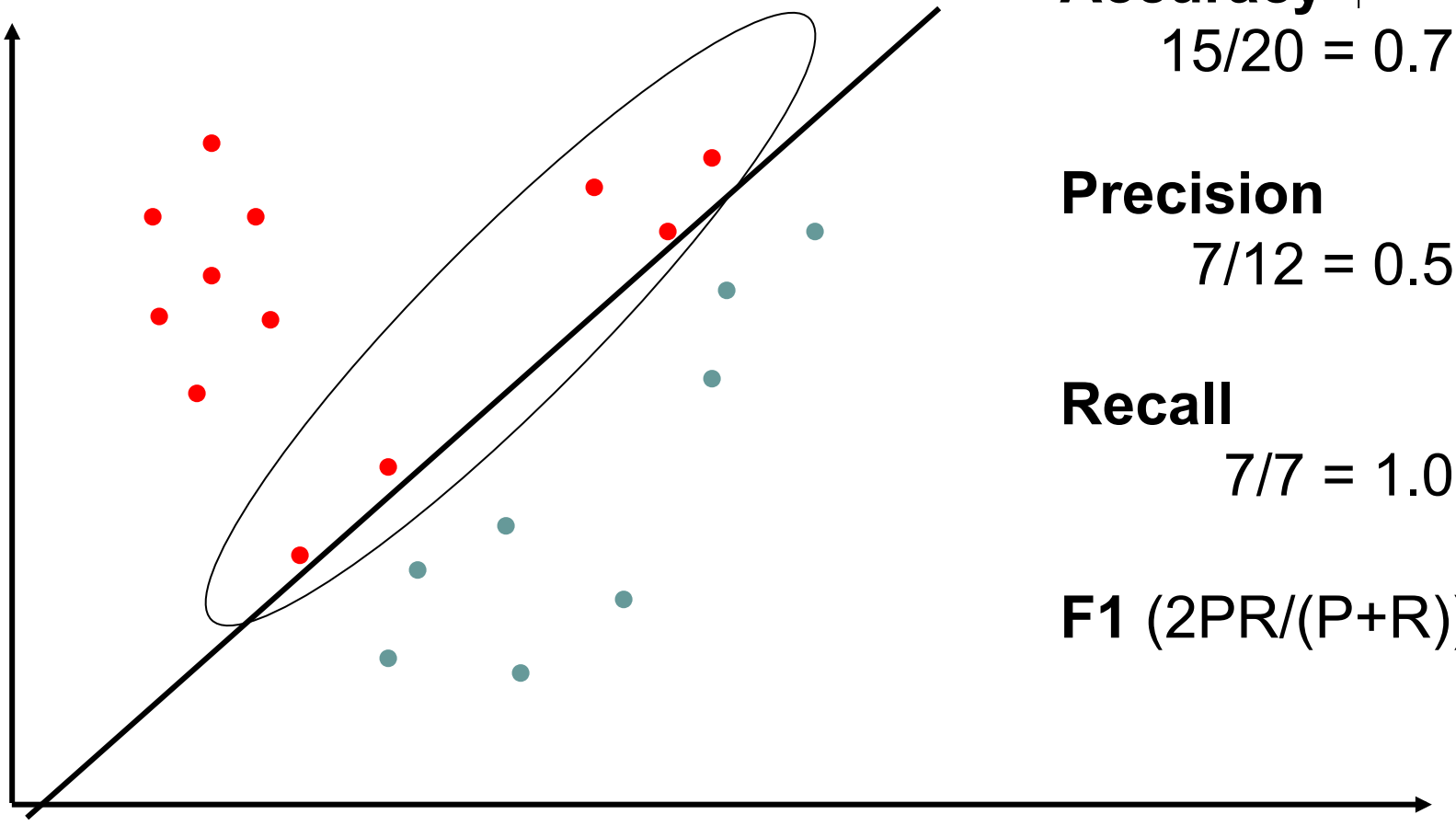
Precision

7/12 = 0.58

Recall

7/7 = 1.0

Supervised Classification Example



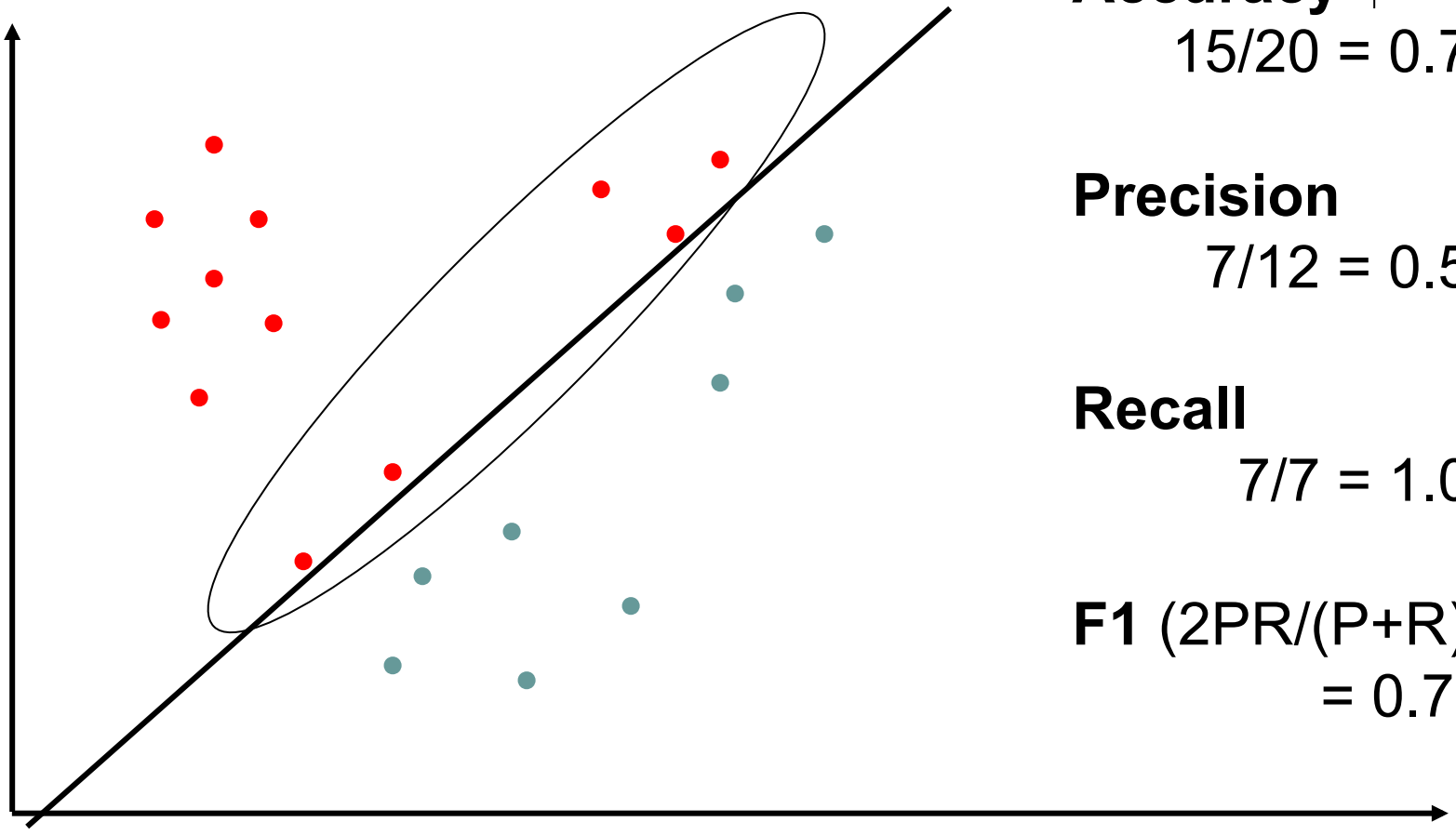
Accuracy
15/20 = 0.75

Precision
7/12 = 0.58

Recall
7/7 = 1.0

F1 (2PR/(P+R))

Supervised Classification Example



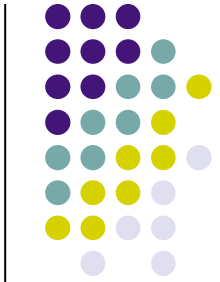
Accuracy
15/20 = 0.75

Precision
7/12 = 0.58

Recall
7/7 = 1.0

F1 (2PR/(P+R))
= 0.73

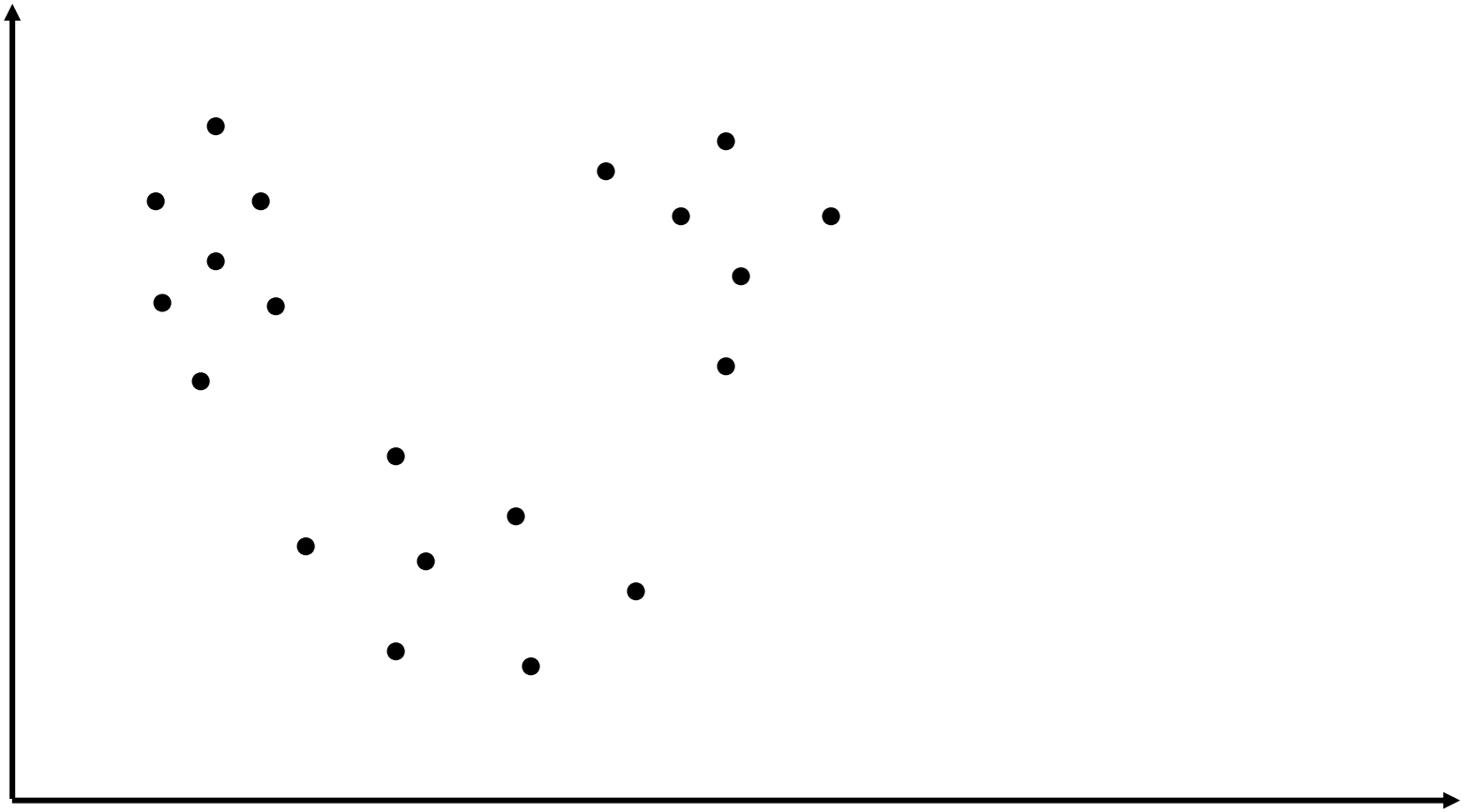
Supervised ML in practice



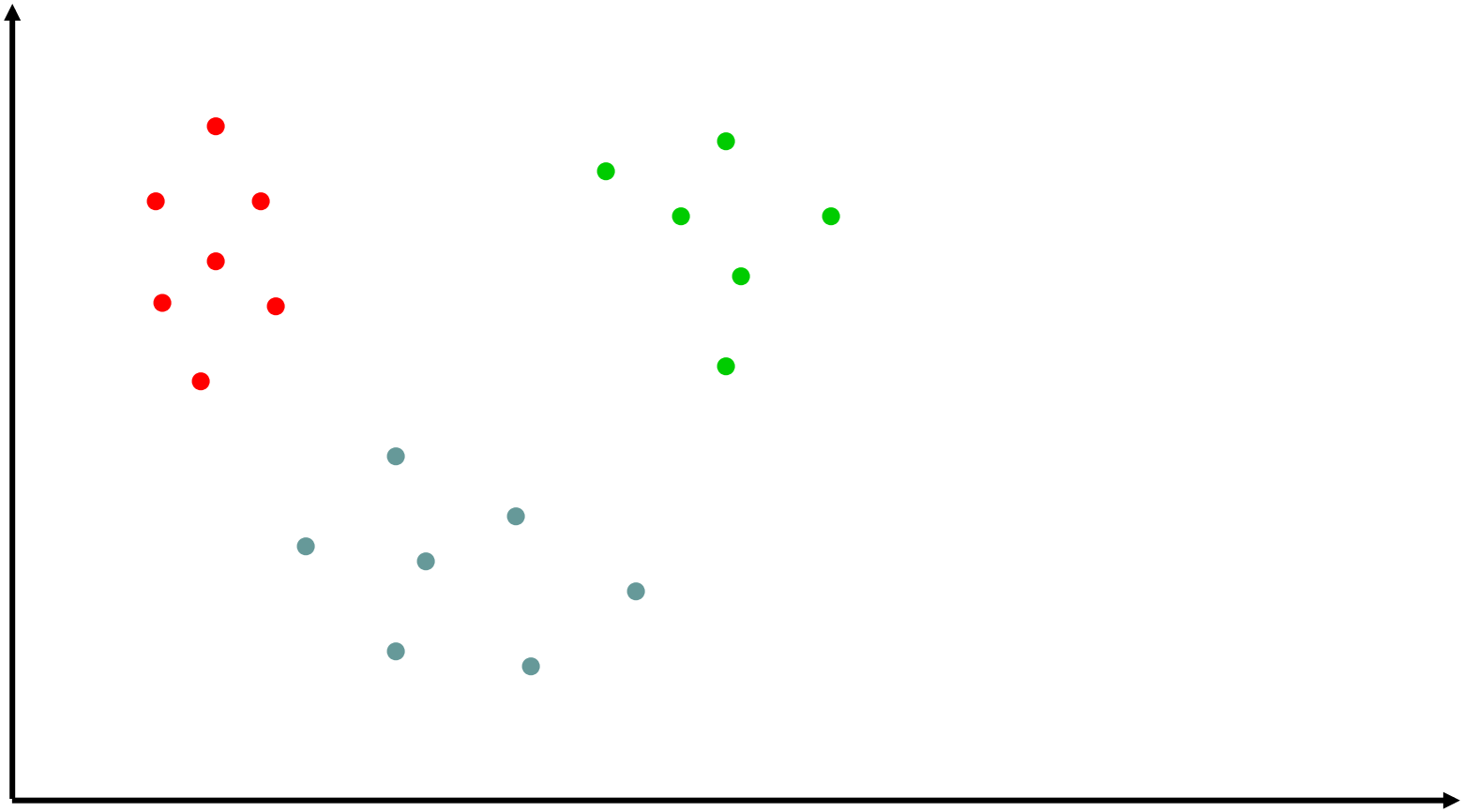
- Supervised learning algorithm choice
  - Support Vector Machines
  - Naïve Bayes
  - Neural Networks
  - Decision Trees
- Configuration
- Feature Selection
- Training Data

=

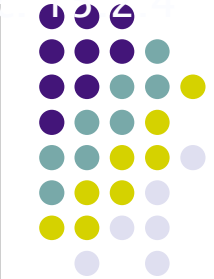
Unsupervised Clustering Example



Unsupervised Clustering Example



Evaluation: Classic Reuters-21578 Data Set

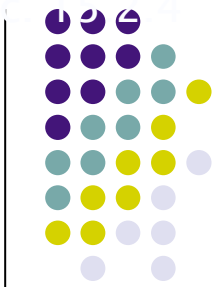


- Most (over)used data set
- 21578 documents
- 9603 training, 3299 test articles (ModApte split)
- 118 categories
  - An article can be in more than one category
  - Learn 118 binary category distinctions (118 2-class classifiers)
- Average number of classes assigned
  - 1.24 for docs with at least one category
- Only about 10 out of 118 categories are large

Common categories (#train, #test)

- Earn (2877, 1087)
- Acquisitions (1650, 179)
- Money-fx (538, 179)
- Grain (433, 149)
- Crude (389, 189)
- Trade (369,119)
- Interest (347, 131)
- Ship (197, 89)
- Wheat (212, 71)
- Corn (182, 56)

Reuters Text Categorization data set (Reuters-21578) document



<REUTERS TOPICS="YES" LEWISSPLIT="TRAIN" CGISPLIT="TRAINING-SET" OLDID="12981" NEWID="798">

<DATE> 2-MAR-1987 16:51:43.42</DATE>

<TOPICS><D>livestock</D><D>hog</D></TOPICS>

<TITLE>AMERICAN PORK CONGRESS KICKS OFF TOMORROW</TITLE>

<DATELINE> CHICAGO, March 2 - </DATELINE><BODY>The American Pork Congress kicks off tomorrow, March 3, in Indianapolis with 160 of the nations pork producers from 44 member states determining industry positions on a number of issues, according to the National Pork Producers Council, NPPC.

Delegates to the three day Congress will be considering 26 resolutions concerning various issues, including the future direction of farm policy and the tax law as it applies to the agriculture sector. The delegates will also debate whether to endorse concepts of a national PRV (pseudorabies virus) control and eradication program, the NPPC said.

A large trade show, in conjunction with the congress, will feature the latest in technology in all areas of the industry, the NPPC added. Reuter

&#3;</BODY></TEXT></REUTERS>

Per class evaluation measures

- Recall: Fraction of docs in class *i* classified correctly:

$$\frac{c_{ii}}{\sum_j c_{ij}}$$

- Precision: Fraction of docs assigned class *i* that are actually about class *i*:

$$\frac{c_{ii}}{\sum_j c_{ji}}$$

- F Measure (F1) = 2PR/(P + R)

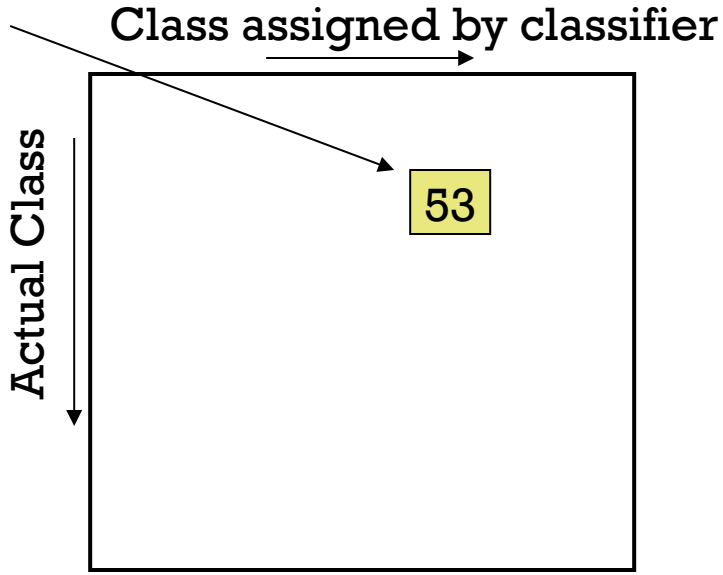
$$F_\beta = (1+\beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}$$

- Accuracy: Fraction of docs classified correctly:

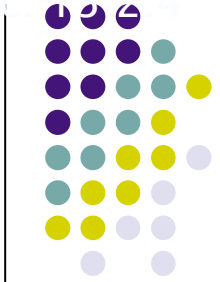
$$\frac{\sum_i c_{ii}}{\sum_j \sum_i c_{ij}}$$

Confusion Matrix

This (*i*, *j*) entry means 53 of the docs actually in class *i* were put in class *j* by the classifier.

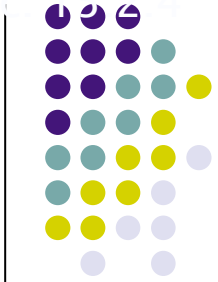


Micro- vs. Macro-Averaging



- If we have more than one class, how do we combine multiple performance measures into one quantity?
- Macroaveraging: Compute performance for each class, then average.
- Microaveraging: Collect decisions for all classes, compute contingency table, evaluate.

Micro- vs. Macro-Averaging: Example



Class 1

	Truth: yes	Truth: no
Classifier: yes	10	10
Classifier: no	10	970

Class 2

	Truth: yes	Truth: no
Classifier: yes	90	10
Classifier: no	10	890

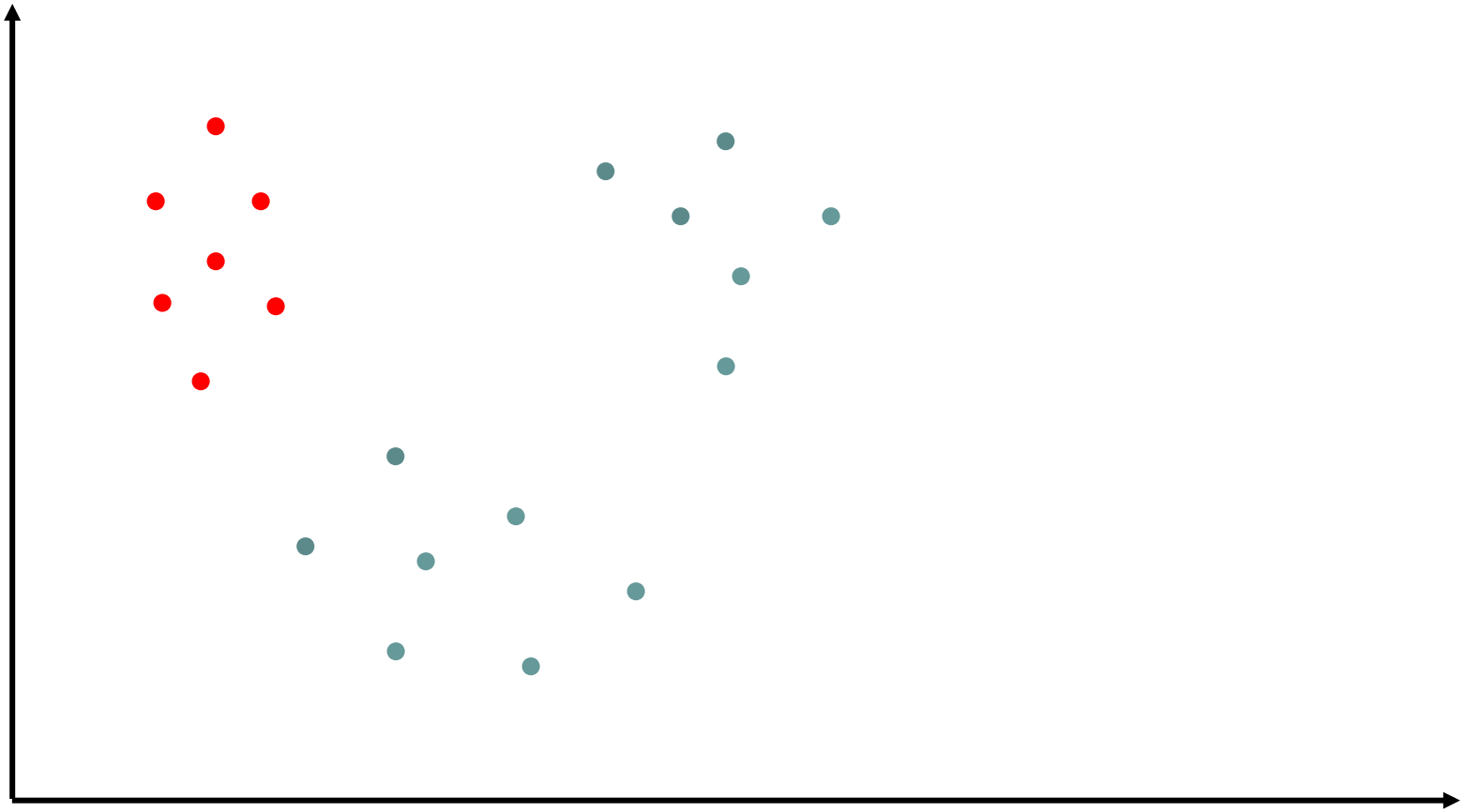
Micro Ave. Table

	Truth: yes	Truth: no
Classifier: yes	100	20
Classifier: no	20	1860

- Macroaveraged precision: (0.5 + 0.9)/2 = 0.7
- Microaveraged precision: 100/120 = .83

- Microaveraged score is dominated by score on common classes

kNN – k nearest neighbors



(a)

	NB	Rocchio	kNN	SVM
micro-avg-L (90 classes)	80	85	86	89
macro-avg (90 classes)	47	59	60	60

(b)

	NB	Rocchio	kNN	trees	SVM
earn	96	93	97	98	98
acq	88	65	92	90	94
money-fx	57	47	78	66	75
grain	79	68	82	85	95
crude	80	70	86	85	89
trade	64	65	77	73	76
interest	65	63	74	67	78
ship	85	49	79	74	86
wheat	70	69	77	93	92
corn	65	48	78	92	90
micro-avg (top 10)	82	65	82	88	92
micro-avg-D (118 classes)	75	62	n/a	n/a	87

Evaluation measure: F_1

Yang&Liu: SVM vs. Other Methods

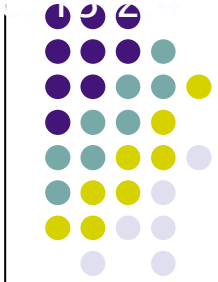


Table 1: Performance summary of classifiers

method	miR	miP	miF1	maF1	error
SVM	.8120	.9137	.8599	.5251	.00365
KNN	.8339	.8807	.8567	.5242	.00385
LSF	.8507	.8489	.8498	.5008	.00414
NNet	.7842	.8785	.8287	.3765	.00447
NB	.7688	.8245	.7956	.3886	.00544

miR = micro-avg recall; miP = micro-avg prec.;
miF1 = micro-avg F1; maF1 = macro-avg F1.