Bag Dissimilarities for Multiple Instance Learning

David M.J. Tax, Marco Loog, Wan-Jui Lee,
Veronika Cheplygina, Robert P.W. Duin

SIMBAD, 2011

TUDelft

How to label this image?

... "Red chilli"?

四川
青麻椒
白胡椒
黑胡椒
BLACK PEPPER
紅辣椒乾
RED CHILLI
草果
KEMBANG RUMBUT
桂皮
CINNAMON
甘草
LIQUORICE
FENNEL
角
SEED
什果皮
丁香
CLOVE
野山椒干
PERENNIAL CHILLI

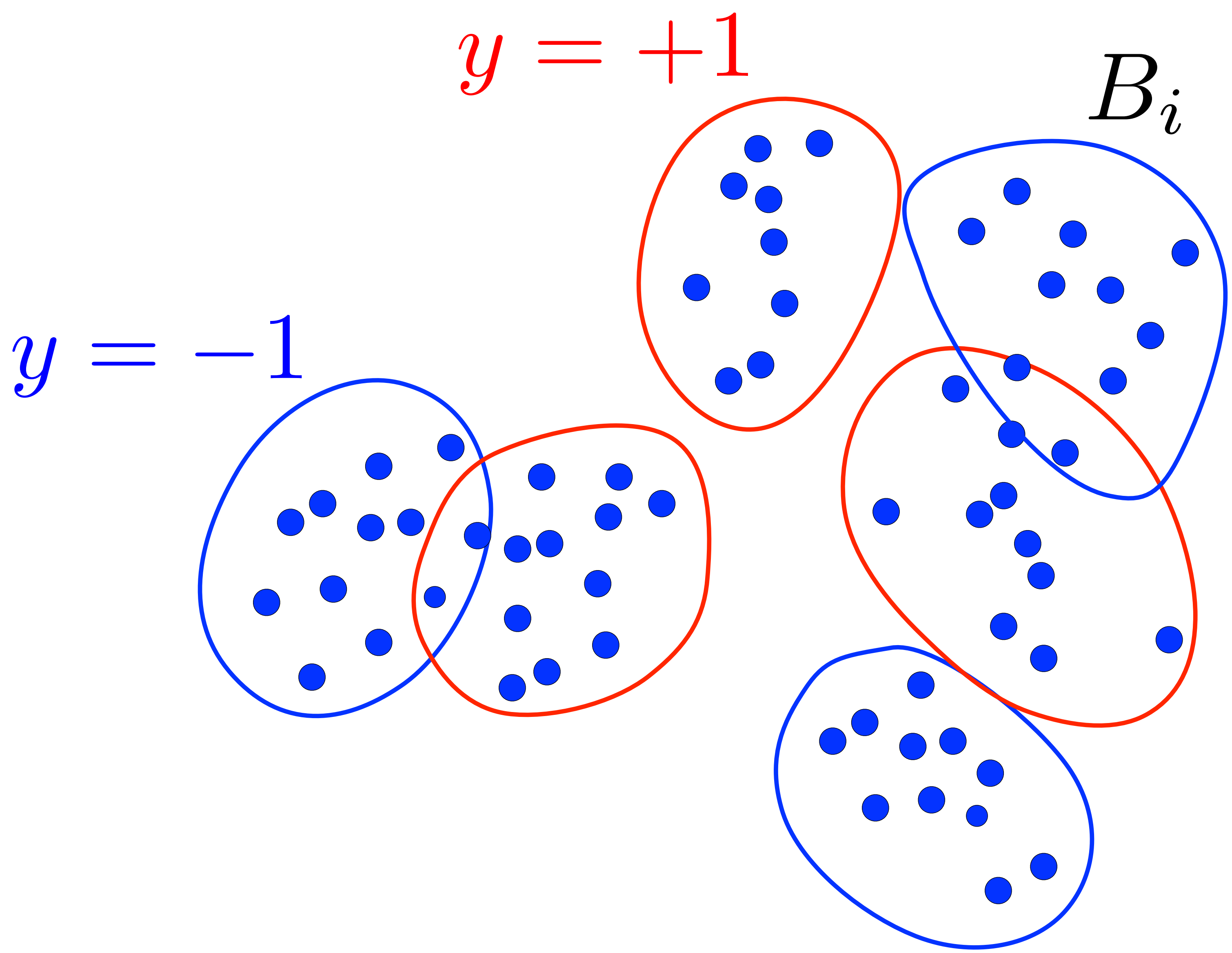
… "Red chilli"?

TUDelft

How to REPRESENT this image?

Multiple Instance Learning (MIL)

- Represent an object (a **bag**) by a collection of feature vectors (or **instances**)
- Each bag is labeled



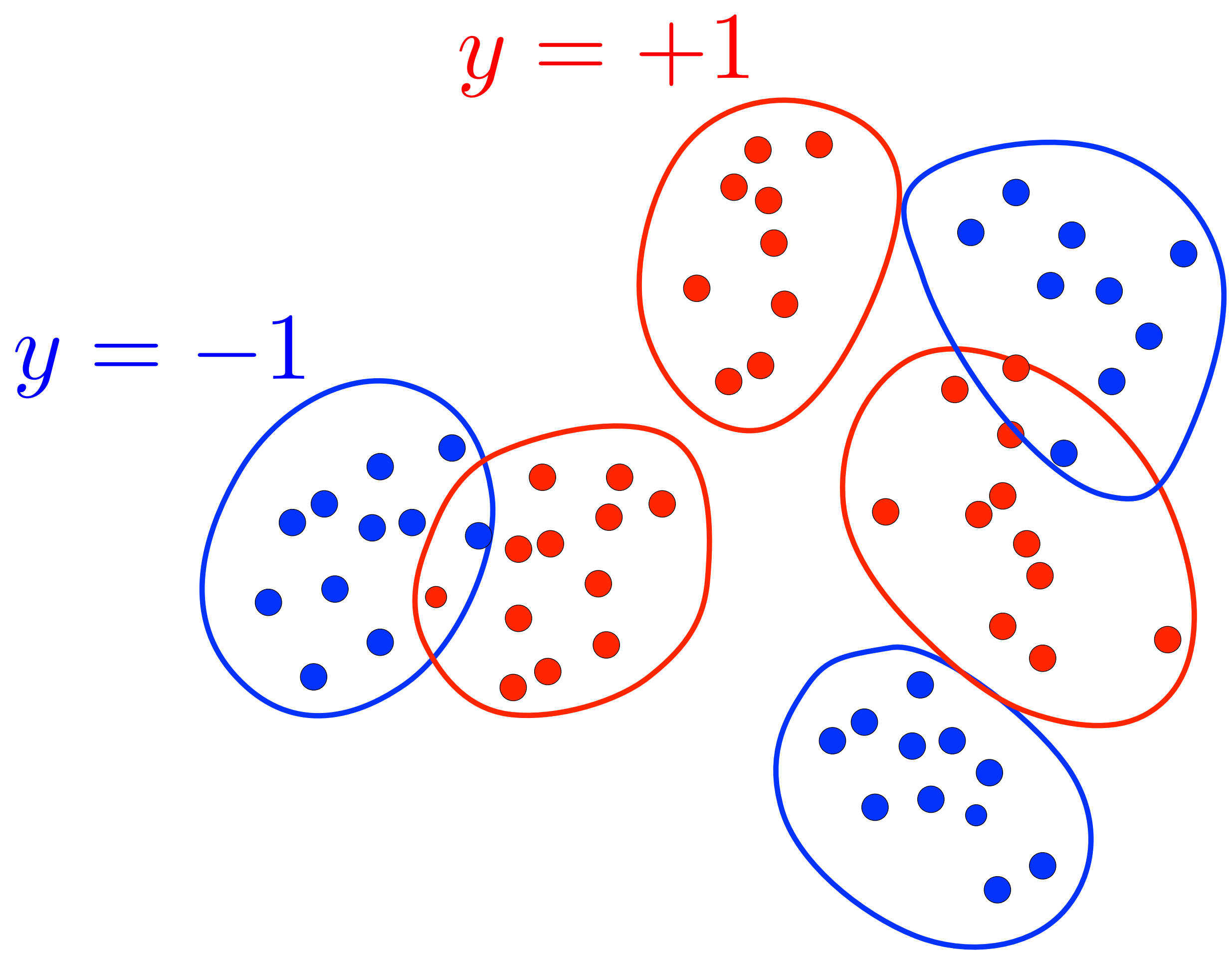
Contents

- Multiple-Instance Learning (MIL)
- Classical MIL
- Bag dissimilarities
  - based on pairwise comparisons
  - based on distribution differences
- Experiments
- Observations, open questions/challenge to you
- Conclusions

Multiple Instance Learning (MIL)

- Represent an object (a **bag**) by a collection of feature vectors (or **instances**)
- Each bag is labeled

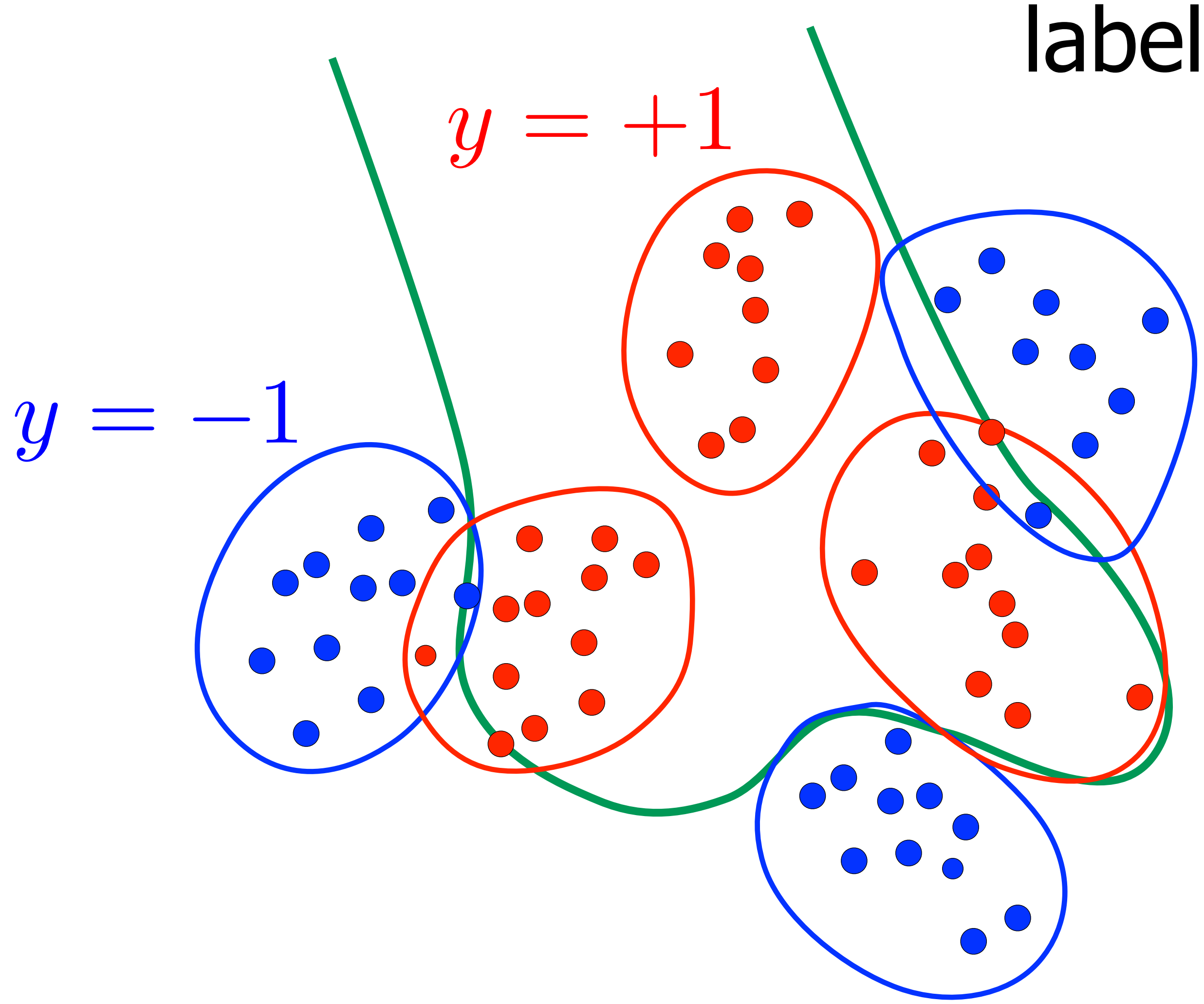
naive approach:
label according to bag label



Multiple Instance Learning (MIL)

- Represent an object (a **bag**) by a collection of feature vectors (or **instances**)
- Each bag is labeled

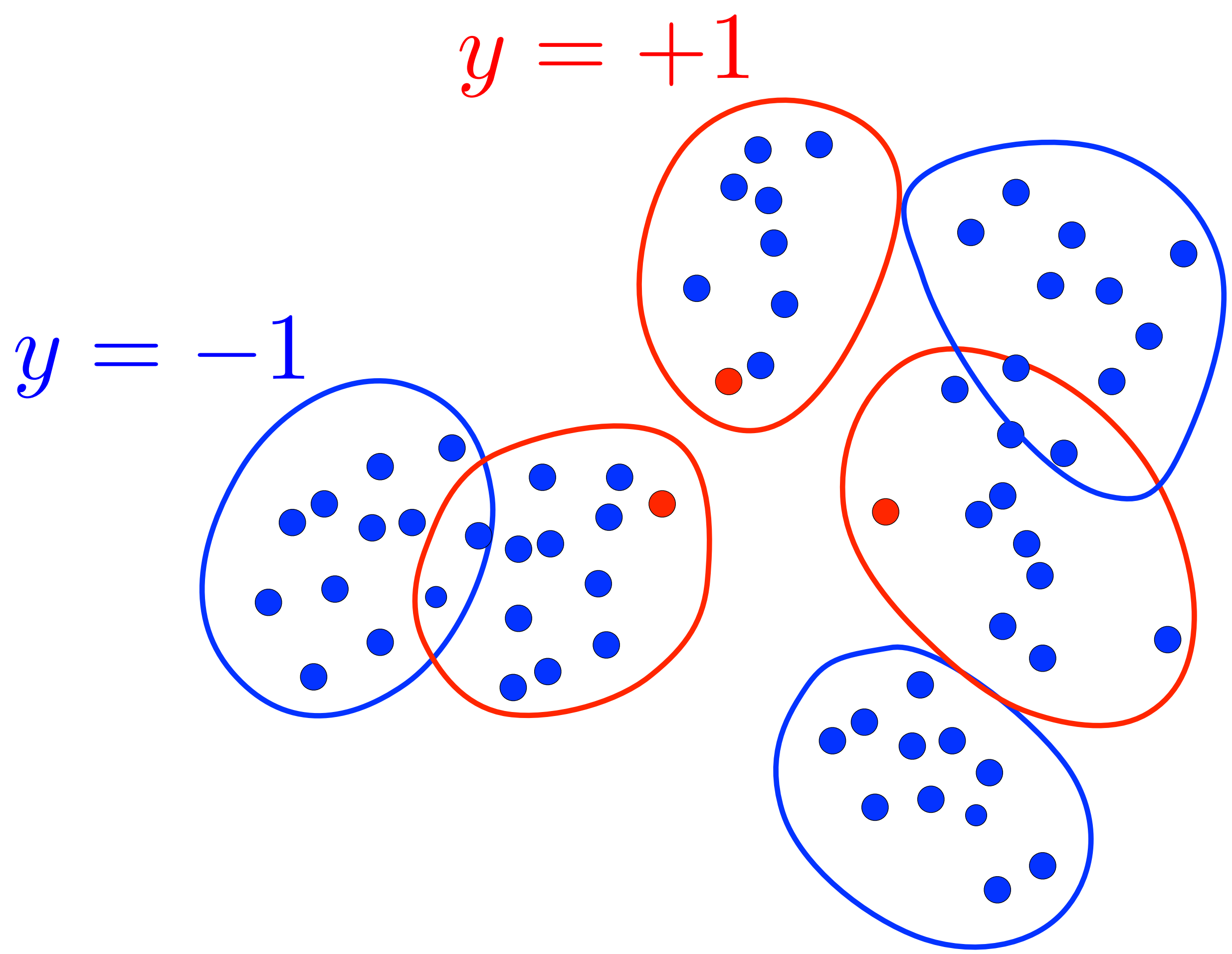
naive approach:
label according to bag
label



Multiple Instance Learning (MIL)

- Represent an object (a **bag**) by a collection of feature vectors (or **instances**)
- Each bag is labeled

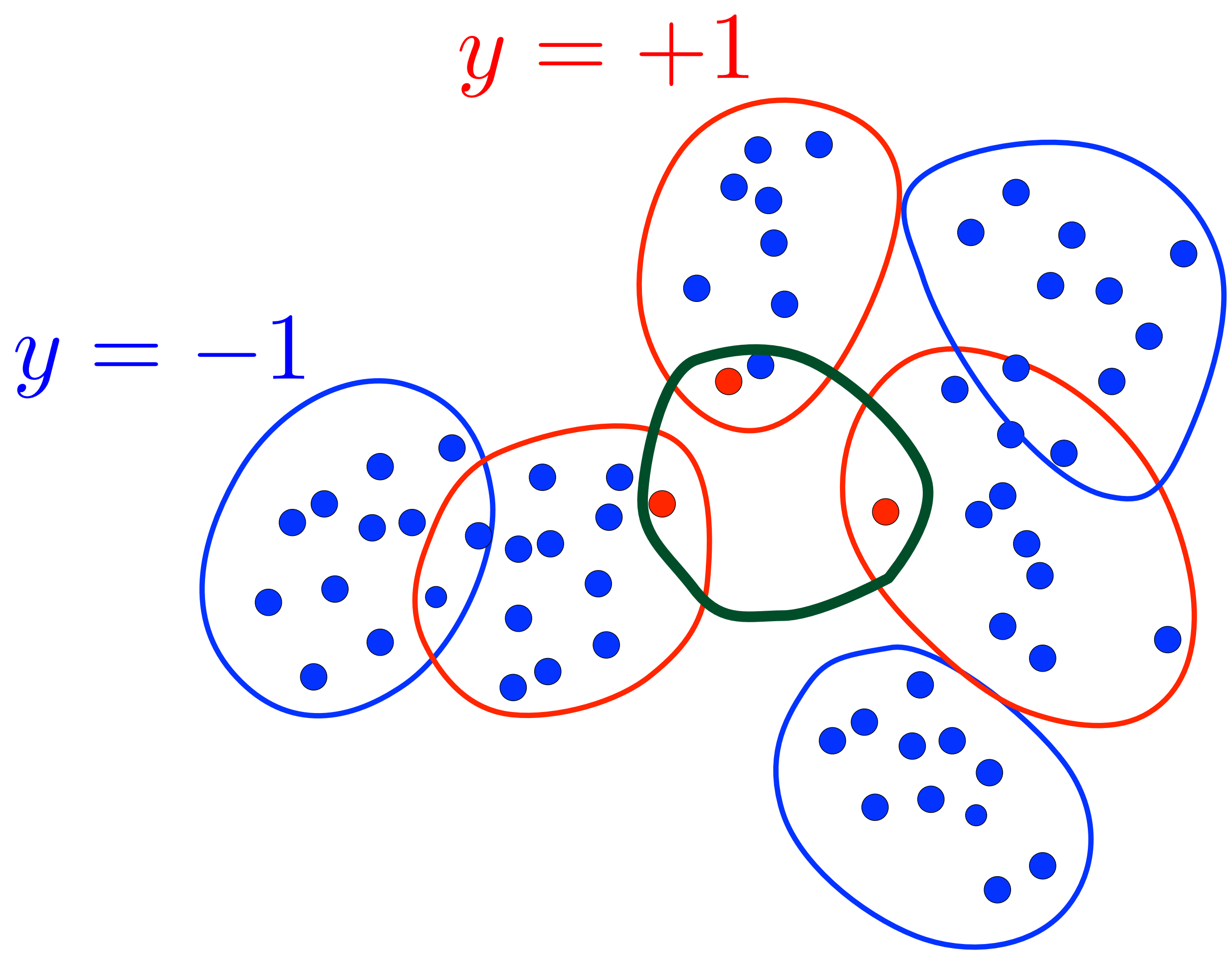
MIL approach:
model a concept



Multiple Instance Learning (MIL)

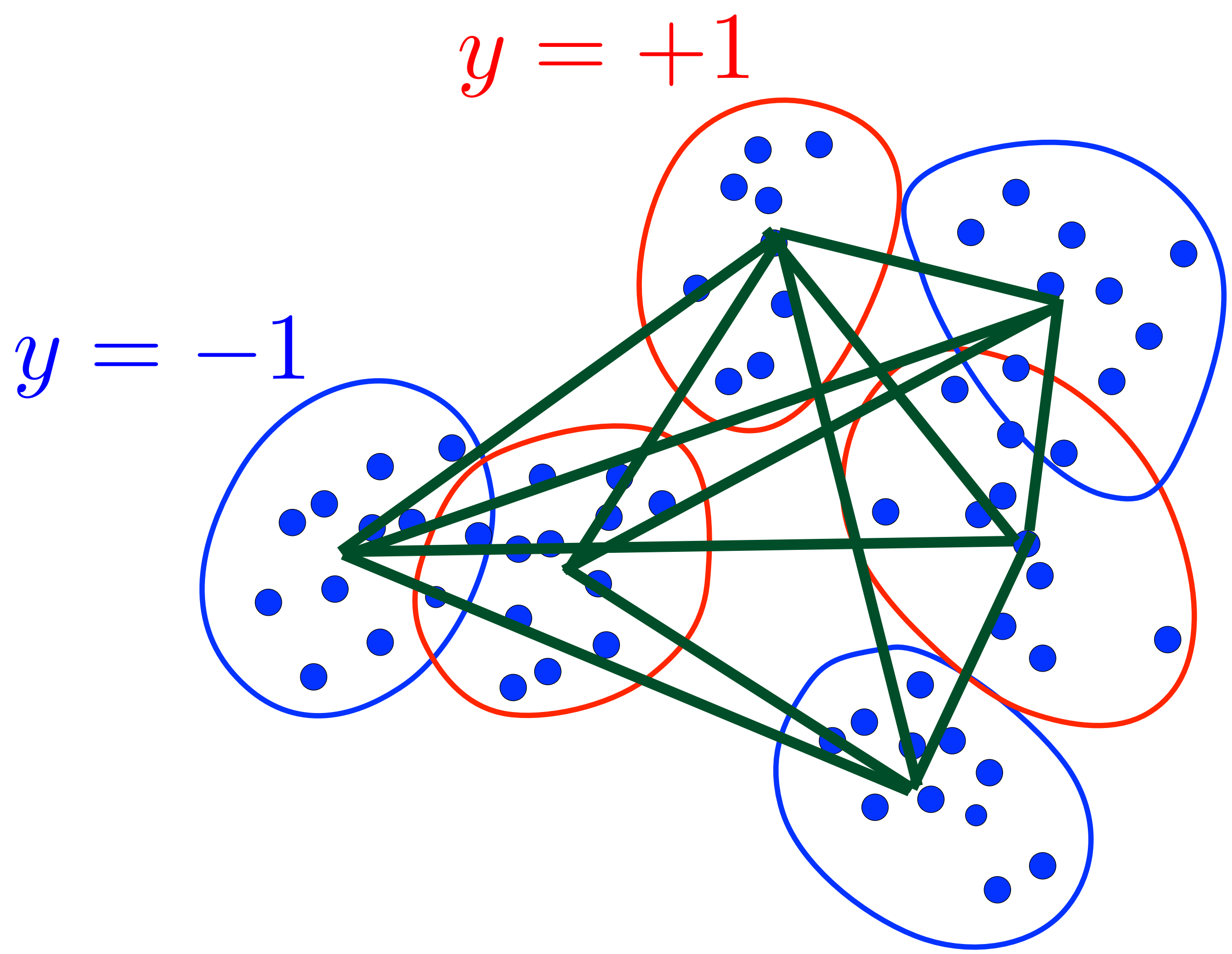
- Represent an object (a **bag**) by a collection of feature vectors (or **instances**)
- Each bag is labeled

MIL approach:
model a concept



Multiple Instance Learning (MIL)

- Dissimilarity approach: define a distance between bags
$$\tilde{\mathbf{x}}(B_j) = [d(B_j, B_1), \ldots, d(B_j, B_N)]$$
- Train (and eval.) a traditional classifier on these features



Notation

- Assume we have N bags of instances
- Each bag B_i has n_i instances

$$B_i = \{\mathbf{x}_{i1}, ..., \mathbf{x}_{ij}, ..., \mathbf{x}_{in_i}\}$$

- In training, each bag is labeled

$$\{(B_i, y_i), \quad i = 1, .., N\}$$

where

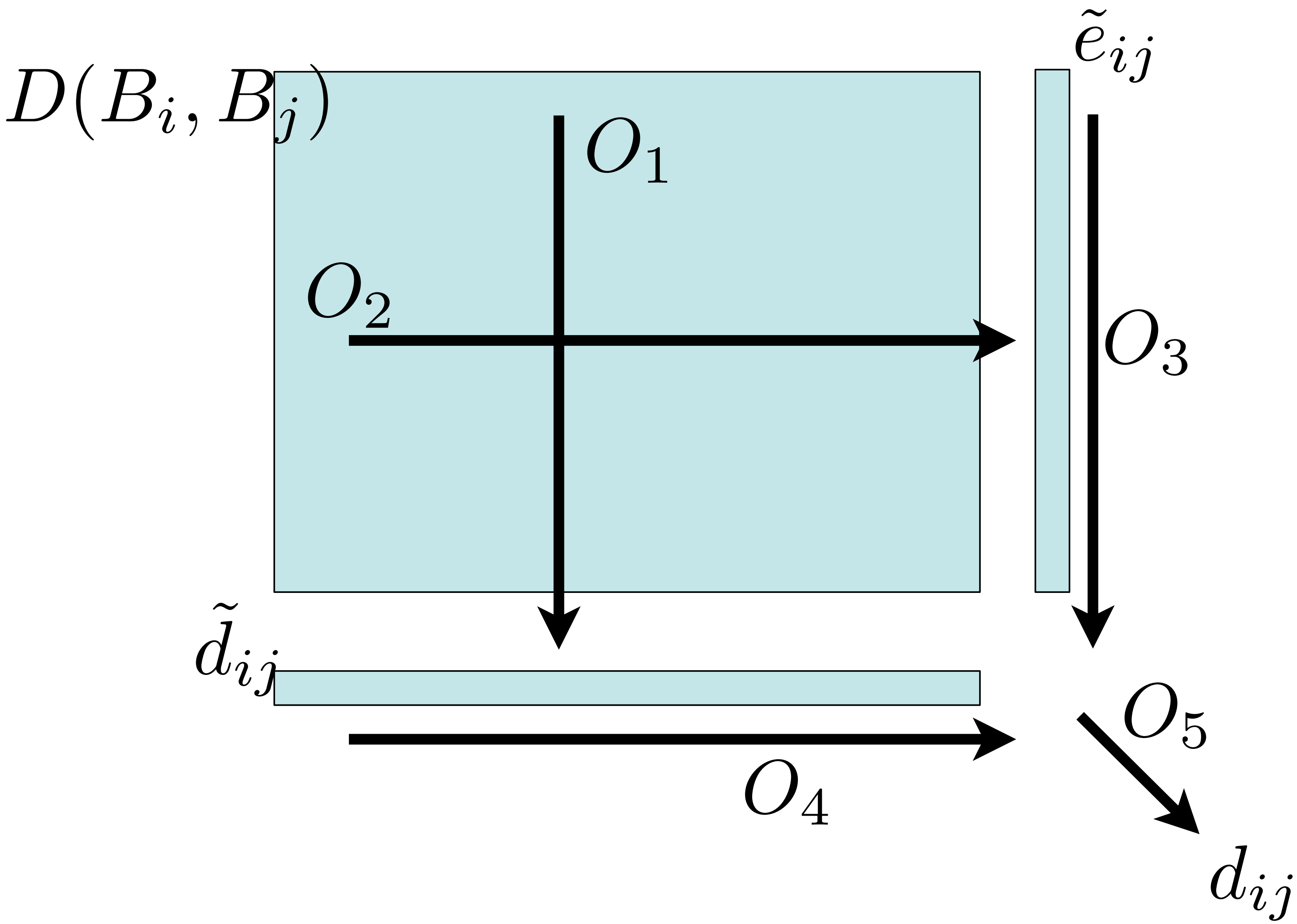
$$y \in \{-1, +1\}$$

- Define:

$$d_{ij} = D(B_i, B_j) = \begin{pmatrix} D(\mathbf{x}_{i1}, \mathbf{x}_{j1}) & \dots & D(\mathbf{x}_{i1}, \mathbf{x}_{jn_j}) \\ D(\mathbf{x}_{i2}, \mathbf{x}_{j1}) & \dots & D(\mathbf{x}_{i2}, \mathbf{x}_{jn_j}) \\ \vdots & & \vdots \\ D(\mathbf{x}_{in_i}, \mathbf{x}_{j1}) & \dots & D(\mathbf{x}_{in_i}, \mathbf{x}_{jn_j}) \end{pmatrix}$$

Bag dissimilarities

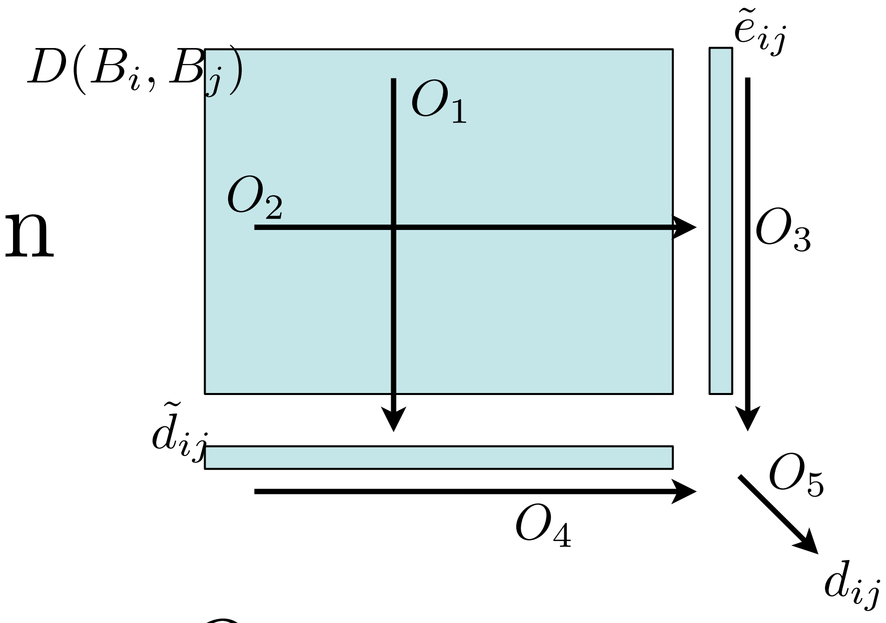
- How to define a bag similarity? Use pairwise distances...



Bag dissim. using pairwise dist.

- Overall minimum distance:

$$O_1 = O_2 = O_3 = O_4 = O_5 = \min$$



- Mean minimum distance:

$$O_1 = O_2 = \min, \quad O_3 = O_4 = \text{mean}, \quad O_5 = \text{mean}$$

- Standard Hausdorff distance:

$$O_1 = O_2 = \min, \quad O_3 = O_4 = \max, \quad O_5 = \max$$

- ... and many more.

Bag distribution dissimilarities

- Bags of instances are samples of a distribution
- Mahalanobis

$$d_{ij} = (\mu_i - \mu_j)^T \left(\frac{1}{2}\Sigma_i + \frac{1}{2}\Sigma_j \right)^{-1} (\mu_i - \mu_j)$$

- Earth Mover's Distance (EMD)
  (minimize the flow f_{kl} to transform one uniform PDF over instances to another)

$$d_{ij} = \min_{f_{kl}} \sum_{k,l} f_{kl} D_{ij}(k,l)$$

Does it make sense?

- Strict MIL: when a single instance belongs to the concept, then label bag positive
  :-(Noise sensitive, and not always applicable

- MIL on distributions: all instances contribute to the bag dissimilarities
  :-(Background may have large influence

- MIL pairwise dissim.: select important pairwise distances as features
  :-(Completely (?) unclear what it models

Results

- Test on a variety of datasets:

dataset	nr.inst.	dim.	pos. bags	neg. bags	min. inst/bag	median inst/bag	max. inst/bag
MUSK 1 [2]	476	166	47	45	2	4	40
MUSK 2 [2]	6598	166	39	63	1	12	1044
Corel African [19]	7947	9	100	1900	2	3	13
Corel Historical [19]	7947	9	100	1900	2	3	13
SIVAL AjaxOrange [23]	47414	30	60	1440	31	32	32
News atheism [22]	5443	200	50	50	22	58	76
News motorcycles [22]	4730	200	50	50	22	49	73
News mideast [22]	3373	200	50	50	15	34	55
Corel Fox [6]	1320	230	100	100	2	6	13
Corel Tiger [6]	1220	230	100	100	1	6	13
Corel Elephant [6]	1391	230	100	100	2	7	13
Web recomm.[24]	2212	5863	17	58	4	24	141

Results

classifier	Musk 1	Musk 2	Corel African	Corel Historical
Standard MIL classifiers				
APR $\tau = 0.999$	81.8 (1.3)	82.5 (1.2)	50.5 (0.0)	50.5 (0.1)
APR $\tau = 0.995$	78.9 (1.7)	80.8 (2.3)	57.4 (0.8)	61.4 (0.4)
Diverse Density (100 restarts)	89.4 (1.3)	**93.2 (0.0)**	85.6 (0.1)	83.4 (0.7)
MiBoost ($M = 100$ rounds)	80.3 (3.1)	49.3 (3.7)	68.0 (0.0)	80.4 (1.6)
MI-SVM (linear kernel)	70.3 (3.0)	81.5 (2.1)	63.4 (2.0)	78.9 (0.6)
MI-SVM (RBG kernel)	**92.9 (1.3)**	92.9 (1.6)	NaN (0.0)	**90.8 (1.0)**
MILES (linear kernel)	89.3 (1.9)	88.8 (1.8)	**88.5 (0.5)**	NaN (0.0)
MILES (RBF kernel)	**92.8 (1.4)**	**95.3 (1.5)**	58.9 (9.2)	60.8 (12.8)
Simple MIL with LDA, max-comb.	72.9 (3.4)	76.7 (3.4)	68.8 (0.2)	74.4 (0.2)
LDA on mean-inst	85.7 (1.4)	87.6 (2.8)	77.2 (0.3)	86.2 (0.1)
LDA on extremes	**92.4 (1.9)**	88.9 (4.0)	88.5 (0.1)	85.3 (0.2)
BagOfWords (k=10)+linear SVM	72.7 (4.7)	63.7 (6.1)	75.1 (3.2)	78.4 (3.9)
BagOfWords (k=100)+linear SVM	78.7 (5.5)	71.2 (3.1)	83.4 (1.8)	85.6 (2.6)
Distance-based classifiers on bag dissimilarities				
minmin+k-NND	90.1 (1.4)	84.0 (1.9)	86.6 (0.4)	84.1 (1.2)
mindist+k-NND	86.3 (2.0)	83.2 (1.6)	**92.7 (0.7)**	**90.7 (1.1)**
hausd.+k-NND	89.0 (1.6)	84.2 (0.8)	86.7 (0.9)	88.5 (1.0)
mahal.+k-NND	61.8 (2.8)	65.7 (5.7)	67.3 (0.7)	63.2 (1.2)
emd+k-NND	90.1 (2.7)	(1)	**92.0 (0.7)**	88.8 (1.7)
lin.ass.+kNND	84.7 (1.6)	76.5 (2.7)	69.9 (0.6)	87.8 (0.4)
Standard classifiers on bag dissimilarity space				
minmin.+Parzen Classifier	**94.7 (3.0)**	92.3 (2.7)	90.4 (0.6)	84.0 (0.6)
mindist.+Parzen Classifier	61.2 (6.0)	50.0 (0.0)	83.4 (0.9)	86.0 (0.5)
hausd.+Parzen Classifier	86.9 (0.7)	92.1 (2.5)	79.1 (0.6)	84.3 (0.5)
mahal.+Parzen Classifier	52.1 (0.9)	65.8 (2.4)	46.3 (2.4)	52.4 (1.3)
emd+Parzen Classifier	87.4 (3.4)	(1)	89.4 (0.4)	85.4 (0.7)
lin.ass.+Parzen Classifier	83.3 (2.7)	72.2 (2.9)	83.5 (0.7)	86.2 (0.5)
minmin.+k-NN	**93.3 (1.5)**	90.7 (3.9)	88.7 (0.8)	83.5 (1.3)
mindist.+k-NN	88.8 (3.0)	83.8 (1.4)	81.7 (1.1)	85.5 (1.0)
hausd.+k-NN	89.2 (2.7)	91.6 (1.0)	77.0 (0.7)	80.0 (1.3)
mahal.+k-NN	72.0 (3.1)	61.6 (2.7)	53.3 (1.6)	57.0 (0.8)
emd+k-NN	**92.4 (1.4)**	(1)	86.9 (1.1)	79.6 (1.5)
lin.ass.+k-NN	88.6 (2.1)	72.6 (3.7)	81.5 (1.4)	84.7 (1.4)

Zoom in (1)

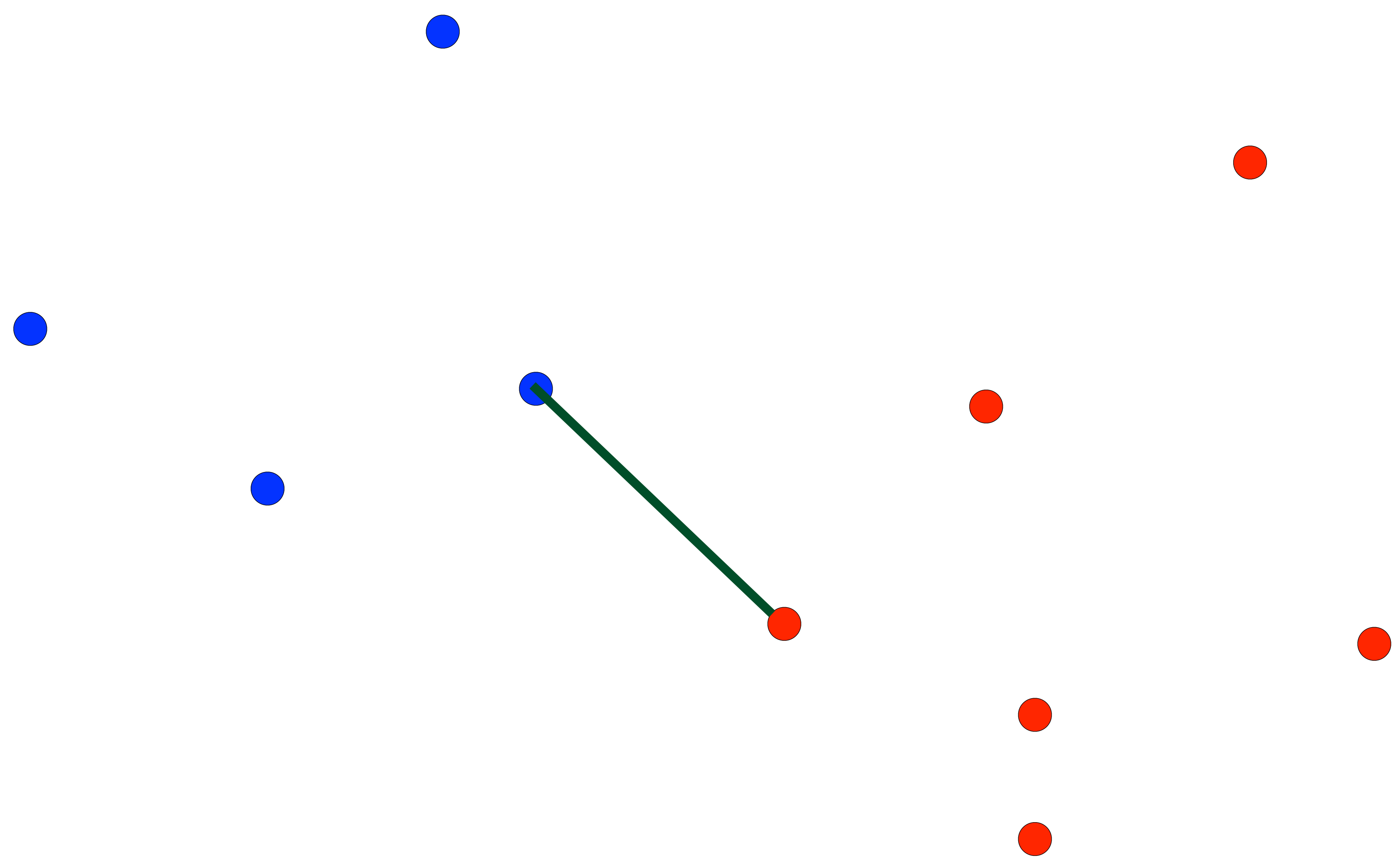
- The classic approaches:

classifier	Musk 1	Musk 2	Corel African	Corel Historical
		Standard MIL classifiers		
APR $\tau = 0.999$	81.8 (1.3)	82.5 (1.2)	50.5 (0.0)	50.5 (0.1)
APR $\tau = 0.995$	78.9 (1.7)	80.8 (2.3)	57.4 (0.8)	61.4 (0.4)
Diverse Density (100 restarts)	89.4 (1.3)	**93.2 (0.0)**	85.6 (0.1)	83.4 (0.7)
MiBoost ($M = 100$ rounds)	80.3 (3.1)	49.3 (3.7)	68.0 (0.0)	80.4 (1.6)
MI-SVM (linear kernel)	70.3 (3.0)	81.5 (2.1)	63.4 (2.0)	78.9 (0.6)
MI-SVM (RBG kernel)	**92.9 (1.3)**	92.9 (1.6)	NaN (0.0)	**90.8 (1.0)**
MILES (linear kernel)	89.3 (1.9)	88.8 (1.8)	**88.5 (0.5)**	NaN (0.0)
MILES (RBF kernel)	**92.8 (1.4)**	**95.3 (1.5)**	58.9 (9.2)	60.8 (12.8)
Simple MIL with LDA, max-comb.	72.9 (3.4)	76.7 (3.4)	68.8 (0.2)	74.4 (0.2)
LDA on mean-inst	85.7 (1.4)	87.6 (2.8)	77.2 (0.3)	86.2 (0.1)
LDA on extremes	**92.4 (1.9)**	88.9 (4.0)	88.5 (0.1)	85.3 (0.2)
BagOfWords (k=10)+linear SVM	72.7 (4.7)	63.7 (6.1)	75.1 (3.2)	78.4 (3.9)
BagOfWords (k=100)+linear SVM	78.7 (5.5)	71.2 (3.1)	83.4 (1.8)	85.6 (2.6)

Zoom in (2)

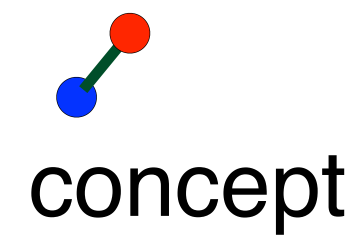
Distance-based classifiers on bag dissimilarities				
minmin+k-NND	90.1 (1.4)	84.0 (1.9)	86.6 (0.4)	84.1 (1.2)
mindist+k-NND	86.3 (2.0)	83.2 (1.6)	**92.7 (0.7)**	**90.7 (1.1)**
hausd.+k-NND	89.0 (1.6)	84.2 (0.8)	86.7 (0.9)	88.5 (1.0)
mahal.+k-NND	61.8 (2.8)	65.7 (5.7)	67.3 (0.7)	63.2 (1.2)
emd+k-NND	90.1 (2.7)	(1)	**92.0 (0.7)**	88.8 (1.7)
lin.ass.+kNND	84.7 (1.6)	76.5 (2.7)	69.9 (0.6)	87.8 (0.4)
Standard classifiers on bag dissimilarity space				
minmin.+Parzen Classifier	**94.7 (3.0)**	92.3 (2.7)	90.4 (0.6)	84.0 (0.6)
mindist.+Parzen Classifier	61.2 (6.0)	50.0 (0.0)	83.4 (0.9)	86.0 (0.5)
hausd.+Parzen Classifier	86.9 (0.7)	92.1 (2.5)	79.1 (0.6)	84.3 (0.5)
mahal.+Parzen Classifier	52.1 (0.9)	65.8 (2.4)	46.3 (2.4)	52.4 (1.3)
emd+Parzen Classifier	87.4 (3.4)	(1)	89.4 (0.4)	85.4 (0.7)
lin.ass.+Parzen Classifier	83.3 (2.7)	72.2 (2.9)	83.5 (0.7)	86.2 (0.5)
minmin.+k-NN	**93.3 (1.5)**	90.7 (3.9)	88.7 (0.8)	83.5 (1.3)
mindist.+k-NN	88.8 (3.0)	83.8 (1.4)	81.7 (1.1)	85.5 (1.0)
hausd.+k-NN	89.2 (2.7)	91.6 (1.0)	77.0 (0.7)	80.0 (1.3)
mahal.+k-NN	72.0 (3.1)	61.6 (2.7)	53.3 (1.6)	57.0 (0.8)
emd+k-NN	**92.4 (1.4)**	(1)	86.9 (1.1)	79.6 (1.5)
lin.ass.+k-NN	88.6 (2.1)	72.6 (3.7)	81.5 (1.4)	84.7 (1.4)

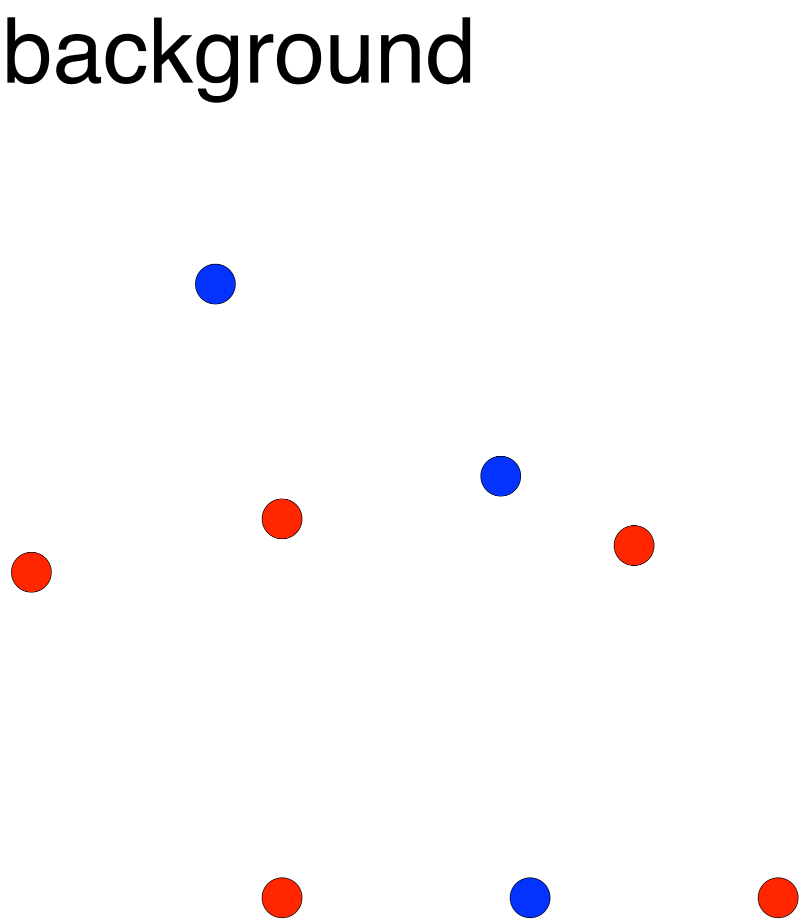
Min-min distance?



Min-min distance?

Min-min distance?

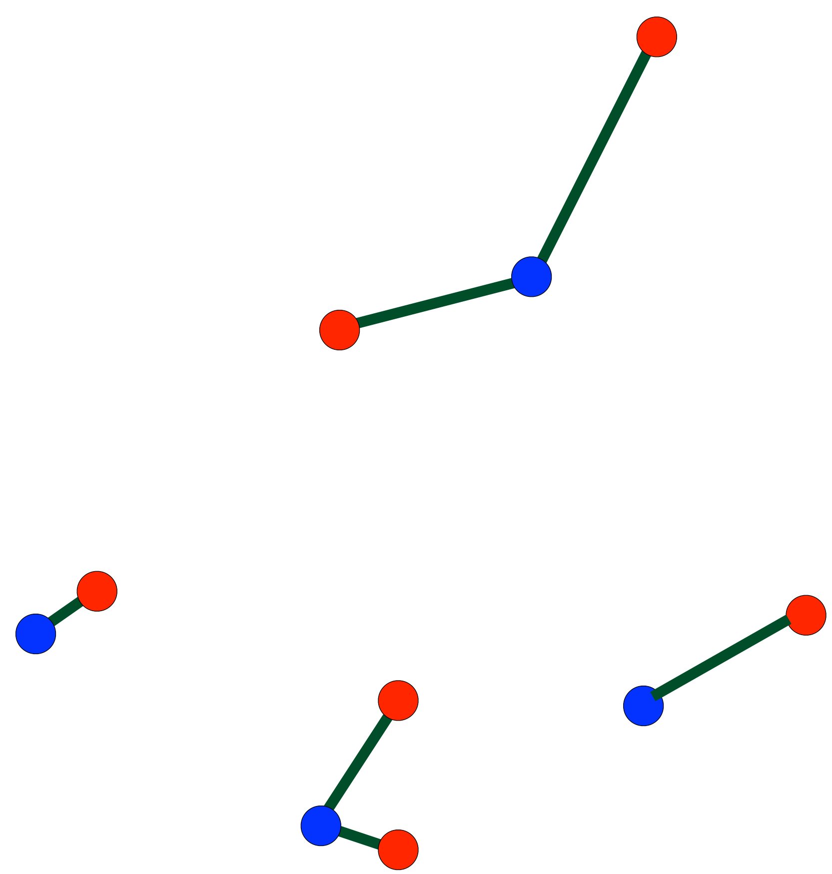




Noise sensitive?!!
Instances in the concept have be more similar than in the background.

Mean-min. distance?

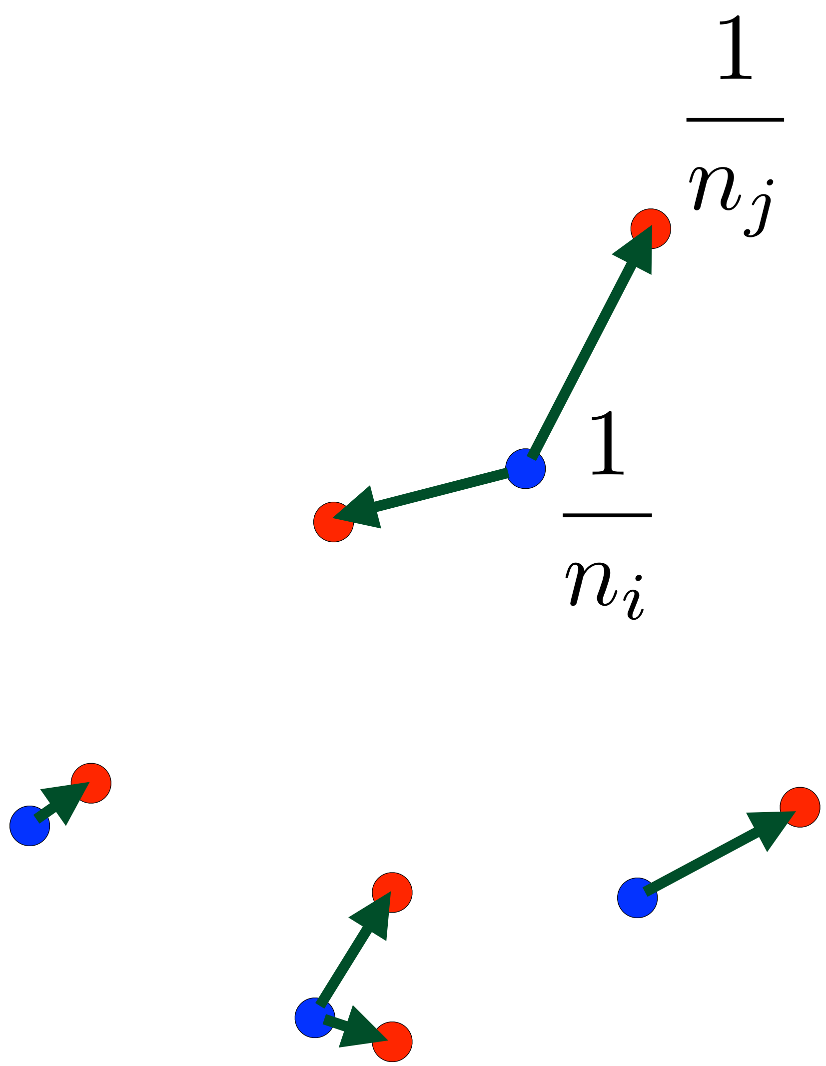
- Average the minimum distances: take all instances into account



no concept?!!

Earth movers distance?

- How much work does it take to move probability mass from B_i to B_j



no concept?!!

More results

classifier	AjaxOrange	alt.atheism	rec.motorcycles	politics.mideast
Standard MIL classifiers				
APR $\tau = 0.995$	48.4 (0.8)	50.0 (0.0)	50.0 (0.0)	49.8 (0.4)
Diverse Density (100 restarts)	55.5 (2.9)	52.2 (2.4)	46.4 (2.9)	40.2 (2.5)
MiBoost ($M = 100$ rounds)	56.5 (2.4)	50.0 (0.0)	NaN (0.0)	50.3 (1.5)
MI-SVM (linear kernel)	**93.6 (2.6)**	69.8 (2.8)	76.4 (4.0)	79.8 (2.3)
MI-SVM (RBG kernel)	NaN (0.0)	45.5 (7.1)	49.7 (5.4)	46.1 (2.4)
MILES (linear kernel)	(2)	**80.4 (1.2)**	**77.4 (1.9)**	**79.9 (3.4)**
MILES (RBF kernel)	(2)	47.1 (4.5)	44.7 (4.8)	54.1 (1.8)
Simple MIL with LDA, max-comb.	89.3 (0.3)	81.6 (1.2)	80.4 (2.1)	75.0 (3.1)
LDA on mean-inst	82.3 (0.9)	83.7 (2.1)	84.4 (1.8)	78.1 (1.7)
LDA on extremes	90.3 (0.3)	50.0 (0.0)	51.2 (0.4)	65.0 (1.8)
BagOfWords (k=100)+linear SVM	81.2 (2.5)	54.0 (0.0)	65.2 (9.3)	58.6 (6.8)
Distance-based classifiers on bag dissimilarities				
minmin+k-NND	53.6 (1.2)	50.0 (0.0)	50.0 (0.0)	52.8 (2.2)
mindist+k-NND	62.9 (1.3)	59.2 (1.9)	58.4 (0.5)	53.4 (1.1)
hausd.+k-NND	72.4 (1.3)	72.8 (3.0)	68.7 (3.2)	67.1 (1.8)
mahal.+k-NND	64.0 (1.6)	47.7 (4.4)	45.0 (3.4)	58.5 (6.0)
emd+k-NND	77.6 (2.6)	56.0 (1.2)	60.8 (0.4)	57.2 (1.3)
lin.ass.+kNND	71.6 (1.4)	69.2 (1.7)	53.7 (2.9)	58.5 (3.2)
Standard classifiers on bag dissimilarity space				
minmin.+Parzen Classifier	55.7 (1.6)	49.8 (0.4)	50.0 (0.0)	50.4 (2.3)
mindist.+Parzen Classifier	78.0 (1.3)	78.9 (2.8)	78.4 (0.5)	75.2 (1.9)
hausd.+Parzen Classifier	71.8 (0.9)	73.8 (2.0)	82.0 (2.2)	73.8 (0.9)
mahal.+Parzen Classifier	75.3 (0.9)	54.2 (3.3)	43.7 (3.5)	61.9 (1.8)
emd+Parzen Classifier	78.7 (1.1)	**89.7 (1.3)**	77.6 (1.5)	**87.8 (1.1)**
lin.ass.+Parzen Classifier	78.9 (0.6)	80.1 (2.4)	84.2 (2.8)	84.3 (3.1)
minmin.+k-NN	56.0 (1.6)	50.0 (0.0)	50.0 (0.0)	47.8 (2.7)
mindist.+k-NN	70.6 (2.6)	84.9 (1.6)	**86.6 (2.0)**	82.2 (1.5)
hausd.+k-NN	68.9 (1.9)	85.6 (2.1)	**89.2 (3.5)**	77.2 (3.2)
mahal.+k-NN	70.8 (1.5)	51.2 (3.6)	56.3 (3.8)	55.8 (4.6)
emd+k-NN	72.0 (2.4)	**90.0 (1.4)**	**86.7 (0.7)**	82.6 (1.7)
lin.ass.+k-NN	70.1 (0.8)	82.1 (2.3)	82.9 (2.4)	80.8 (3.8)

Zoomed...

Distance-based classifiers on bag dissimilarities				
minmin+k-NND	53.6 (1.2)	50.0 (0.0)	50.0 (0.0)	52.8 (2.2)
mindist+k-NND	62.9 (1.3)	59.2 (1.9)	58.4 (0.5)	53.4 (1.1)
hausd.+k-NND	72.4 (1.3)	72.8 (3.0)	68.7 (3.2)	67.1 (1.8)
mahal.+k-NND	64.0 (1.6)	47.7 (4.4)	45.0 (3.4)	58.5 (6.0)
emd+k-NND	77.6 (2.6)	56.0 (1.2)	60.8 (0.4)	57.2 (1.3)
lin.ass.+kNND	71.6 (1.4)	69.2 (1.7)	53.7 (2.9)	58.5 (3.2)
Standard classifiers on bag dissimilarity space				
minmin.+Parzen Classifier	55.7 (1.6)	49.8 (0.4)	50.0 (0.0)	50.4 (2.3)
mindist.+Parzen Classifier	78.0 (1.3)	78.9 (2.8)	78.4 (0.5)	75.2 (1.9)
hausd.+Parzen Classifier	71.8 (0.9)	73.8 (2.0)	82.0 (2.2)	73.8 (0.9)
mahal.+Parzen Classifier	75.3 (0.9)	54.2 (3.3)	43.7 (3.5)	61.9 (1.8)
emd+Parzen Classifier	78.7 (1.1)	**89.7 (1.3)**	77.6 (1.5)	**87.8 (1.1)**
lin.ass.+Parzen Classifier	78.9 (0.6)	80.1 (2.4)	84.2 (2.8)	84.3 (3.1)
minmin.+k-NN	56.0 (1.6)	50.0 (0.0)	50.0 (0.0)	47.8 (2.7)
mindist.+k-NN	70.6 (2.6)	84.9 (1.6)	**86.6 (2.0)**	82.2 (1.5)
hausd.+k-NN	68.9 (1.9)	85.6 (2.1)	**89.2 (3.5)**	77.2 (3.2)
mahal.+k-NN	70.8 (1.5)	51.2 (3.6)	56.3 (3.8)	55.8 (4.6)
emd+k-NN	72.0 (2.4)	**90.0 (1.4)**	**86.7 (0.7)**	82.6 (1.7)
lin.ass.+k-NN	70.1 (0.8)	82.1 (2.3)	82.9 (2.4)	80.8 (3.8)

Are there different MIL problems?

- 'Spice shop'   OR   'Red chilli' ??

Conclusions

- Bag dissimilarities offer possibilities for MIL
- The idea of a ‘concept’ is often not clear, the bag distribution is more important

- Promising: the ‘mindist’ and the Earth Movers Distance

  ? This suggest that
  (1) the full bag distribution is informative, or
  (2) there may be insufficient nr. of instances to describe the concept well

TU Delft