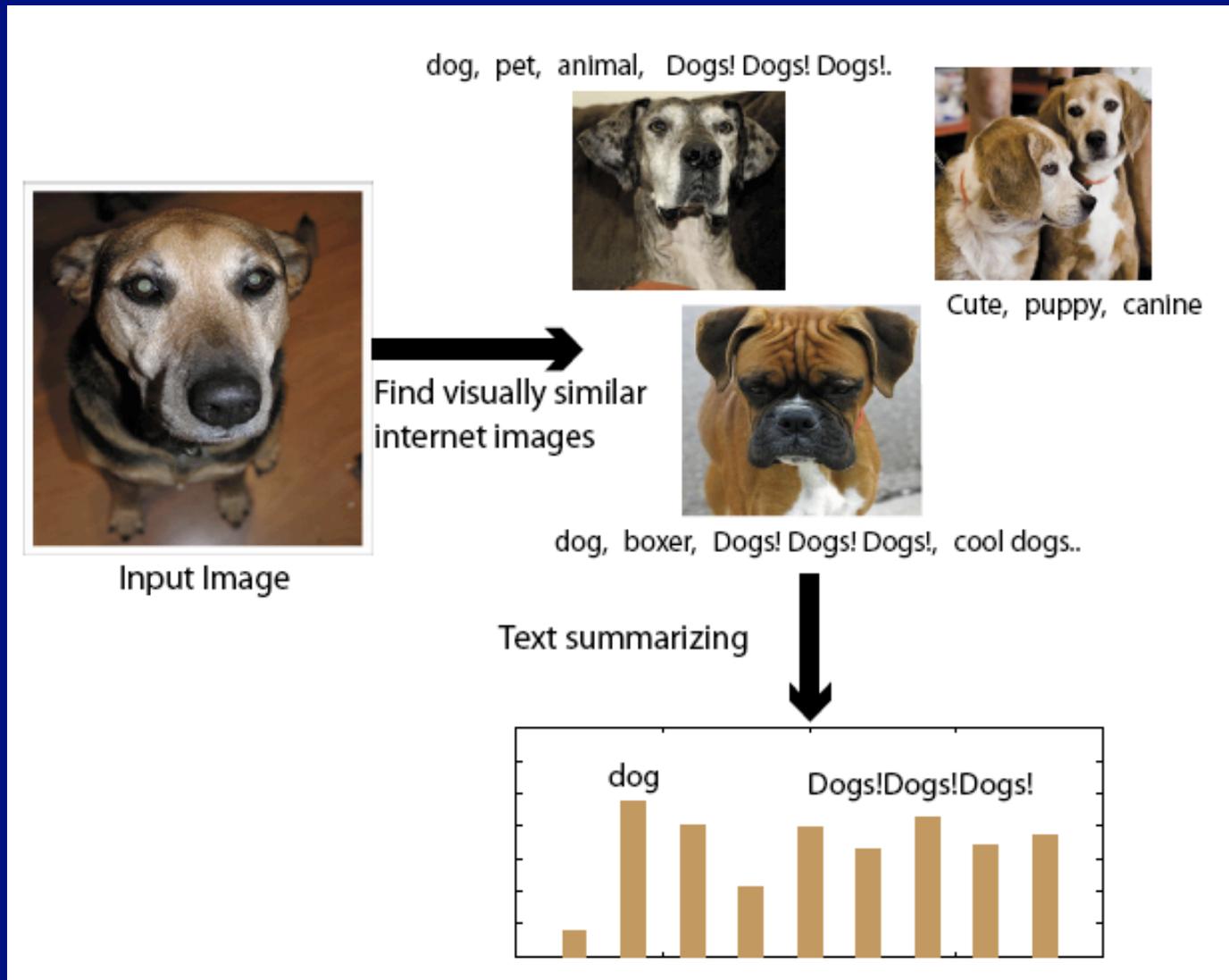# Words and Pictures: Variants

D.A. Forsyth, UIUC
with: Kobus Barnard, U.Arizona; Pinar Duygulu, Bilkent U.;
Nando de Freitas, UBC; Tamara Berg, UIUC; Derek Hoiem,
UIUC; Ian Endres, UIUC; Ali Farhadi, UIUC; Gang Wang,
UIUC;

A=picture, B=words

Names + Faces - special case

President George W. Bush makes a statement in the Rose Garden while Secretary of Defense Donald Rumsfeld looks on, July 23, 2003. Rumsfeld said the United States would release graphic photographs of the dead sons of Saddam Hussein to prove they were killed by American troops. Photo by Larry Downing/Reuters

Classical problem

- Edges with explicit location
  - eg "Smith (left) and Brown (right)", Govindaraju 89, Srihari 95
- Smoothed association
  - cf multiple instance learning, Satoh 97, 99, Yang 04
- Multiple instance learning
  - anchors Song 04
  - news video Yang 05

News dataset

- Approx 5e5 news images, with captions
  - Easily collected by script from Yahoo over the last 18 months or so
- Mainly people
  - politicians, actors, sportsplayers
  - long, long tails distribution
- Face pictures captured "in the wild"
- Correspondence problem
  - some images have many (resp. few) faces, few (resp. many) names (cf. Srihari 95)

Data examples

Doctor Nikola shows a fork that was removed from an Israeli woman who swallowed it while trying to catch a bug that flew in to her mouth, in Poriah Hospital northern Israel July 10, 2003. Doctors performed emergency surgery and removed the fork. (Reuters)

President George W. Bush waves as he leaves the White House for a day trip to North Carolina, July 25, 2002. A White House spokesman said that Bush would be compelled to veto Senate legislation creating a new department of homeland security unless changes are made. (Kevin Lamarque/Reuters)

Process

- Extract proper names
  - rather crudely, at present
- Detect faces
  - with Cordelia Schmid's face detector, (Vogelhuber Schmid 00)
- Rectify faces
  - by finding eye, nose, mouth patches, affine transformation
- Kernel PCA rectified faces
- Estimate linear discriminants
- Now have (face vector; name_1,...., name_k)

Scale

44773 big face responses

34623 properly rectified

27742   for k<=4

Building a face dictionary

- Compute linear discriminants
  - using single name, single face data items
  - we now have a set of clusters

- Now break correspondence with modified k-means
  - assign face to cluster with closest center,
    - chosen from associated names
  - recompute centers, iterate
  - using distance in LD space

- Now recompute discriminants, recluster with modified k-means

President George
State Colin Powell

US President George W. Bush (L) makes remarks while Secretary of State Colin Powell (R) listens before signing the US Leadership Against HIV /AIDS , Tuberculosis and Malaria Act of 2003 at the Department of State in Washington, DC. The five-year plan is designed to help prevent and treat AIDS, especially in more than a dozen African and Caribbean nations(AFP/ Luke Frazza)


Claudia Schiffer

German supermodel Claudia Schiffer gave birth to a baby boy by Caesarian section January 30, 2003, her spokeswoman said. The baby is the first child for both Schiffer, 32, and her husband, British film producer Matthew Vaughn, who was at her side for the birth. Schiffer is seen on the German television show 'Bet It...?!' ('Wetten Dass...?!') in Braunschweig, on January 26, 2002. (Alexandra Winkler/Reuters)

Kate Winslet
Sam Mendes

British director Sam Mendes and his partner actress Kate Winslet arrive at the London premiere of 'The Road to Perdition', September 18, 2002. The films stars Tom Hanks as a Chicago hit man who has a separate family life and co-stars Paul Newman and Jude Law. REUTERS/Dan Chung

Pruning

- Using a likelihood model
- Tradeoff: size vs accuracy
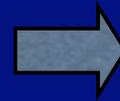
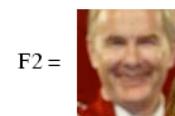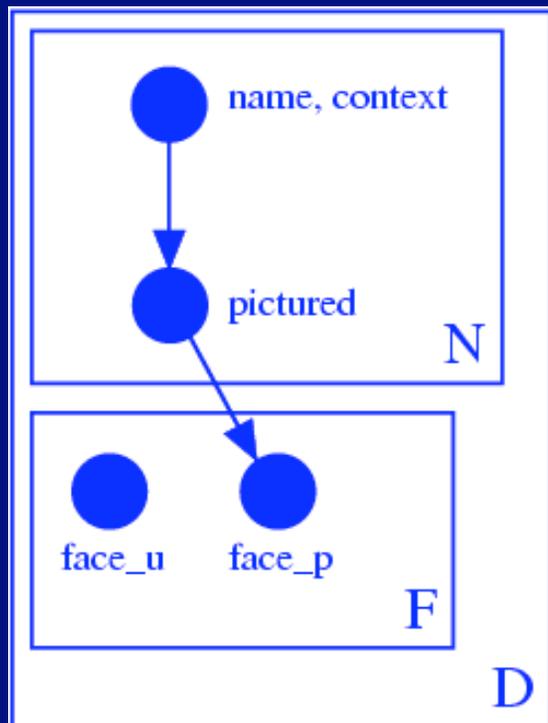# Merging

Venezuelan
President Chavez

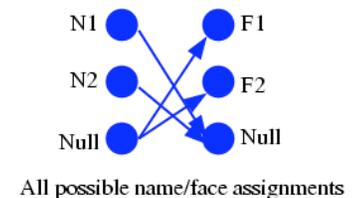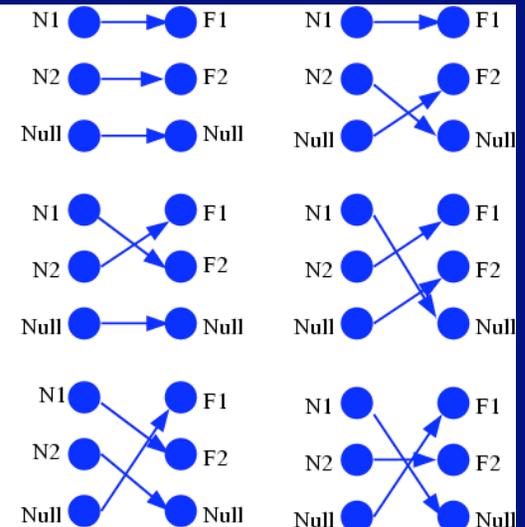
Hugo Chavez

Works - but

- We are missing language cues

Sahar Aziz, left, a law student at the University of Texas, hands the business card identifying Department of the Army special agent Jason D. Treesh to one of her attorneys, Bill Allison, right, during a news conference on Friday, Feb. 13, 2004, in Austin, Texas. In the background is Jim Harrington, director of the Texas Civil Rights Project. (AP Photo Harry Cabluck)

Training a language module

- Idea:
  - a set of named faces is supervised training data for a "who's in the picture" module
 - actually, do EM (or maximize?) over missing correspondences

President and Chief Operating Officer of the Vancouver, British Columbia 2010 Bid Corporation John Furlong (rear) smiles while celebrating with compatriots their victory in obtaining the 2010 Winter Olympics bid on late July 2, 2003 in Prague. Vancouver won with 56 votes against 53 votes for Pyeonchang in the second round of balloting at an IOC gathering in Prague. REUTERS/Petr Josek

N1 = John Furlong,
N2 = Winter Olympics,

All possible name/face assignments

Language improves naming,

before – CEO Summit
after – Martha Stewart
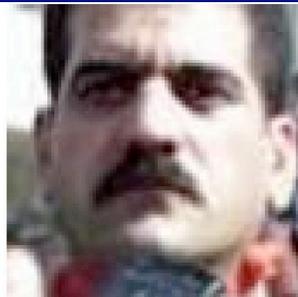
before – U.S. Joint
after – Null

before – Angelina Jolie
after – Jon Voight

before – Ric Pipino
after – Heidi Klum

before – U.S. Open
after – David Nalbandian

before – James Bond
after – Pierce Brosnan

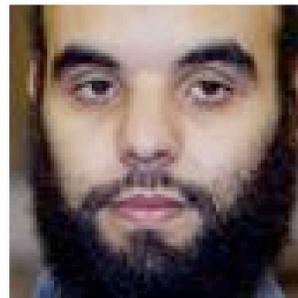before – U.S. House
after – Andrew Fastow

before – Julia Vakulenko
after – Jennifer Capriati

before – Vice President
Dick Cheney
after – President George W.

before – Marcel Avram
after – Michael Jackson

before – al Qaeda
after – Null

before – James Ivory
after – Naomi Watts

Model	EM	MM
Appearance Model, No Lang Model	56%	67%
Appearance Model + Lang Model	72%	77%

Clusters,

Angelina Jolie

Elizabeth Dole

Donald Fehr

U.S. Open

Abraham Lincoln

Anastasia Myskina
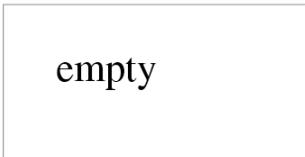
Without language model

Elizabeth Dole
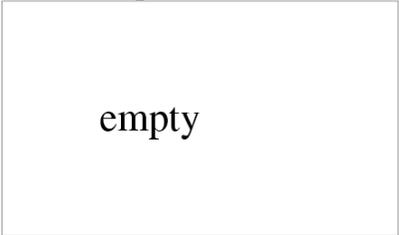
Angelina Jolie

Donald Fehr

Abraham Lincoln

empty

U.S. Open

empty

Anastasia Myskina

With language model

and yields a fun little NLP module, too

IN Pete Sampras IN of the U.S. celebrates his victory over Denmark's **OUT Kristian Pless OUT** at the **OUT U.S. Open OUT** at Flushing Meadows August 30, 2002. Sampras won the match 6-3 7- 5 6-4. REUTERS/Kevin Lamarque

Germany's **IN Chancellor Gerhard Schroeder IN**, left, in discussion with France's **IN President Jacques Chirac IN** on the second day of the EU summit at the European Council headquarters in Brussels, Friday Oct. 25, 2002. EU leaders are to close a deal Friday on finalizing entry talks with 10 candidate countries after a surprise breakthrough agreement on Thursday between France and Germany regarding farm spending.(AP Photo/European Commission/HO)

'The Right Stuff' cast members **IN Pamela Reed IN**, (L) poses with fellow cast member **IN Veronica Cartwright IN** at the 20th anniversary of the film in Hollywood, June 9, 2003. The women played wives of astronauts in the film about early United States test pilots and the space program. The film directed by **OUT Philip Kaufman OUT**, is celebrating its 20th anniversary and is being released on DVD. REUTERS/Fred Prouser

Kraft Foods Inc., the largest U.S. food company, on July 1, 2003 said it would take steps, like capping portion sizes and providing more nutrition information, as it and other companies face growing concern and even lawsuits due to rising obesity rates. In May of this year, San Francisco attorney **OUT Stephen Joseph OUT**, shown above, sought to ban Oreo cookies in California – a suit that was withdrawn less than two weeks later. Photo by Tim Wimborne/Reuters REUTERS/Tim Wimborne

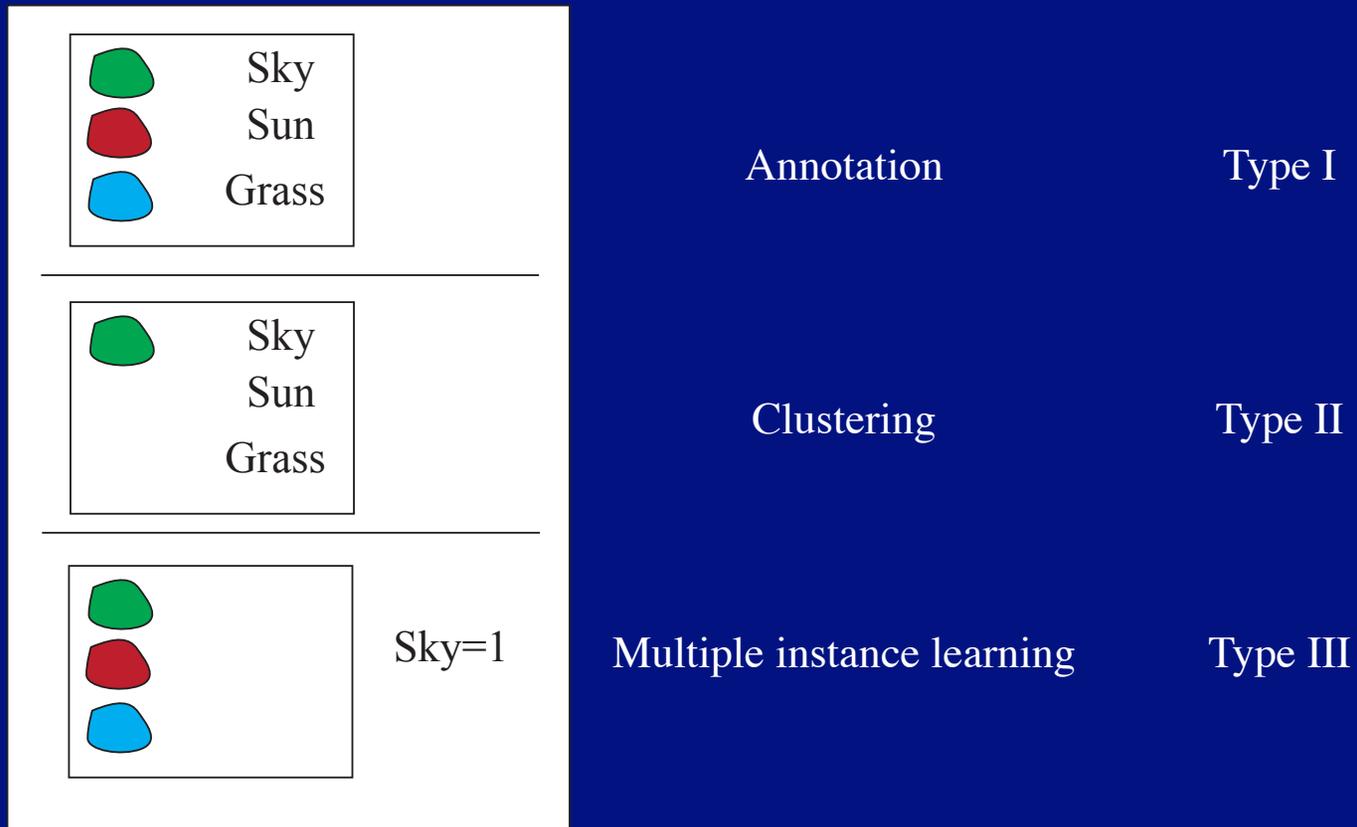| Classifier	labels correct	IN correct	OUT correct
Baseline	67%	100%	0%
EM Labeling with Language Model	76%	95%	56%
MM Labeling with Language Model	84%	87%	76%

Multiple interesting variants

- More textual analysis
  - Pham 08
- Graph based representation of faces, names
  - Ozkan 06, 06a, Guillaimin 08
- Query expansion
  - Mensink 08
- Gender, age, etc...
 - ? someone should have done this?

Datasets

- Faces in the wild
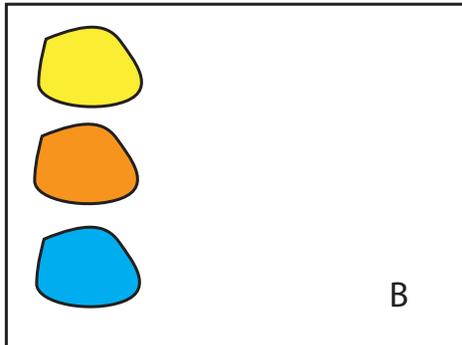 - http://www.tamaraberg.com/faceDataset/index.html
- Labelled faces in the wild
  - http://vis-www.cs.umass.edu/lfw/

Common types of problem

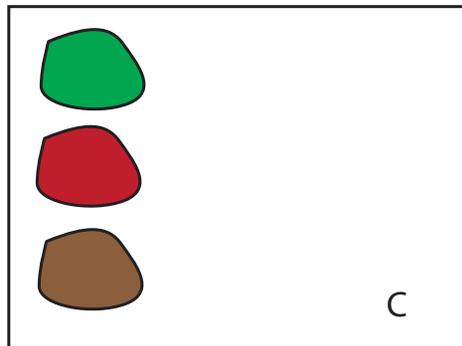|		
Sky Sun Grass	Annotation	Type I
Sky Sun Grass	Clustering	Type II
Sky=1	Multiple instance learning	Type III

Type II and Type II may not be equivalent, because blobs
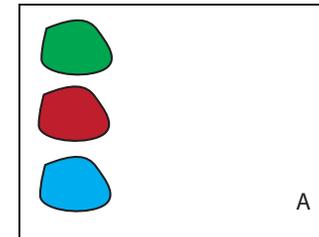come from a big space and labels from a small one
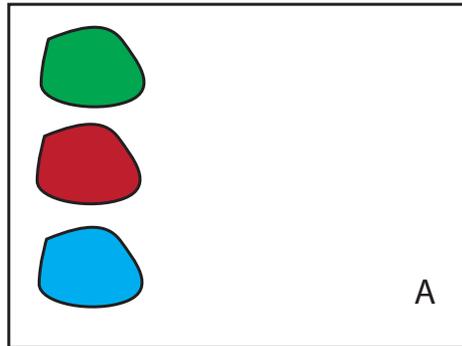
Sky=1

Sky=1

Sky=0

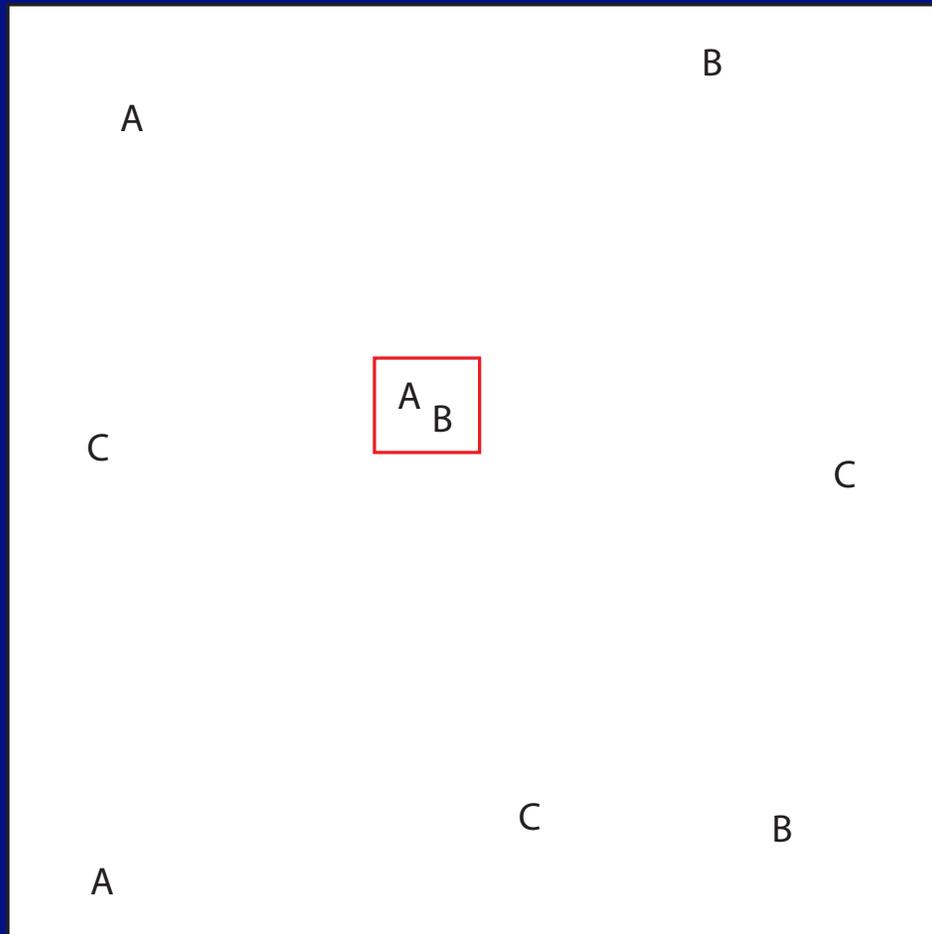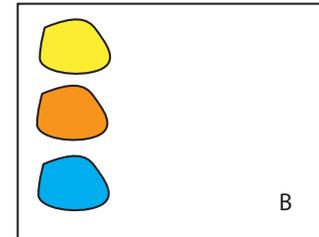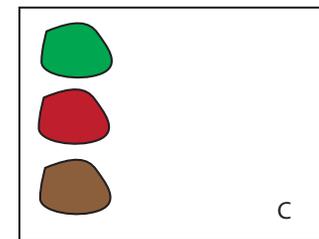# Diverse density

MI-SVM

Sky=1

Sky=1

Sky=0
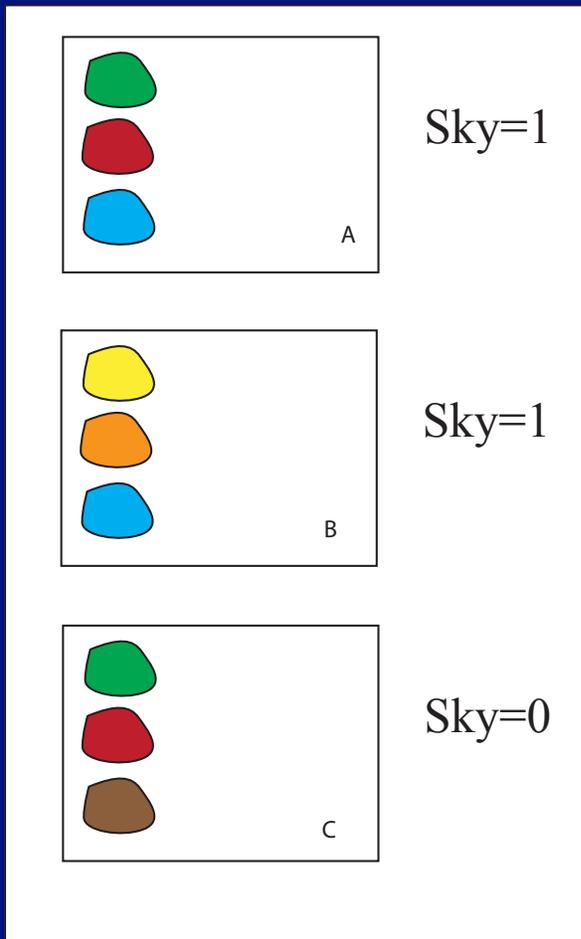
y_{1a}
y_{2a} at least one of these is positive
y_{3a}

all of these are negative

Andrews et al 03

MI-SVM

$$\min_{\xi_{ij}} \min_{w} \left[(1/2)w^T w + C \sum_{i,j} \xi_{ij} \right]$$

$$y_{ij}(w^T x_{ij} + b) \geq 1 - \xi_{ij}$$

$$\xi_{ij} \geq 0$$

$$y_{ij} \in \{-1, 1\}$$

$$y_{ij} = -1 \text{ if } j\text{'th bag is negative}$$

$$\max_{i} y_{ij} = 1 \text{ if } j\text{'th bag is positive}$$

Andrews et al 03