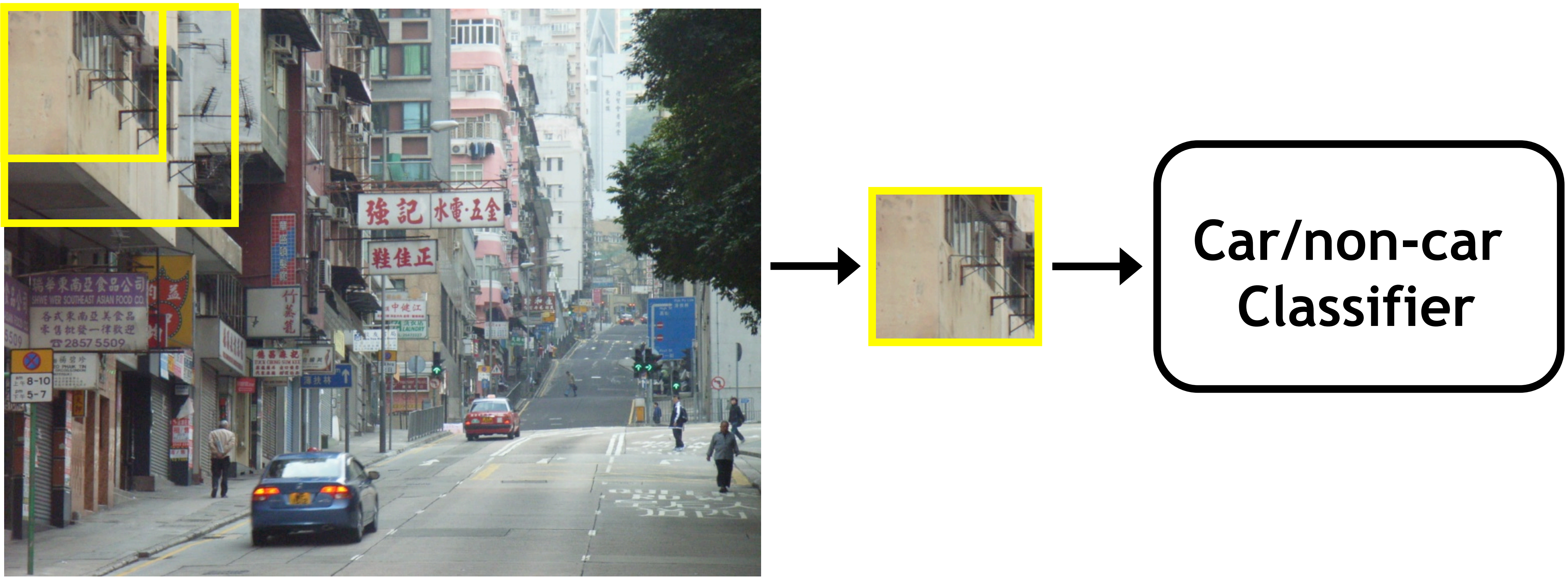
Object Recognition by Discriminative Methods

Sinisa Todorovic

1st Sino-USA Summer School in VLPR
July, 2009

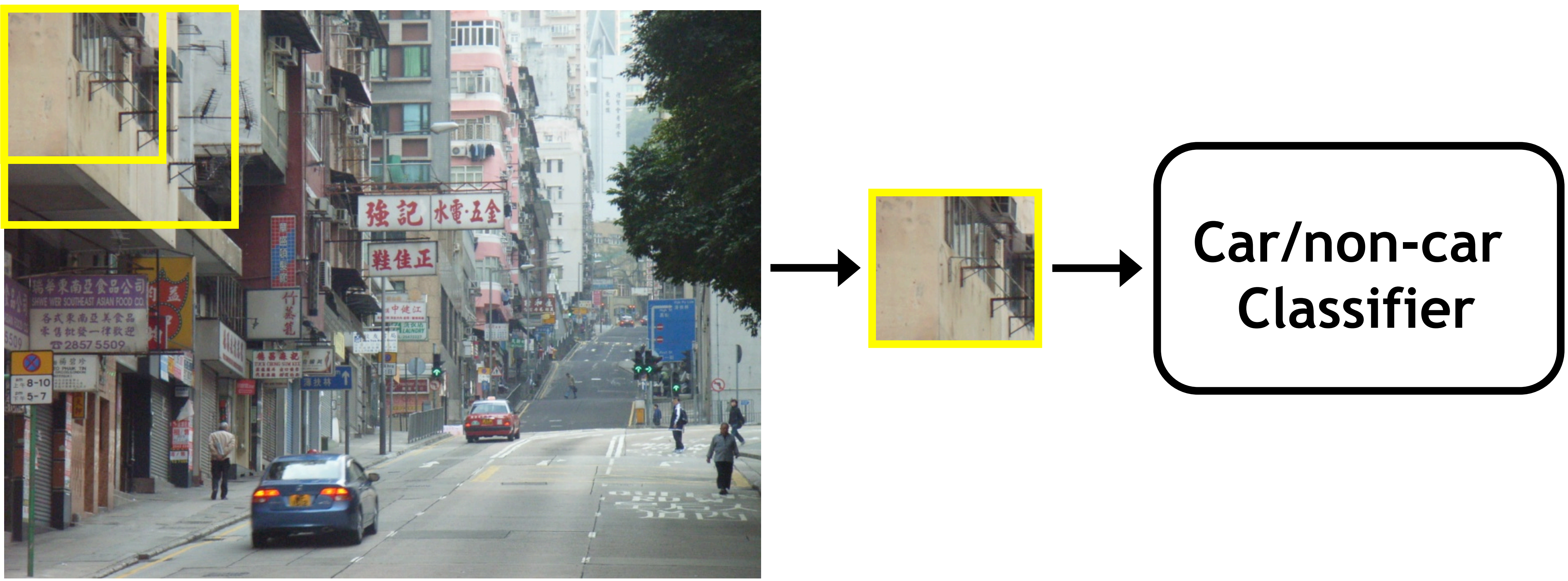
OSU
Oregon State
UNIVERSITY

Motivation



from Kristen Grauman, B. Leibe

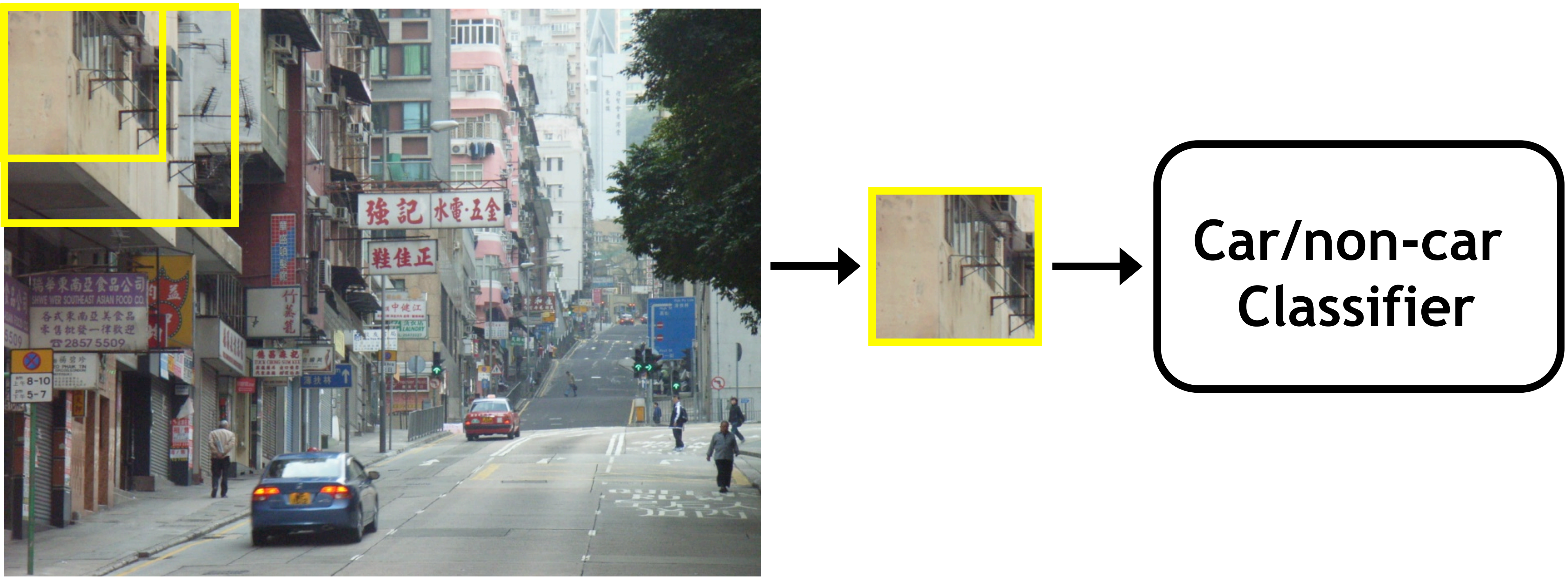
Motivation



Discriminate yes, but what?

from Kristen Grauman, B. Leibe

Motivation

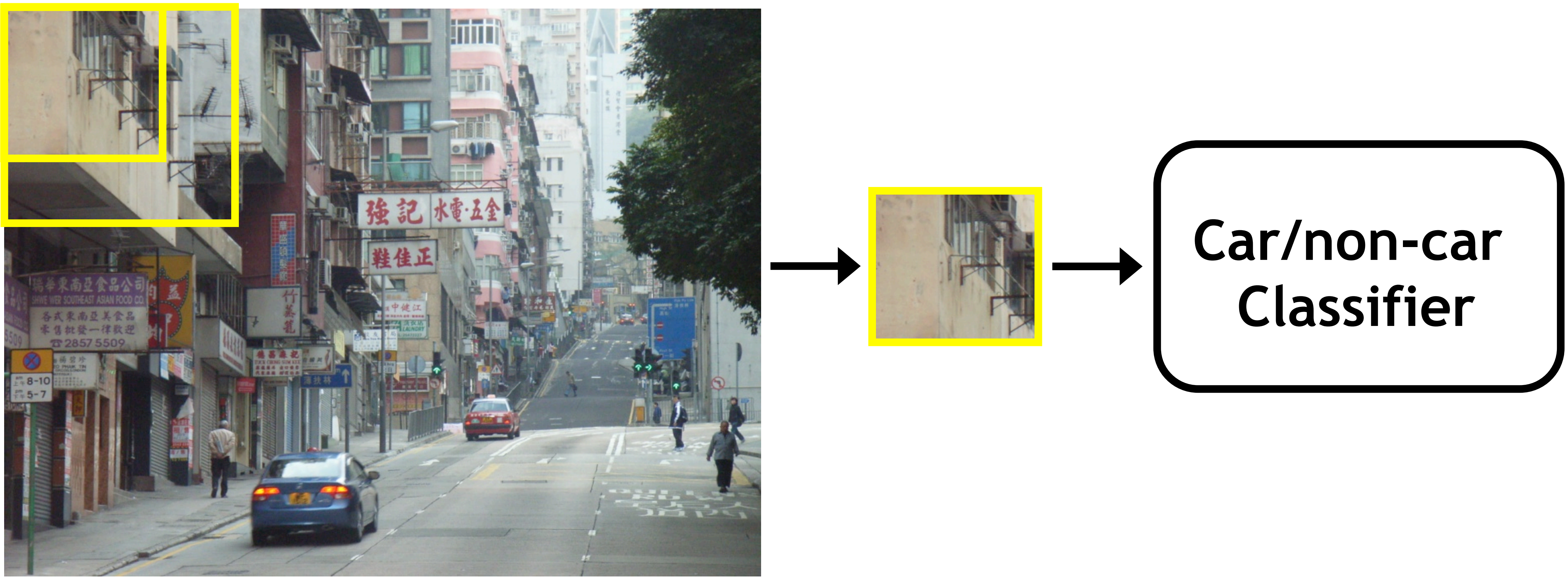


Discriminate yes, but what?

Problem 1: What are good features

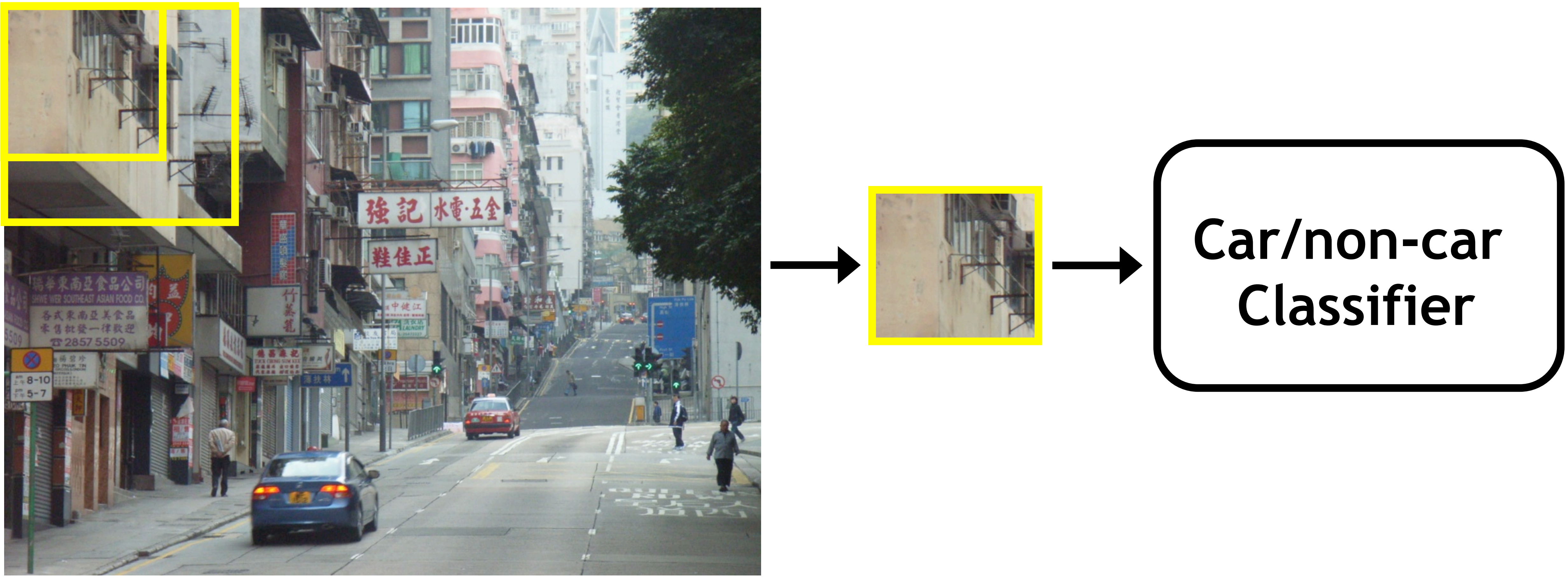
from Kristen Grauman, B. Leibe

Motivation



from Kristen Grauman, B. Leibe

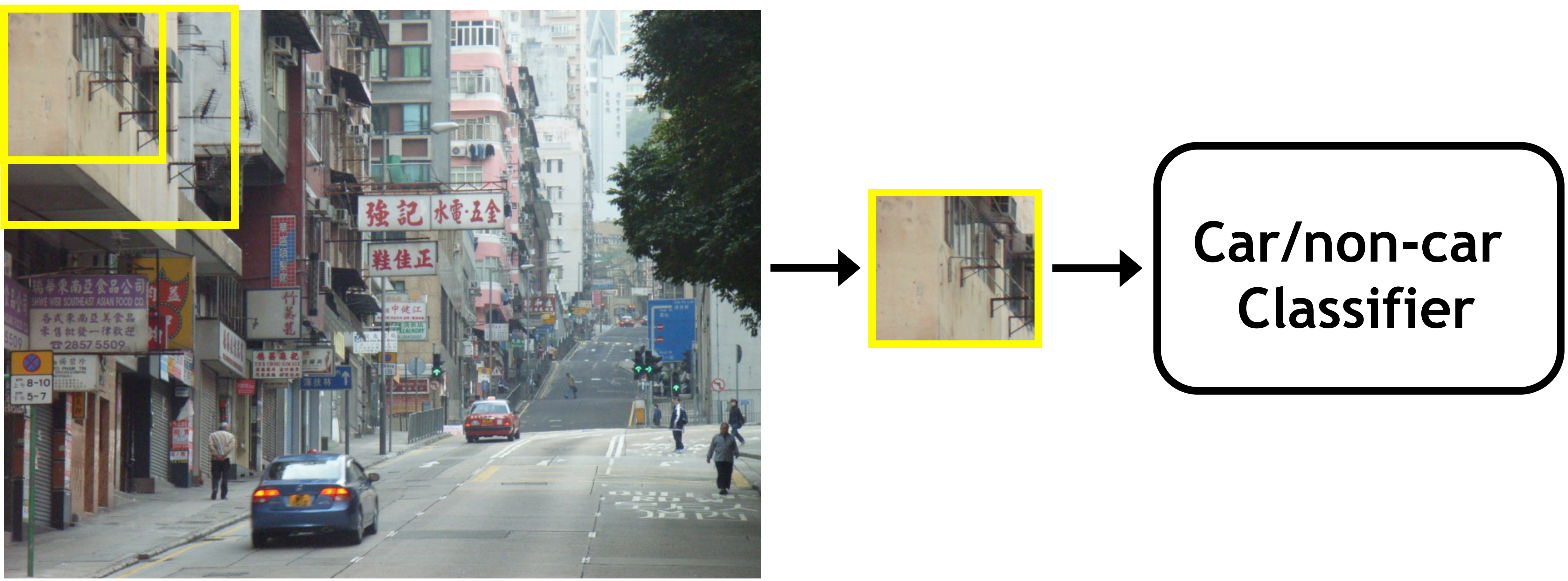
Motivation



Discriminate yes, but against what?

from Kristen Grauman, B. Leibe

Motivation

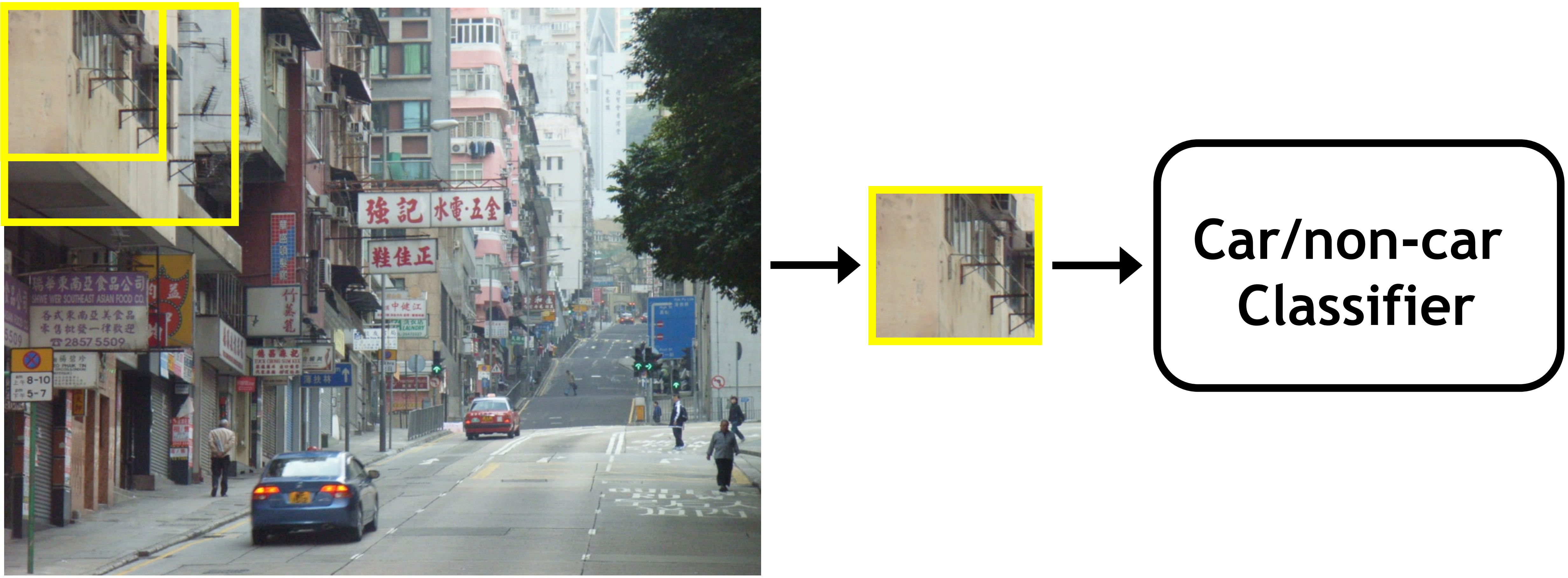


Discriminate yes, but against what?

Problem 2: What are good training examples

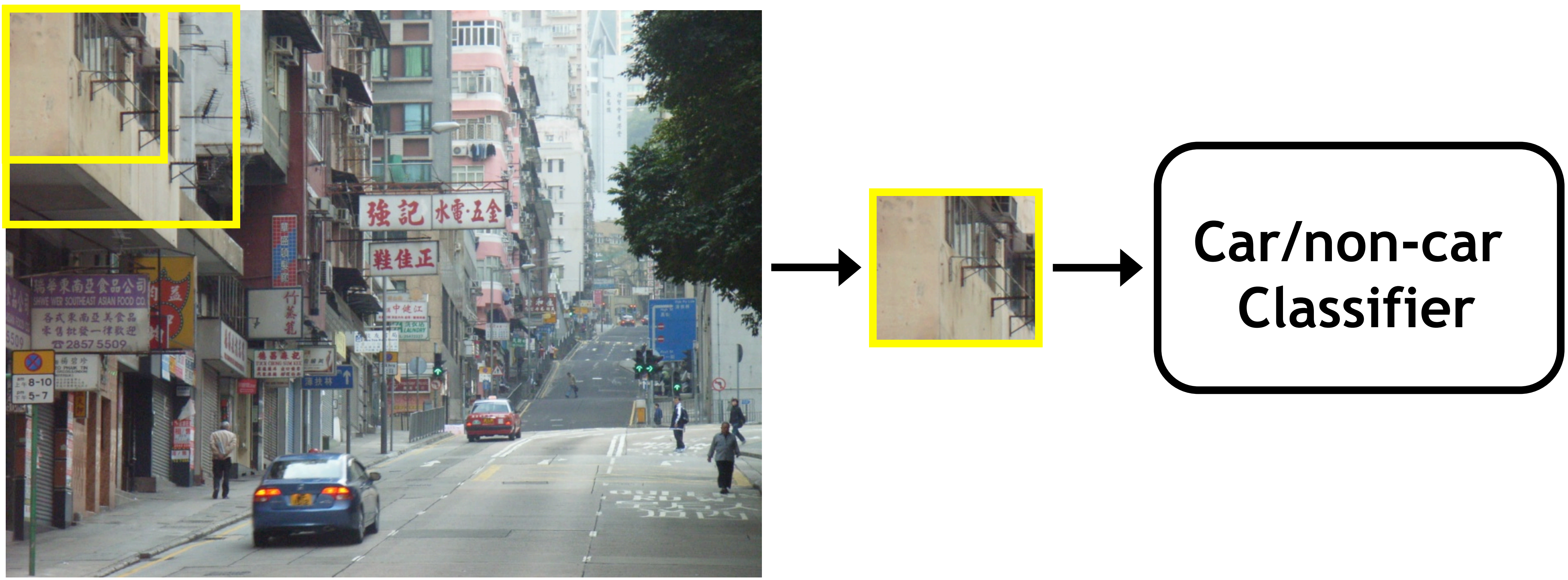
from Kristen Grauman, B. Leibe

Motivation



from Kristen Grauman, B. Leibe

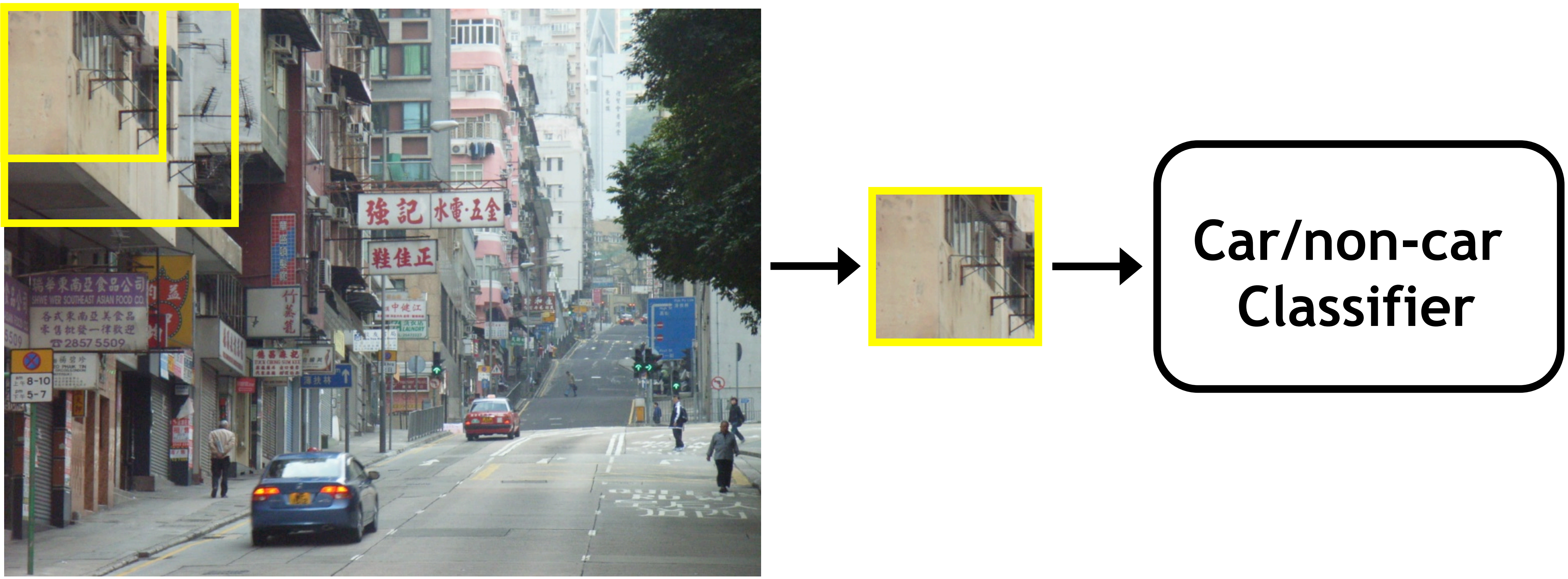
Motivation



Discriminate yes, but how?

from Kristen Grauman, B. Leibe

Motivation



Discriminate yes, but how?

Problem 3: What is a good classifier

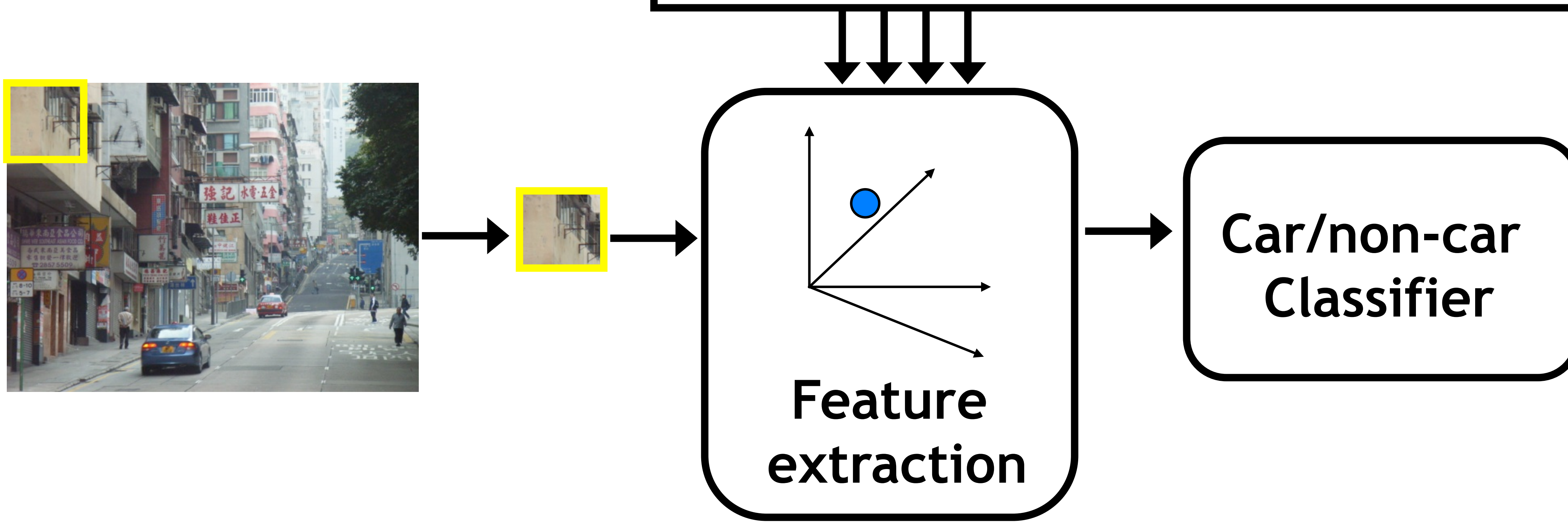
from Kristen Grauman, B. Leibe

Motivation

Sliding windows????

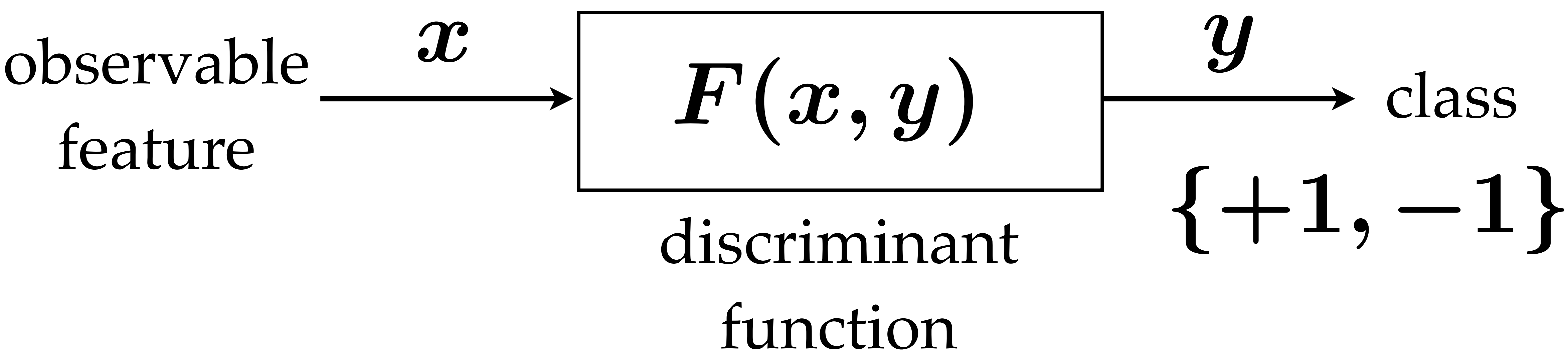
What is a non-car?

Is Bayes decision optimal?



from Kristen Grauman, B. Leibe

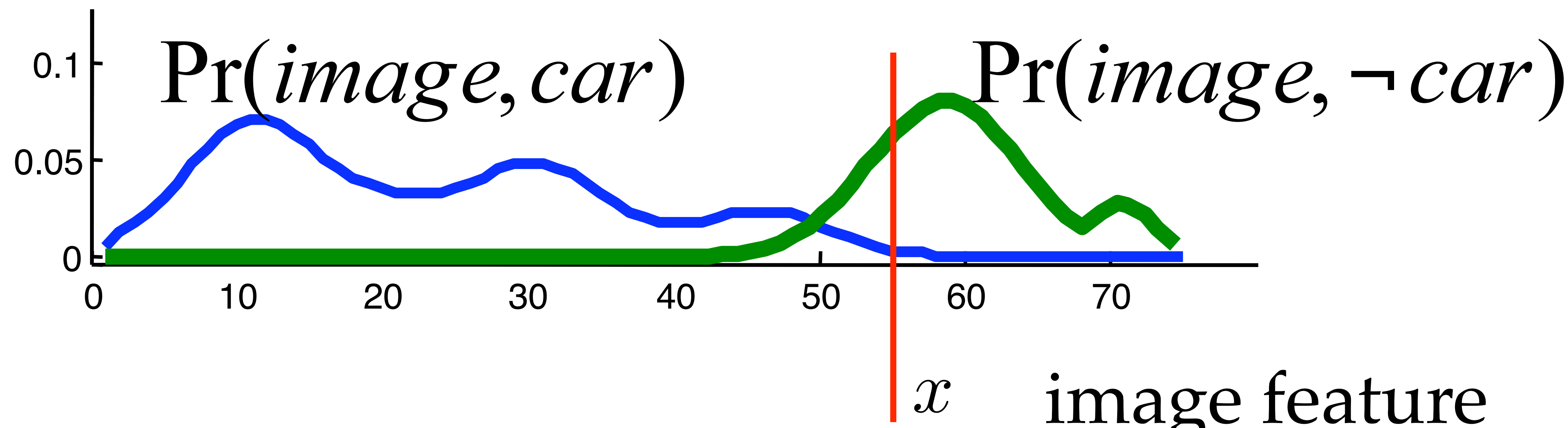
How to Classify?



Discriminant function may have
a probabilistic interpretation

How to Classify?

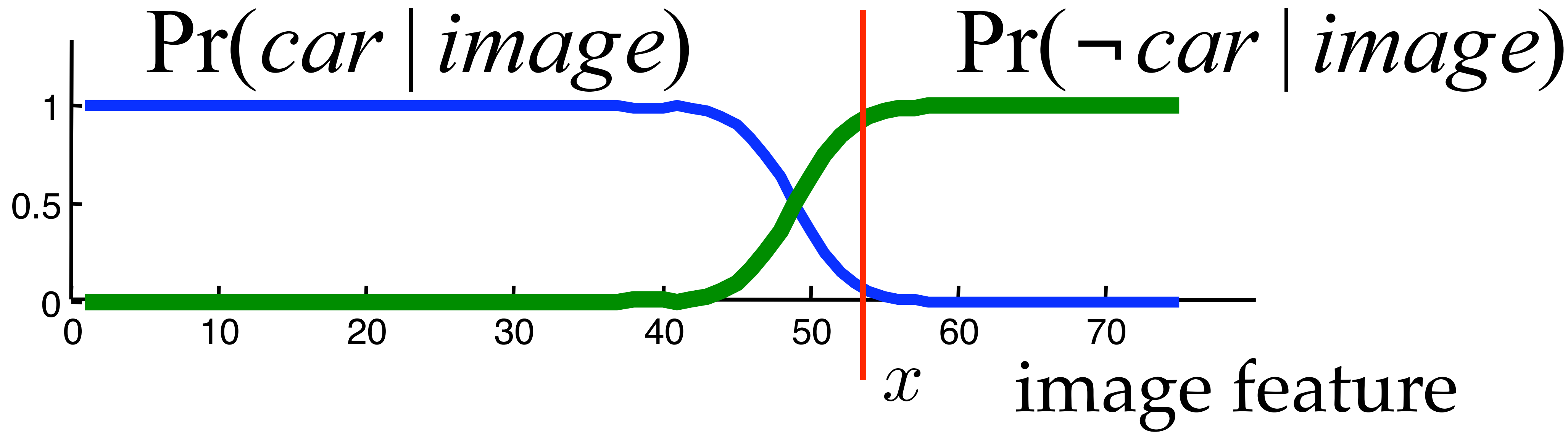
$$F(x, y) = P(x, y)$$



Generative

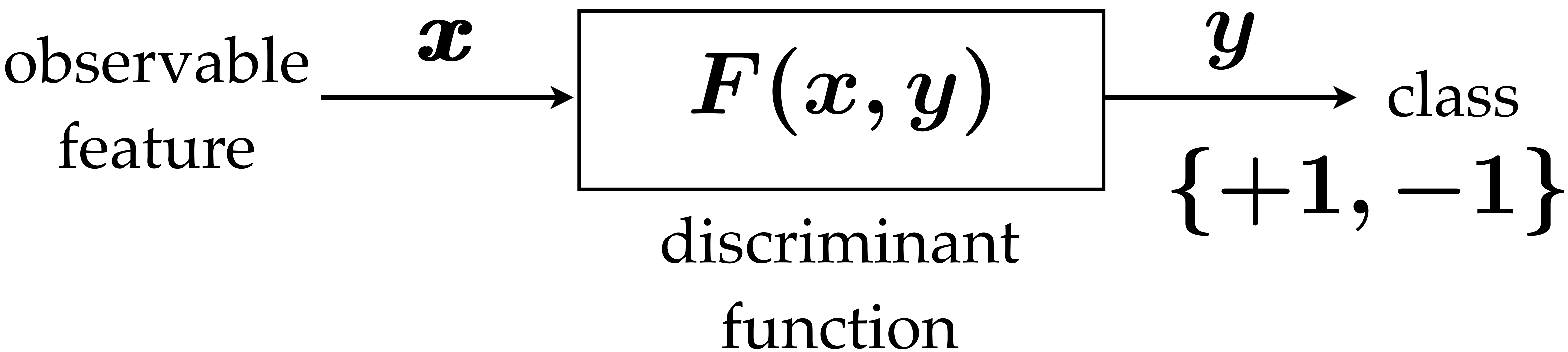
How to Classify?

$$F(x, y) = P(y|x)$$



Discriminative

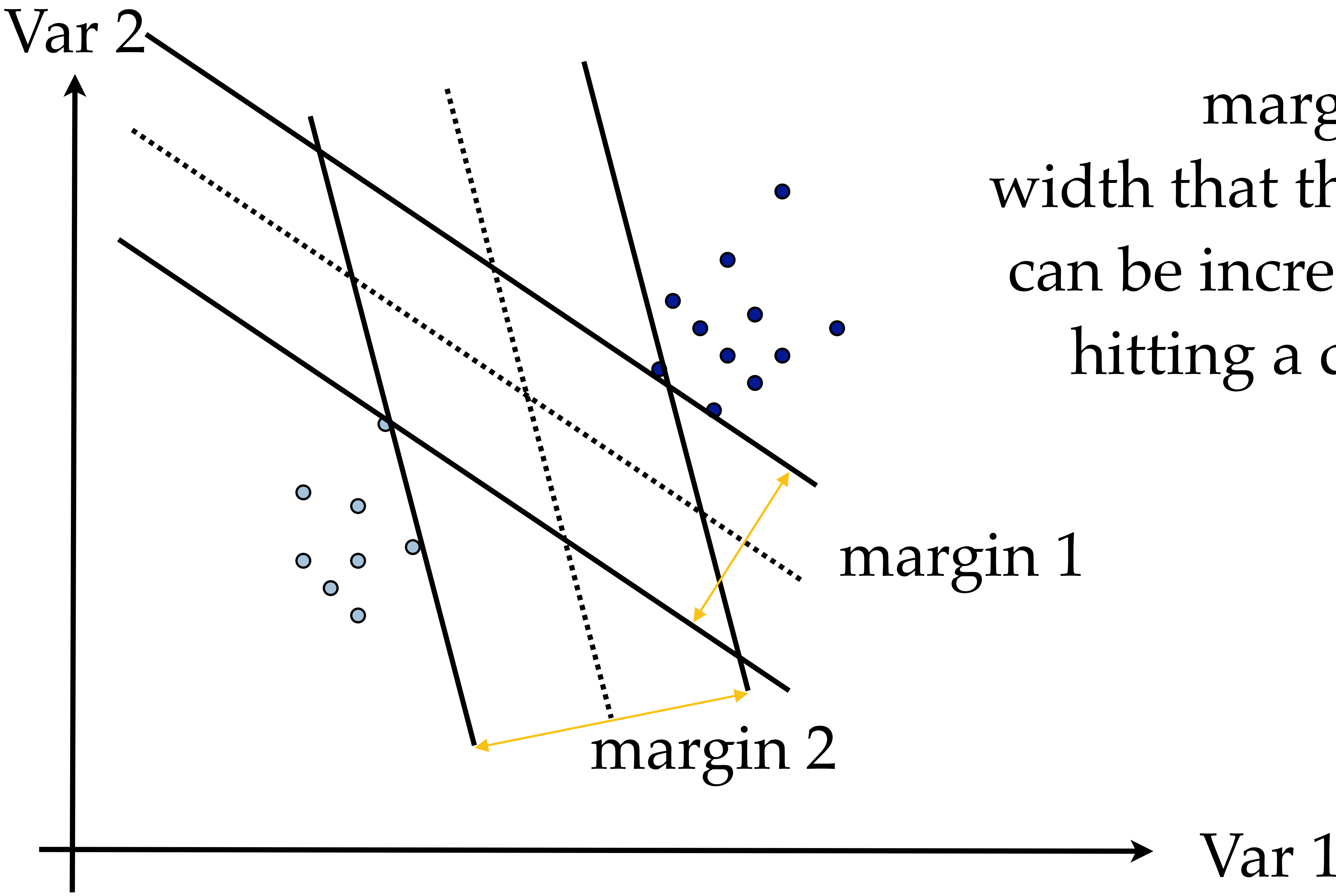
How to Classify?



$$\boldsymbol{F(x, y) = \mathbf{sign}(w.x + b)}$$

Linear discriminant function

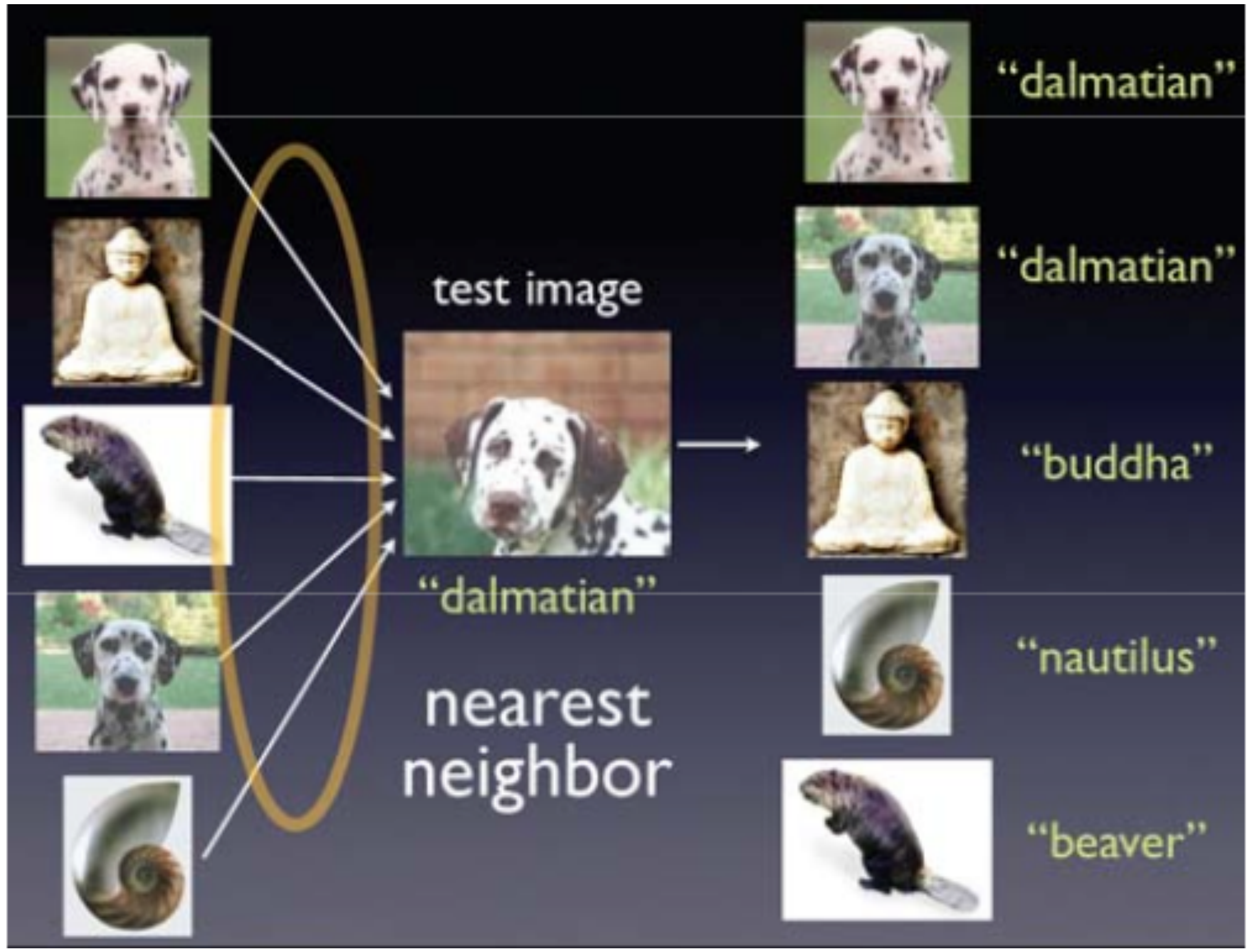
How to Classify?



margin =
width that the boundary
can be increased before
hitting a datapoint

Large margin classifiers

Distance Based Classifiers



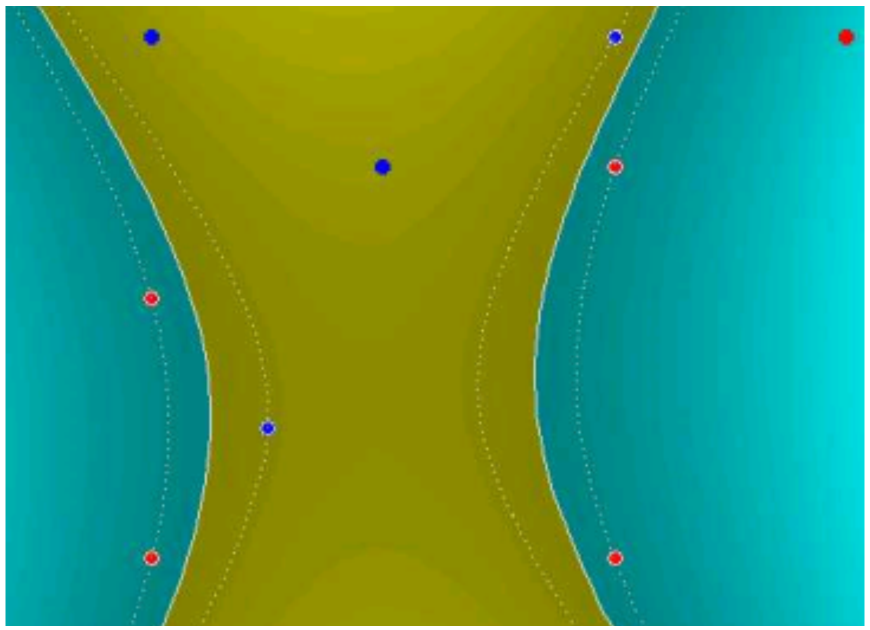
Given: $\{(x_1, y_1), ..., (x_n, y_n)\}$

query: x

$$y = \hat{y} \quad \text{s.t.} \quad \hat{x} = \arg\min_{x_i} D(x, x_i)$$

Outline

Support Vector Machines

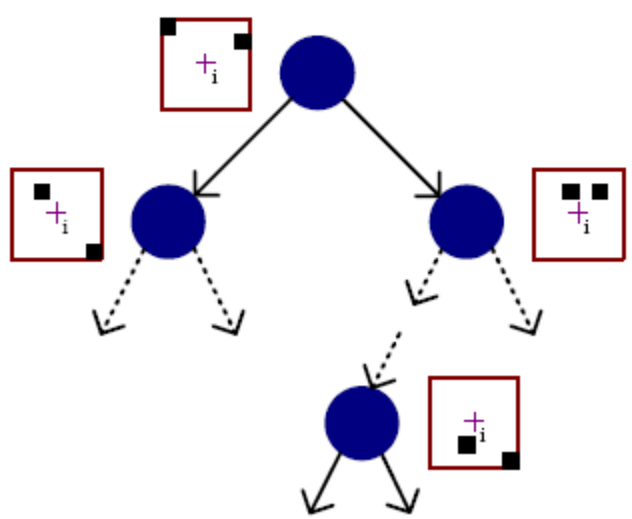


Guyon, Vapnik, Heisele,
Serre, Poggio, Berg...

Nearest Neighbor

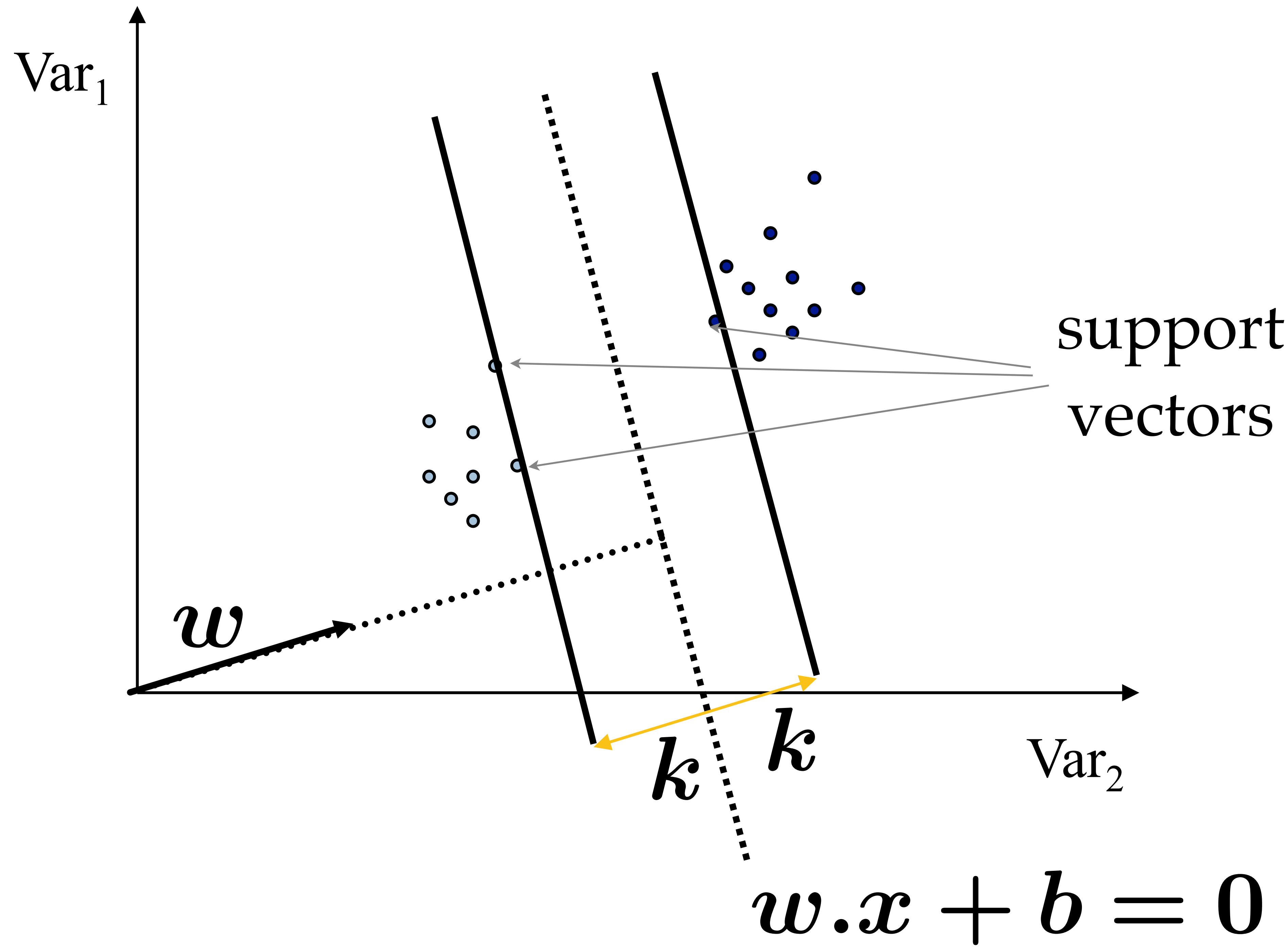
Shakhnarovich, Viola, Darrell,
Torralba, Efros, Berg, Frome,
Malik, Todorovic...

Random Forests

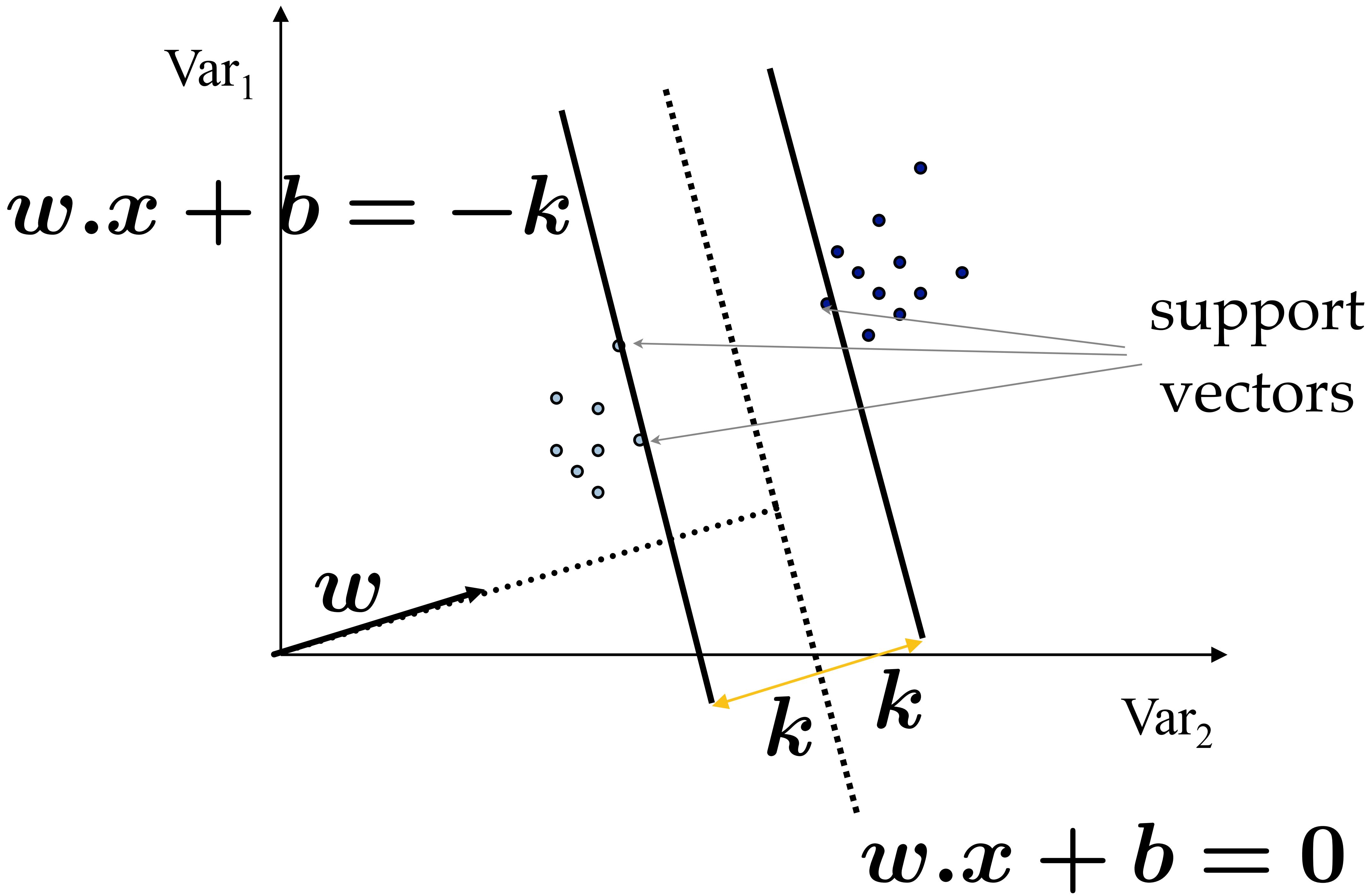


Breiman, Fua,
Criminisi, Cipolla,
Shotton, Lempitsky,
Zisserman, Bosch, ...

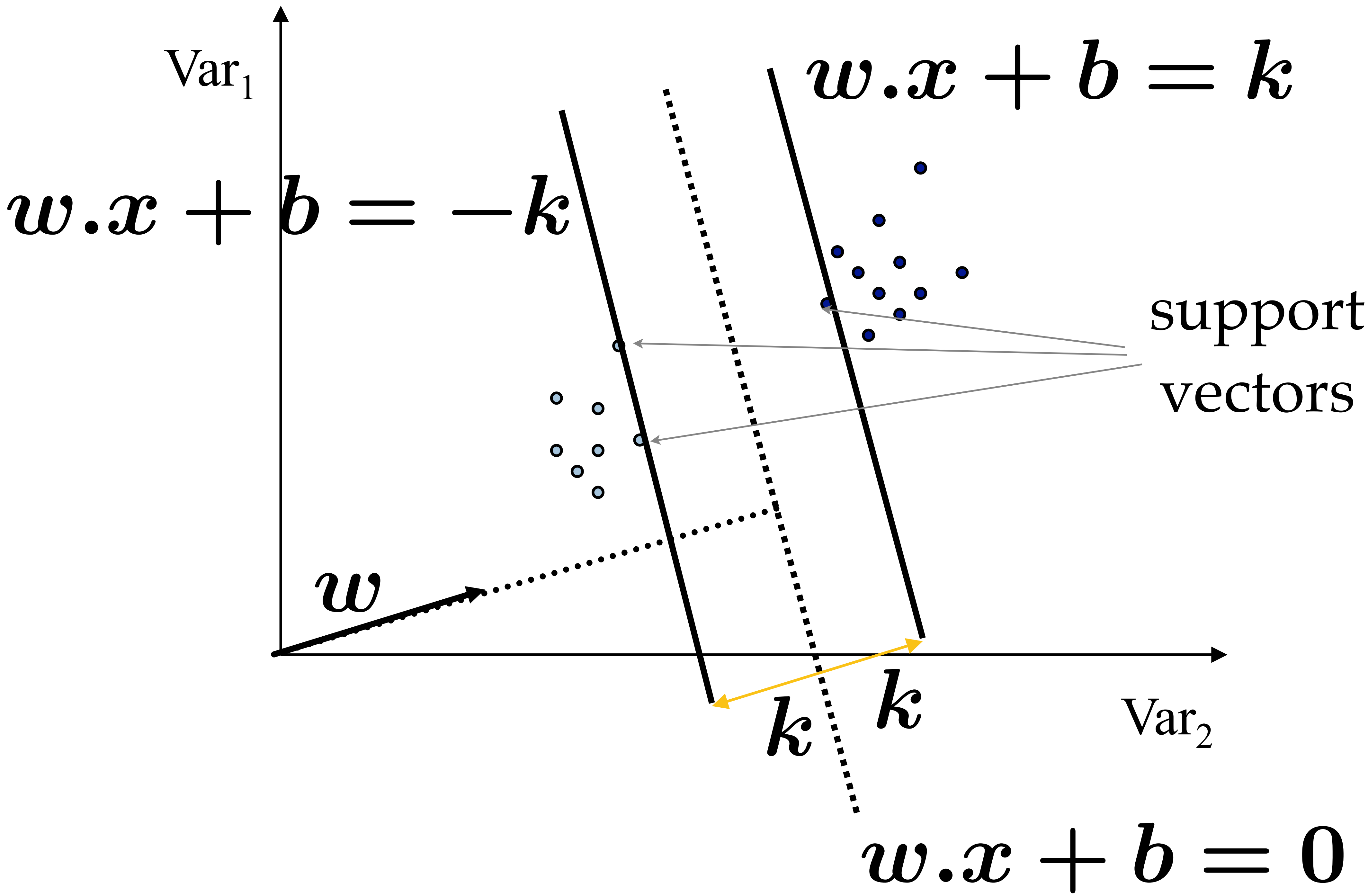
Maxim-Margin Linear Classifier



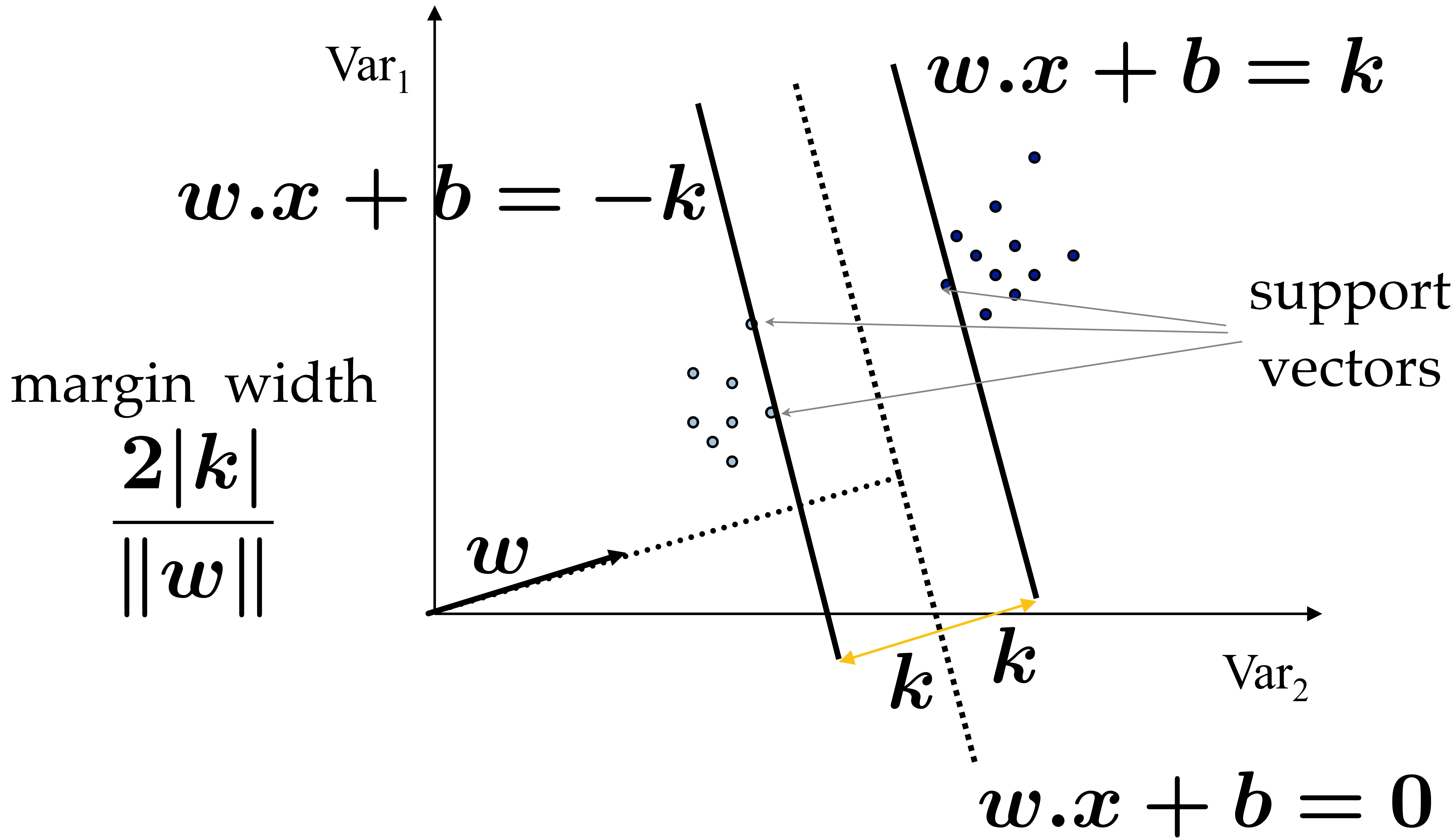
Maxim-Margin Linear Classifier



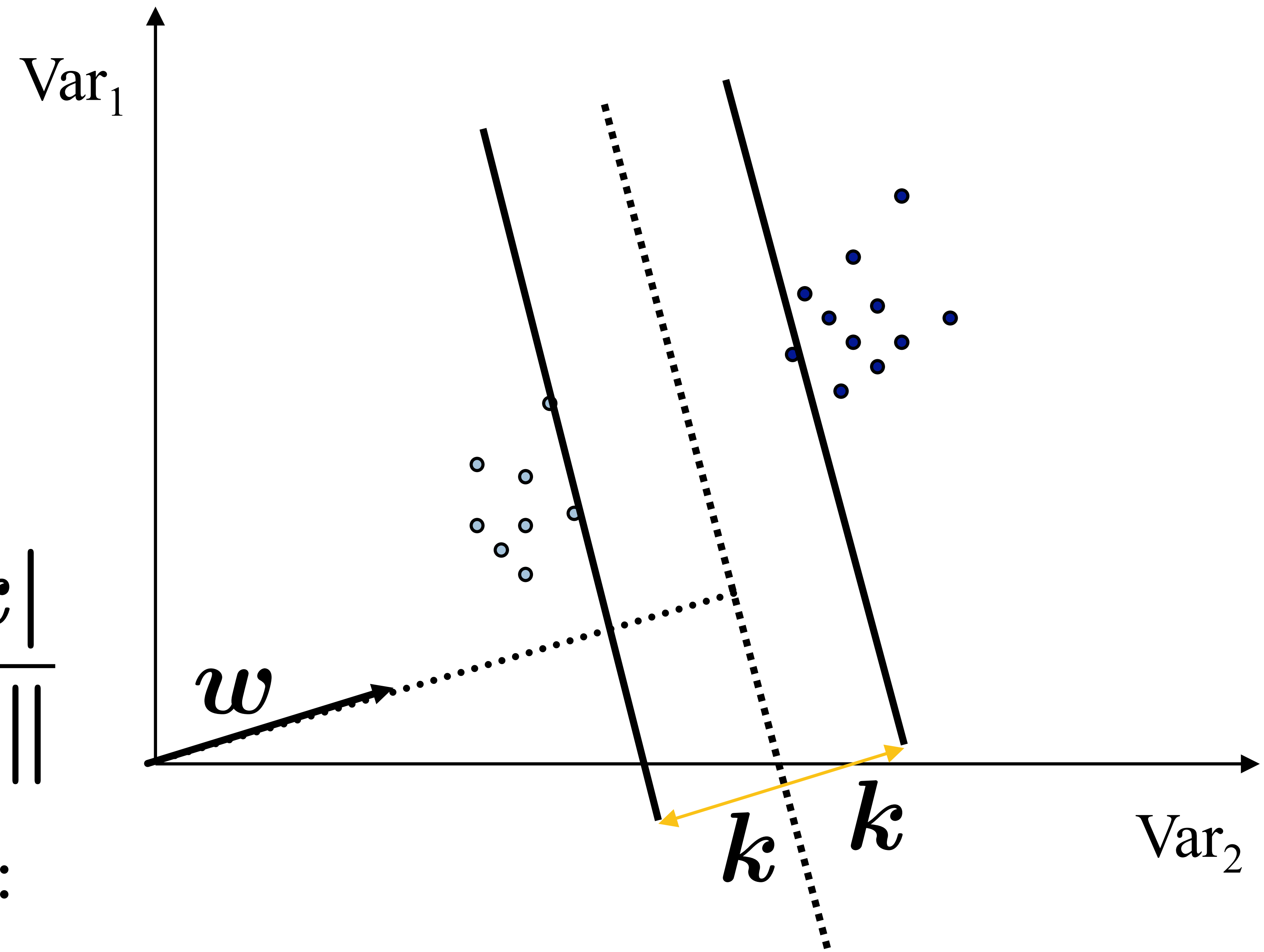
Maxim-Margin Linear Classifier



Maxim-Margin Linear Classifier



Maxim-Margin Linear Classifier



Problem:

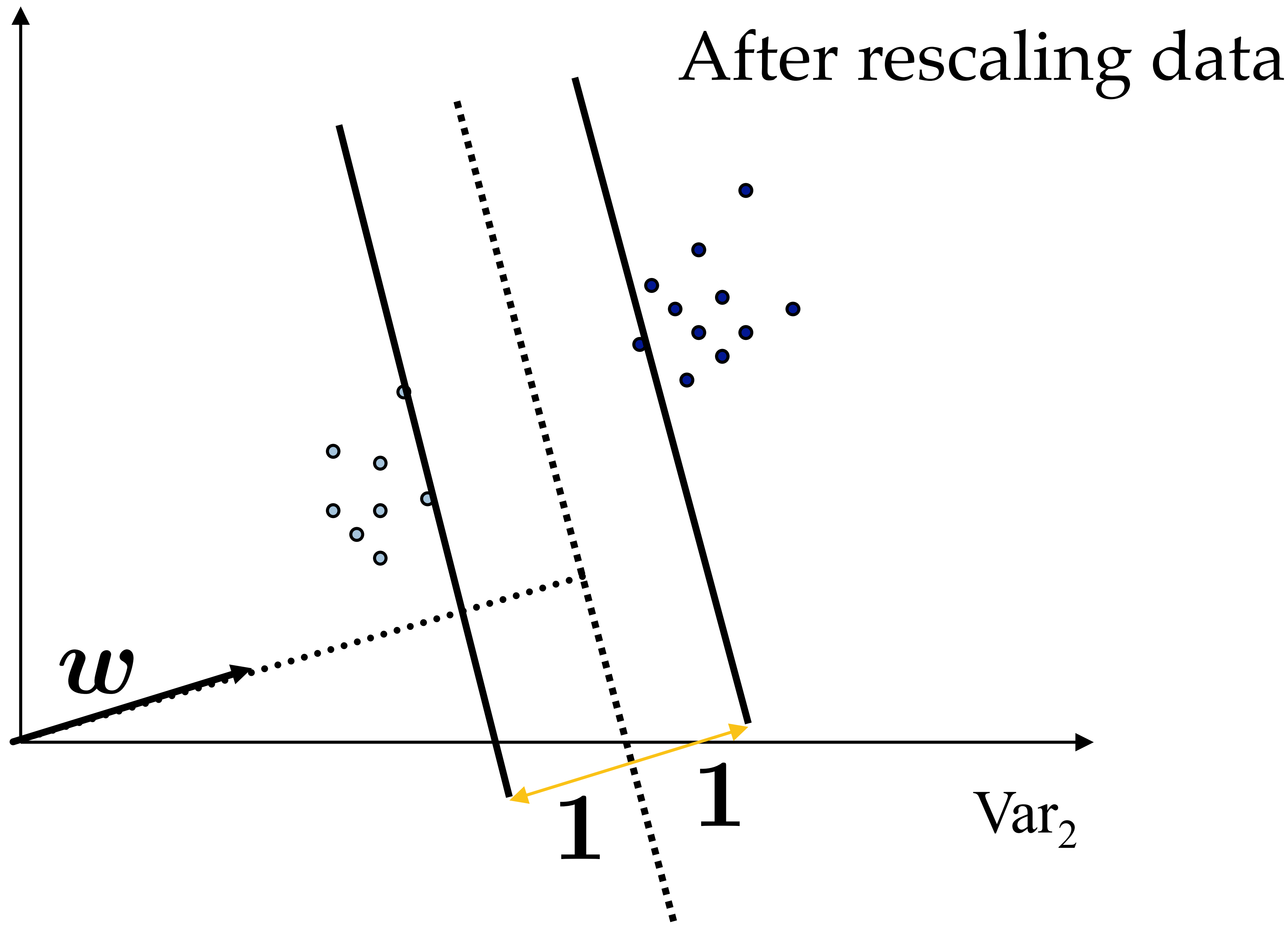
$$\max_{w} \frac{2|k|}{\|w\|}$$

subject to:

$$w.x + b \geq k, \ \forall x \text{ of class 1}$$

$$w.x + b \leq -k, \ \forall x \text{ of class -1}$$

Maxim-Margin Linear Classifier



Problem:

$$\max_{w} \frac{2}{\|w\|}$$

subject to:

$$w.x + b \geq 1, \ \forall x \text{ of class 1}$$

$$w.x + b \leq -1, \ \forall x \text{ of class -1}$$

LSVM Derivation

$$w.x + b \geq 1,\ \forall x \text{ of class 1}$$

$$w.x + b \leq -1,\ \forall x \text{ of class -1}$$

$$\Downarrow \qquad y = \pm 1$$

$$y(w.x + b) \geq 1,\ \forall x$$

LSVM Derivation

Problem:

$$\max_{\boldsymbol{w}} \frac{\mathbf{2}}{\|\boldsymbol{w}\|}$$

s.t.

$$\boldsymbol{y(w.x+b) \geq 1, \ \forall x}$$

LSVM Derivation

Problem:

$$\max_{w} \frac{2}{\|w\|}$$

s.t.

$$y(w.x + b) \geq 1, \; \forall x$$

⇩

$$\min_{w} \frac{1}{2}\|w\|^2$$

s.t.

$$y(w.x + b) \geq 1, \; \forall x$$

Dual Problem

Solve using Lagrangian

$$L = \frac{1}{2} w.w - \sum_i \alpha_i [y_i (w.x_i + b) - 1)]$$

Dual Problem

Solve using Lagrangian

$$L = \frac{1}{2} w.w - \sum_i \alpha_i [y_i (w.x_i + b) - 1)]$$

At solution $\frac{\partial L}{\partial b} = 0 \qquad \frac{\partial L}{\partial w} = 0$

$$L = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i x_j$$

Dual Problem

$$\max_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i x_j$$

s.t.

$$\sum_i y_i \alpha_i = 0, \ \alpha_i \geq 0$$

Dual Problem

$$\max_{\boldsymbol{\alpha}} \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i x_j$$

s.t.

$$\sum_i y_i \alpha_i = 0, \ \alpha_i \geq 0$$

Dual Problem

$$\max_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i x_j$$

s.t.

$$\sum_i y_i \alpha_i = 0, \;\; \alpha_i \geq 0$$

$$\Downarrow \; \hat{\alpha}_i$$

Then compute:

$$\hat{w} = \sum_i \hat{\alpha}_i y_i x_i$$

$\hat{\alpha}_i > 0$ Only for support vectors

$\hat{b} = y_i - \hat{w} x_i$ for support vectors

Linearly Non-Separable Case

trade-off between maximum separation and misclassification

$$\min_{\boldsymbol{w}} \frac{\mathbf{1}}{\mathbf{2}}\|\boldsymbol{w}\|^{\mathbf{2}} + C\sum_i \gamma_i$$

s.t.

$$\boldsymbol{y_i}(\boldsymbol{w}.\boldsymbol{x_i} + \boldsymbol{b}) \geq \mathbf{1} - \boldsymbol{\gamma_i},\ \forall \boldsymbol{x_i}$$

$$\boldsymbol{\gamma_i} \geq \mathbf{0}$$

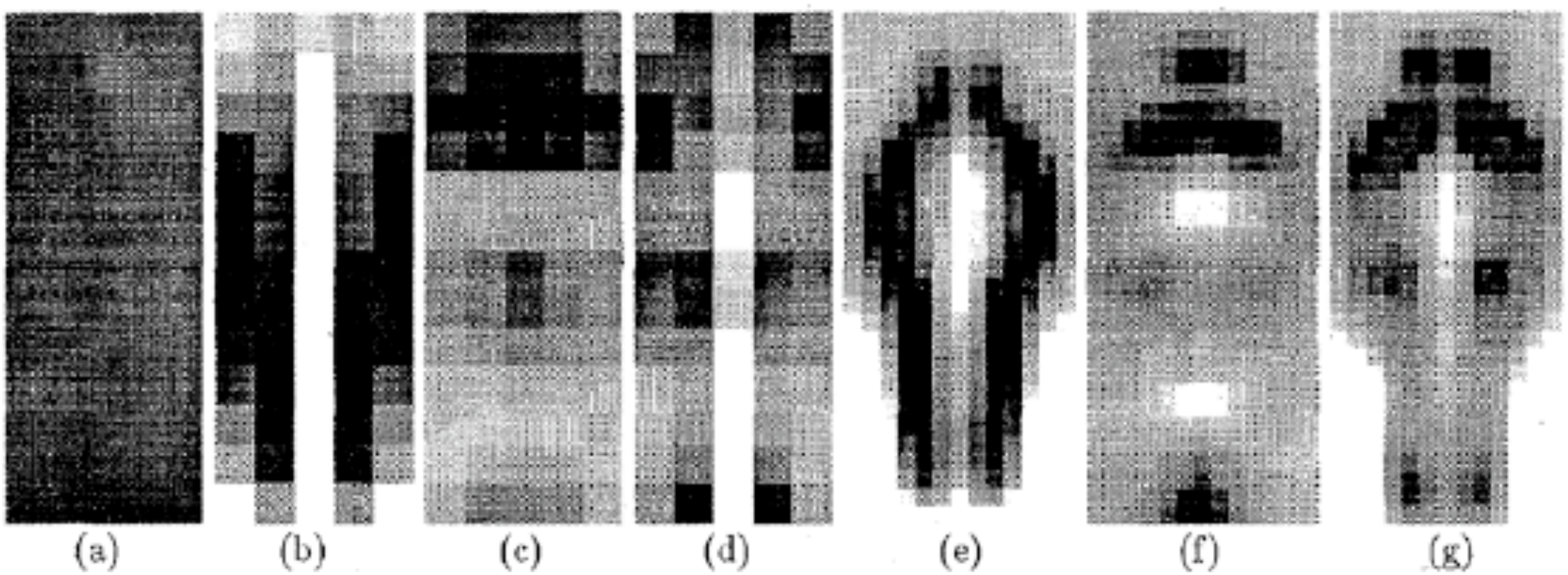
Non-Linear SVMs

- **Non-linear separation by mapping data to another space**
- **In SVM formulation, data appear only in the vector product**
- **No need to compute the vector product in the new space**
- **Mercer kernels**

$$x_i x_j \longrightarrow K(x_i, x_j) = \Phi(x_i)\Phi(x_j)$$

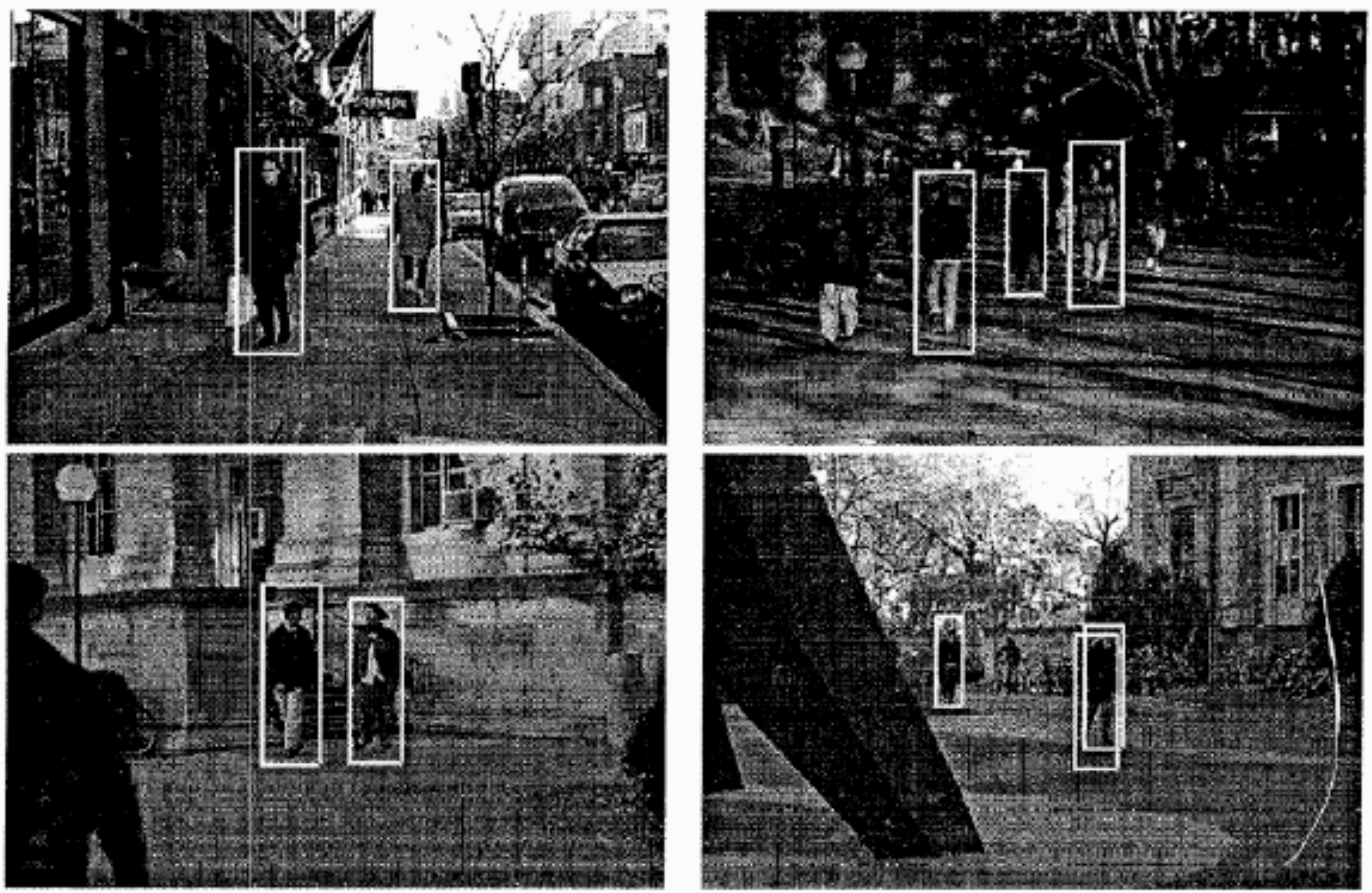
Vision Applications

- **Pedestrian detection**
  - **multiscale scanning windows**
  - **for each window compute the wavelet transform**
  - **classify the window using SVM**



"A general framework for object detection,"
C. Papageorgiou, M. Oren and T. Poggio -- CVPR 98

Vision Applications



"A general framework for object detection,"
C. Papageorgiou, M. Oren and T. Poggio -- CVPR 98

Shortcomings of SVMs

- **Kernelized SVM requires evaluating the kernel for a test vector and each of the support vectors**

- **Complexity = Kernel complexity × number of support vecto**

- **For a class of kernels this can be done more efficiently [Maji,Berg, Malik CVPR 08]**

$$K(x_i, x_j) = \sum_n \min(x_i(n), x_j(n))$$

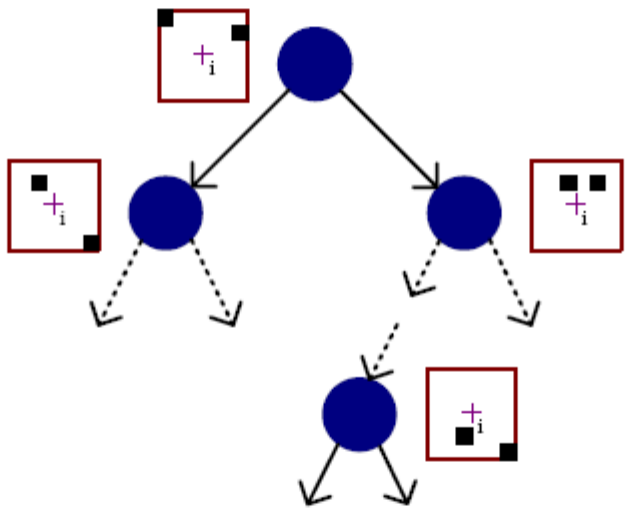
intersection kernel

Outline

Nearest Neighbor

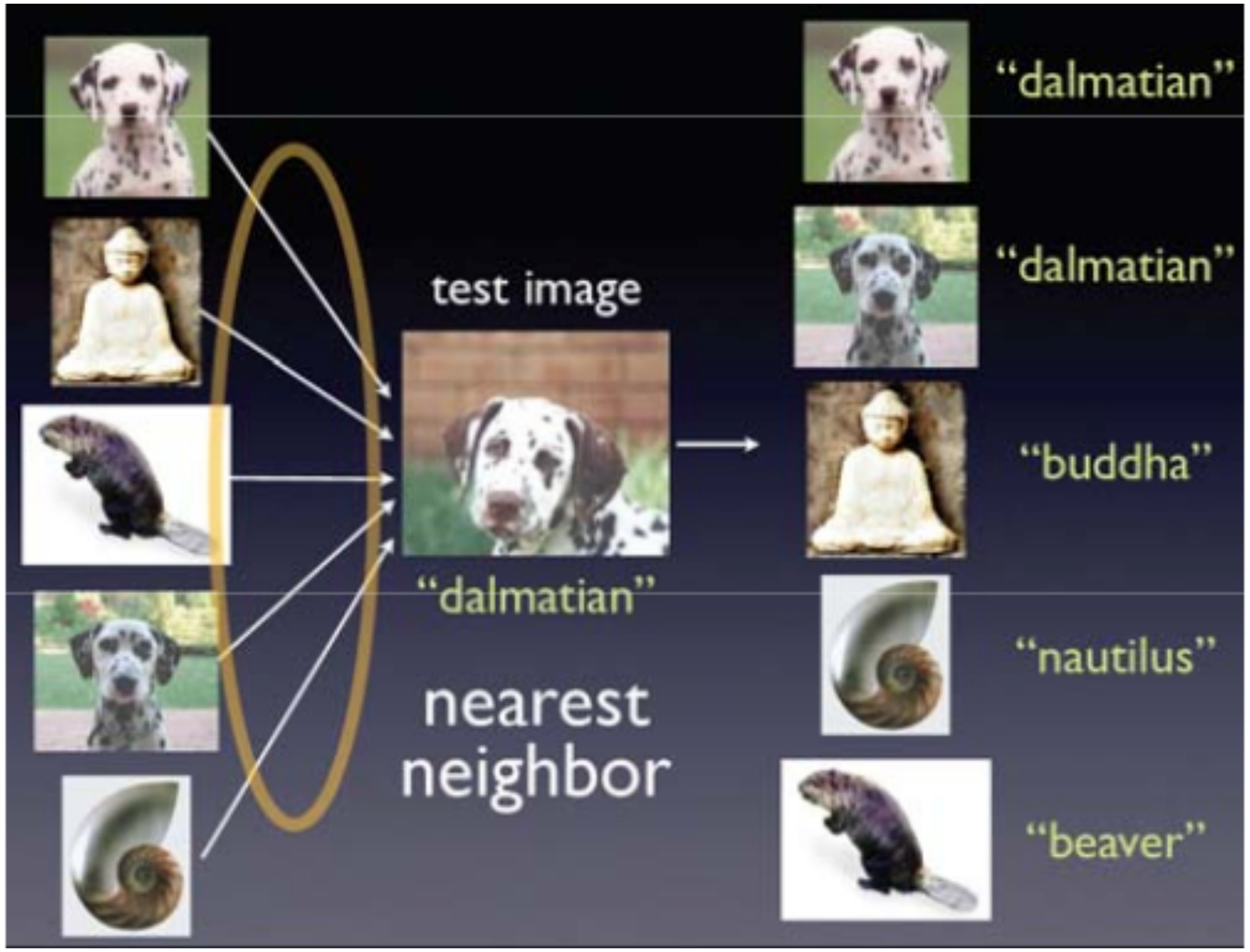
Shakhnarovich, Viola,
Darrell, Berg, Frome,
Malik, Todorovic...

Random Forests



Breiman, Fua,
Criminisi, Cipolla,
Shotton, Lempitsky,
Zisserman, Bosch, ...

Distance Based Classifiers

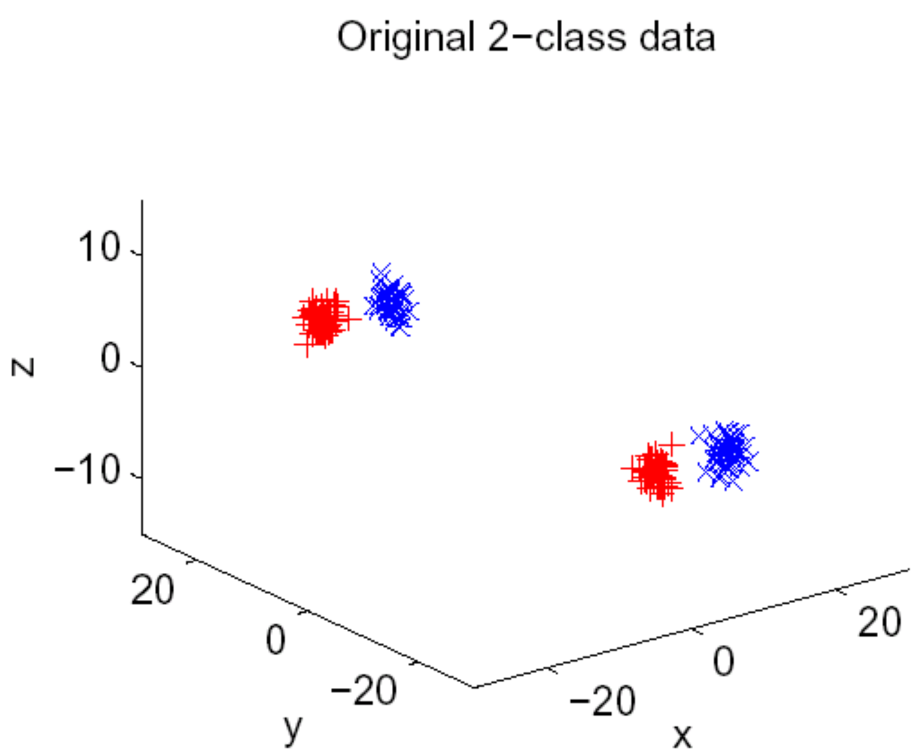


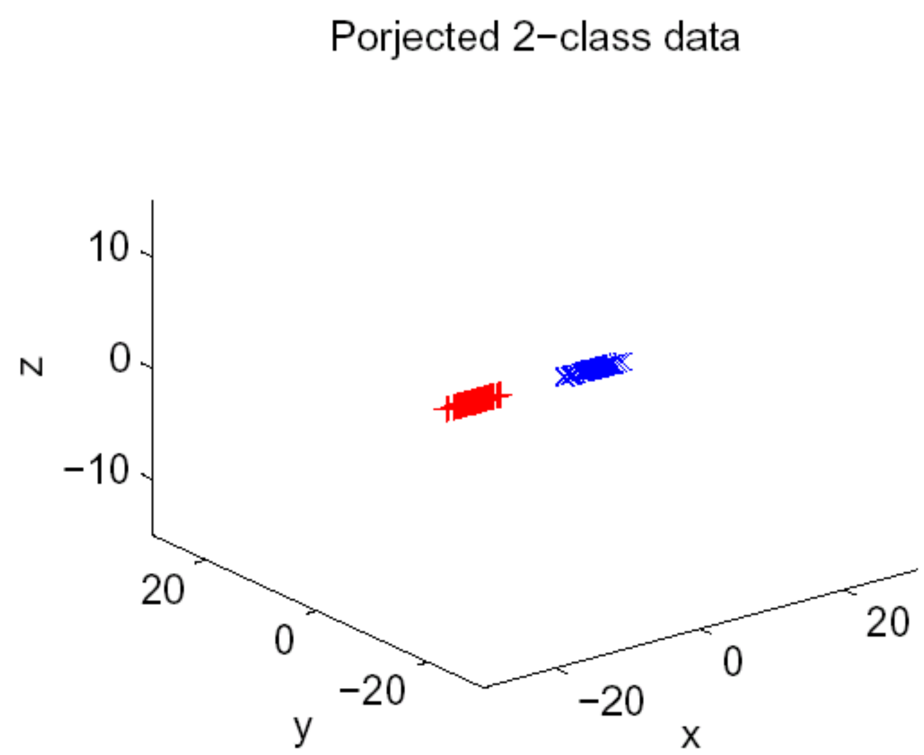
Given: $\{(x_1, y_1), ..., (x_n, y_n)\}$

query: x

$$y = \hat{y} \quad \text{s.t.} \quad \hat{x} = \arg\min_{x_i} D(x, x_i)$$

Learning Global Distance Metric





- **Given query x and datapoint-class pairs $\{(x_1, y_1), \ldots, (x_n, y_n)\}$**
- **Learn a Mahalanobis distance metric that**
  - **brings points from the same class closer, and**
  - **makes points from different classes be far away**

"Distance metric learning with application to clustering with side information"
E. Xing, A. Ng, and M. Jordan, NIPS, 2003.

Learning Global Distance Metric

$$\min_{A} \sum_{(x_i, x_j) \in S} \|x_i - x_j\|_A$$

s.t.

$$\sum_{(x_i, x_j) \in D} \|x_i - x_j\|_A \geq 1$$

Learning Global Distance Metric

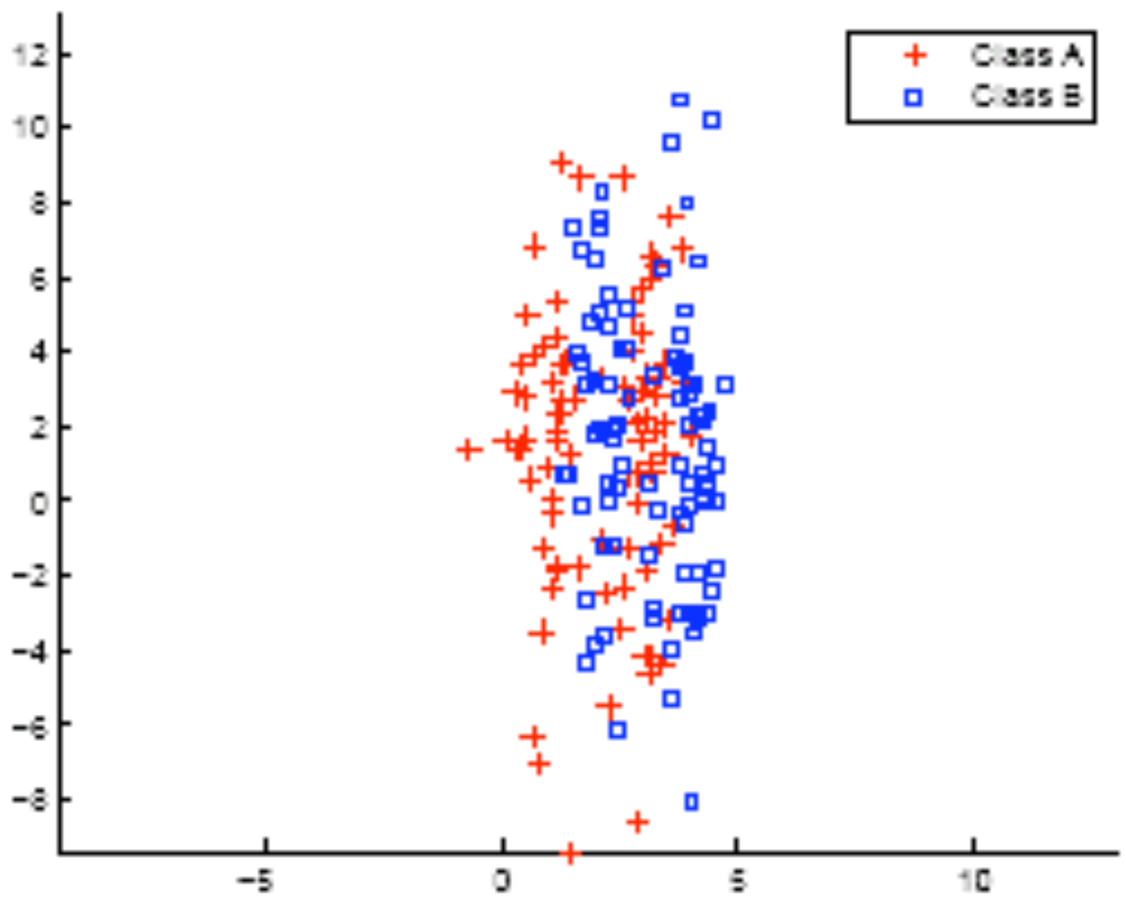
$$\min_{A} \sum_{(x_i, x_j) \in S} \|x_i - x_j\|_A$$

s.t.

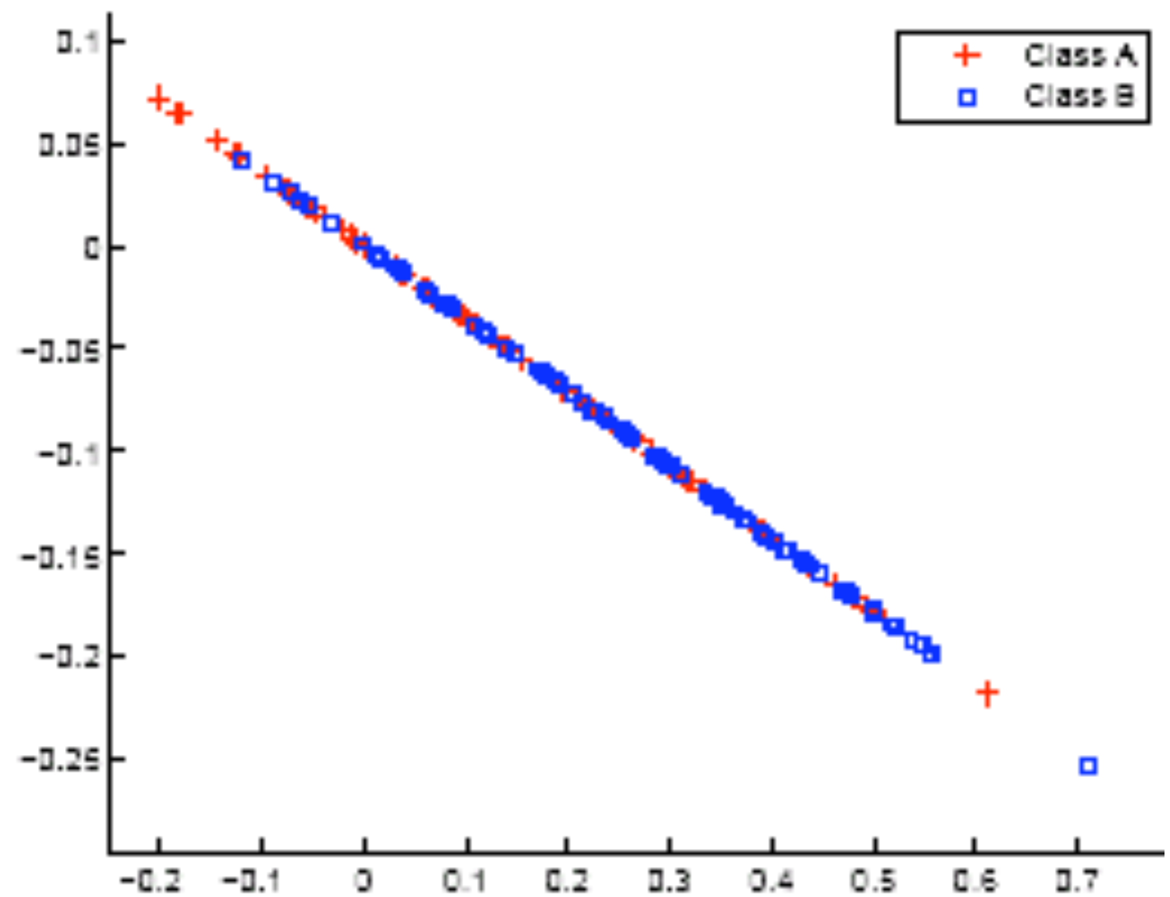
$$\sum_{(x_i, x_j) \in D} \|x_i - x_j\|_A \geq 1$$

PROBLEM!

Problem with multimodal classes

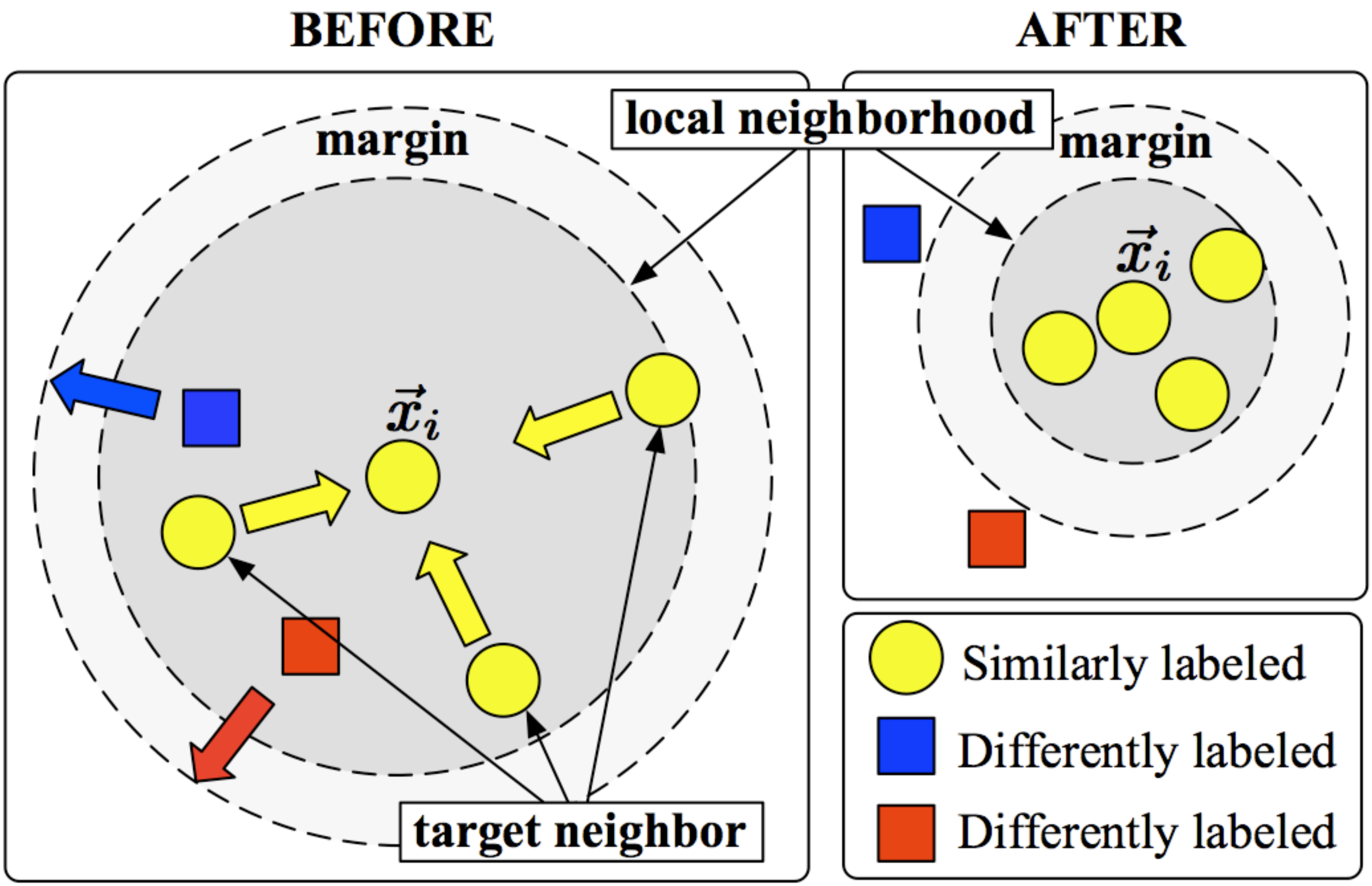


before learning



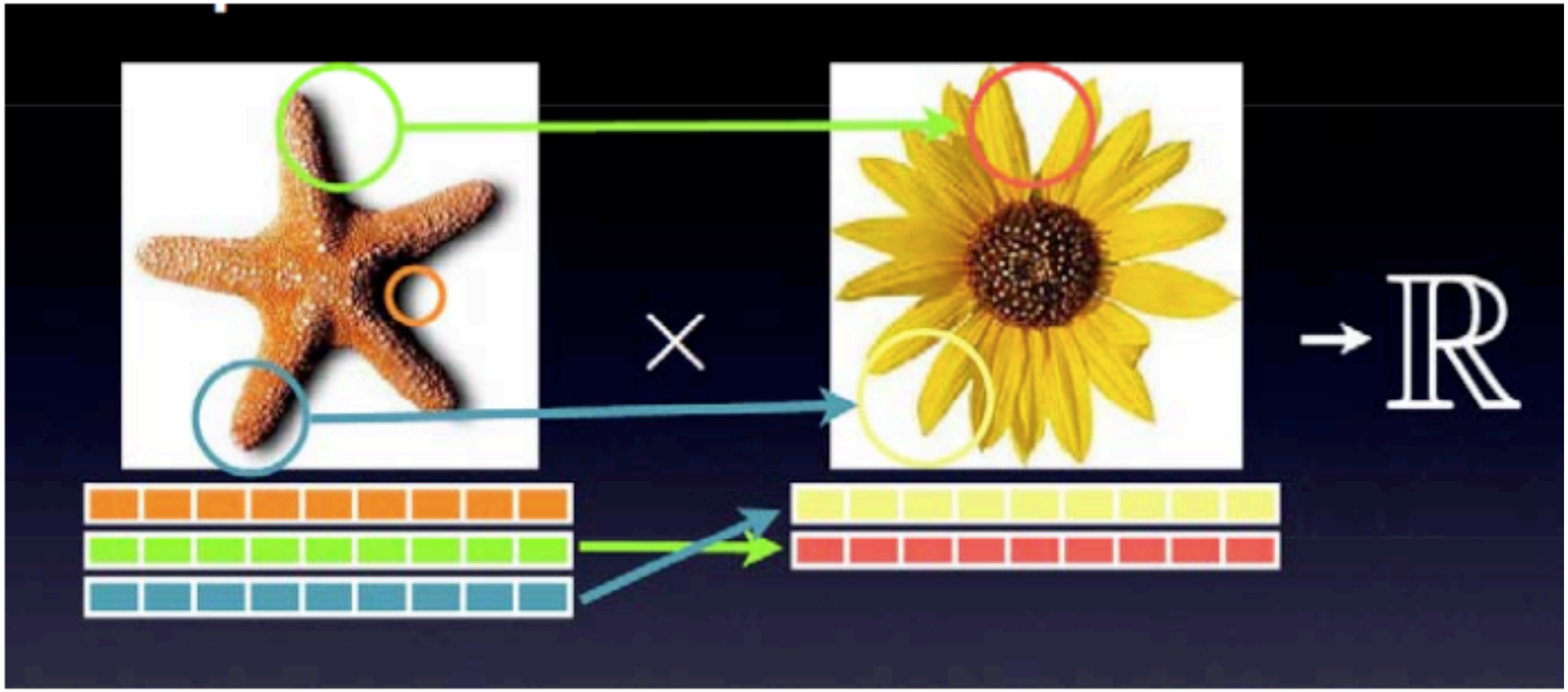
after learning

Per-Exemplar Distance Learning



Frome & Malik ICCV07, Todorovic & Ahuja CVPR08

Distance Between Two Images



$$D(F, I) = \sum_j w_j^F d_j^F(I) = w^F \cdot d^F(I)$$

distance between j-th patch in image F and image I

Learning from Triplets

For each image I in the set we have:

$$\boldsymbol{w}^{\boldsymbol{F}} \cdot \boldsymbol{d}^{\boldsymbol{F}}(\boldsymbol{I}^{\boldsymbol{D}}) > \boldsymbol{w}^{\boldsymbol{F}} \cdot \boldsymbol{d}^{\boldsymbol{F}}(\boldsymbol{I}^{\boldsymbol{S}})$$

Learning from Triplets

For each image I in the set we have:

$$\boldsymbol{w^F \cdot d^F(I^D) > w^F \cdot d^F(I^S)}$$

$$\Downarrow$$

$$\boldsymbol{x_i = d^F(I^D) - d^F(I^S)}$$

Learning from Triplets

For each image I in the set we have:

$$\boldsymbol{w}^{\boldsymbol{F}} \cdot \boldsymbol{d}^{\boldsymbol{F}}(\boldsymbol{I}^{\boldsymbol{D}}) > \boldsymbol{w}^{\boldsymbol{F}} \cdot \boldsymbol{d}^{\boldsymbol{F}}(\boldsymbol{I}^{\boldsymbol{S}})$$

$$\Downarrow$$

$$\boldsymbol{x_i} = \boldsymbol{d}^{\boldsymbol{F}}(\boldsymbol{I}^{\boldsymbol{D}}) - \boldsymbol{d}^{\boldsymbol{F}}(\boldsymbol{I}^{\boldsymbol{S}})$$

$$\Downarrow$$

$$\boldsymbol{w}^{\boldsymbol{F}} \cdot \boldsymbol{x_i} > \boldsymbol{0}$$

Max-Margin Formulation

- **Learn for each focal image F independently**

$$\arg\min_{w^F,\gamma} \frac{1}{2}\|w^F\|^2 + C\sum_i \gamma_i$$

s.t.

$$\forall i, \quad w^F \cdot x_i \geq 1 - \gamma_i$$

$$w^F \geq 0 \qquad \gamma_i \geq 0$$

Max-Margin Formulation

- **Learn for each focal image F independently**

$$\arg\min_{w^F,\gamma} \frac{1}{2}\|w^F\|^2 + C\sum_i \gamma_i$$

s.t.

$$\forall i, \quad w^F \cdot x_i \geq 1 - \gamma_i$$

$$w^F \geq 0 \qquad \gamma_i \geq 0$$

PROBLEM! $\forall i \quad \rightarrow \quad \textbf{only some images}$

They heuristically select 15 closest images

Max-Margin Formulation

$$D(F, I) = \sum_j w_j^F d_j^F(I) = \mathbf{w}^F \cdot \mathbf{d}^F(I)$$

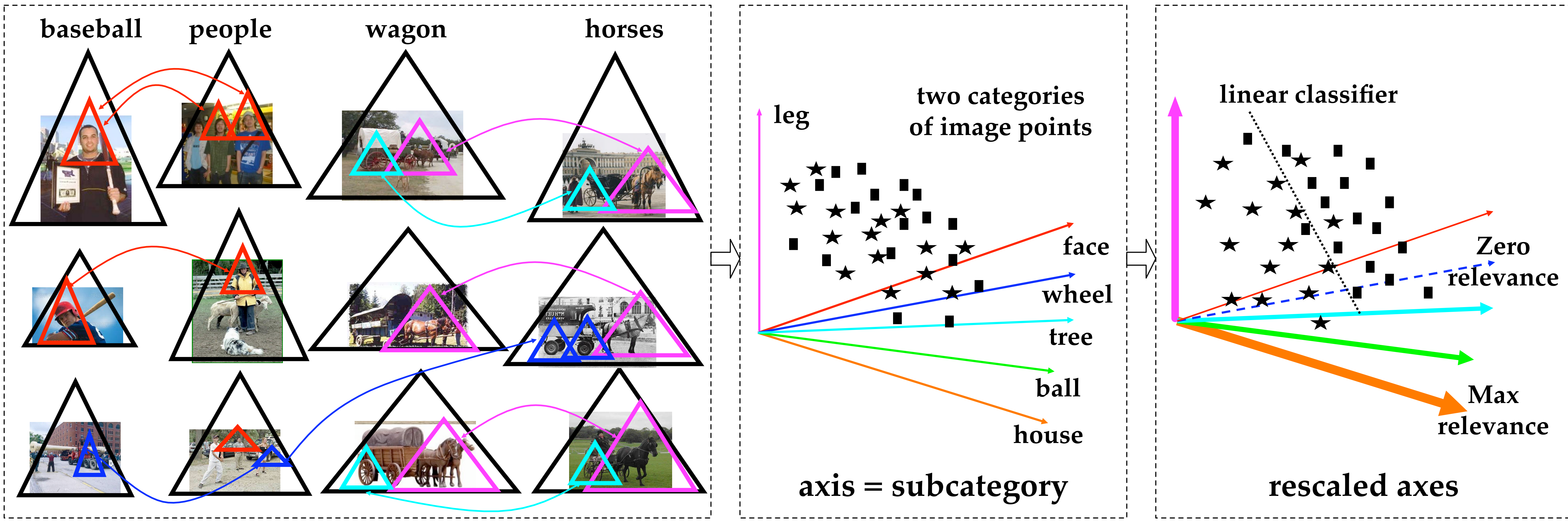
distance between j-th patch in image F and image I

Problem:

After computing $\mathbf{w}^F$

in-class and out-of-class datapoints that have initially been closest may not be closest after learning

EM-based Max-Margin Formulation of Local Distance Learning



discovery of subcategories in segmentation trees   trees = points in the feature space of subcategories

Todorovic&Ahuja CVPR08

Learning from Triplets

For each image I in the set:

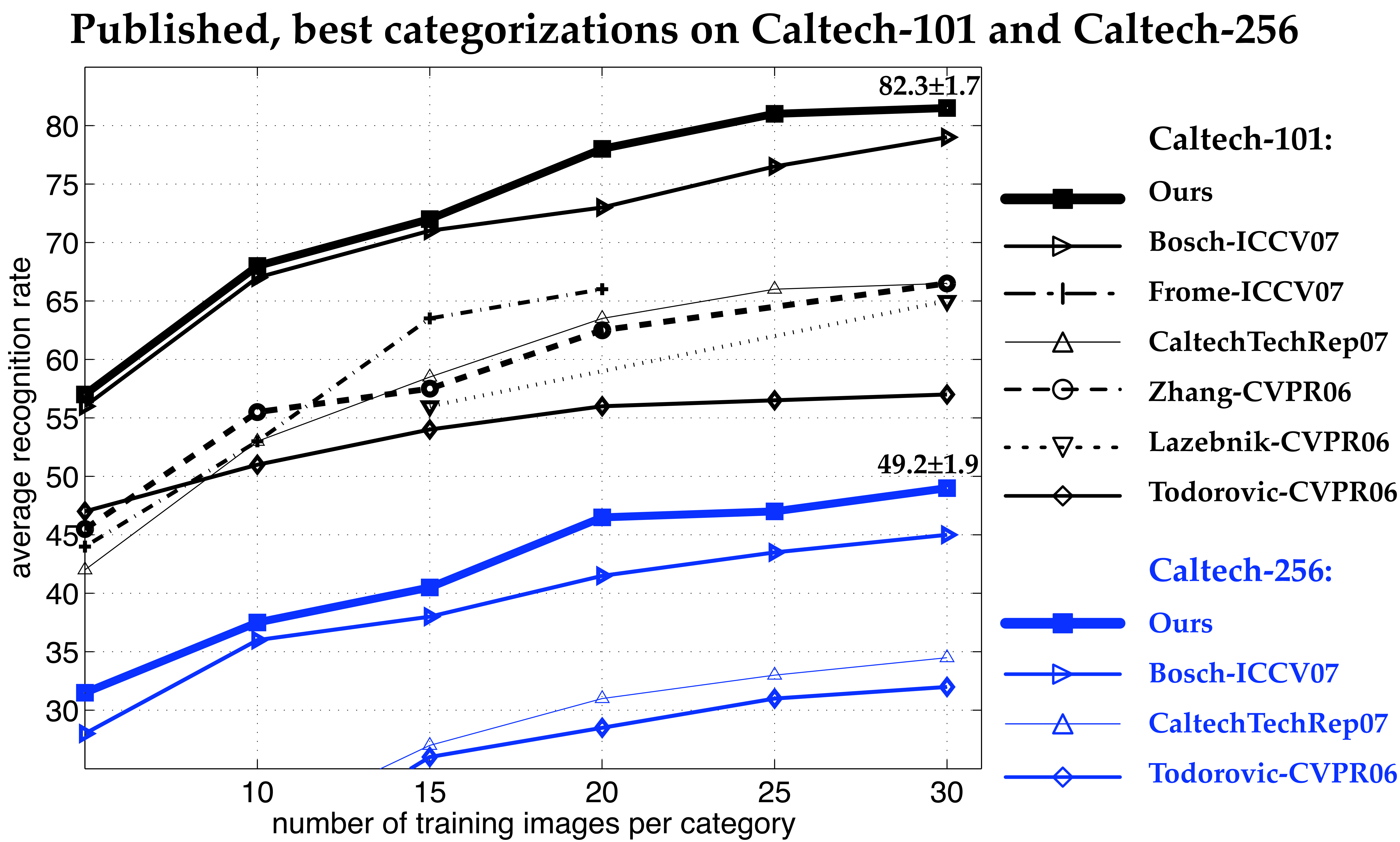
Frome, Malik ICCV07:

$$\boldsymbol{w^F \cdot d^F(I^D) > w^F \cdot d^F(I^S)}$$

Todorovic et al. CVPR08:

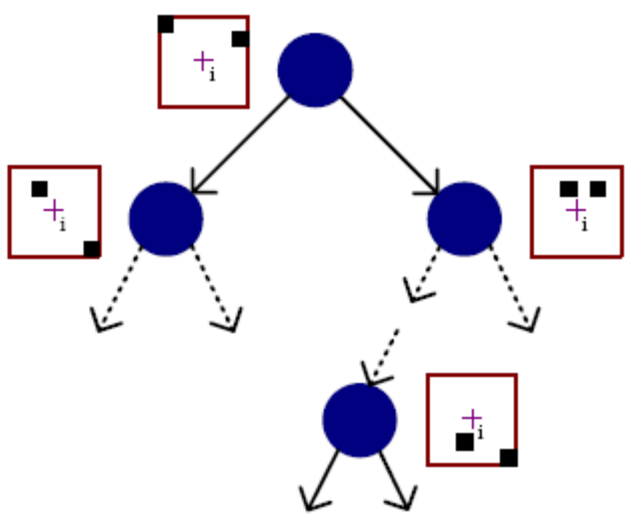
$$\boldsymbol{w^F \cdot d^F(I) Pr(I \in D) > w^F \cdot d^F(I) Pr(I \in S)}$$

CVPR 2008: Results on Caltech-256



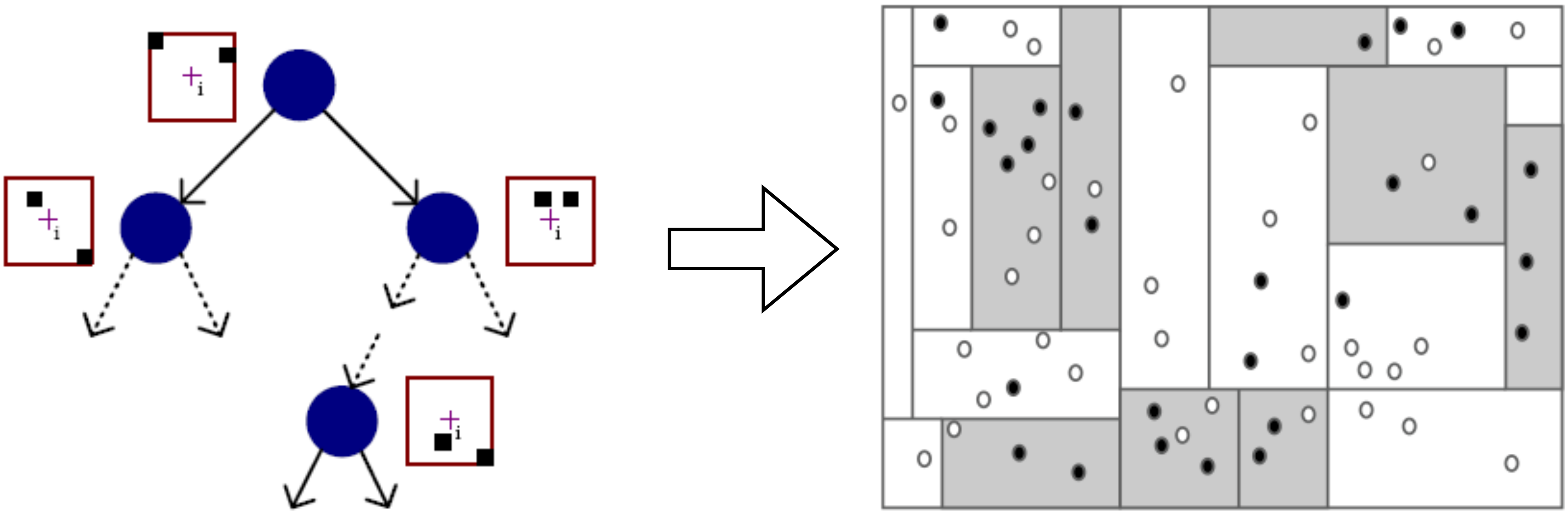
Outline

Random Forests



Breiman, Fua,
Criminisi, Cipolla,
Shotton, Lempitsky,
Zisserman, Bosch, ...

Decision Trees – Not Stable



- **Partitions of data via recursive splitting on a single feature**
- **Result: Histograms based on data-dependent partitioning**
- **Majority voting**
- **Quinlan C4.5; Breiman, Freedman, Olshen, Stone (1984); Devroye, Gyorfi, Lugosi (1996)**

Random Forests (RF)

RF = Set of decision trees such that
each tree depends on a random vector
sampled independently and
with the same distribution
for all trees in RF

Hough Forests

Combine:
spatial info + class info

"Class-Specific Hough Forests for Object Detection"
Juergen Gall and Victor Lempitsky
CVPR 2009

Binary Tests Selection

- Test with optimal split:

$$\underset{k}{\operatorname{argmin}} \left(U_\star(\{p_i \,|\, t^k(\mathcal{I}_i){=}0\}) + U_\star(\{p_i \,|\, t^k(\mathcal{I}_i){=}1\}) \right)$$

- Class-label uncertainty:

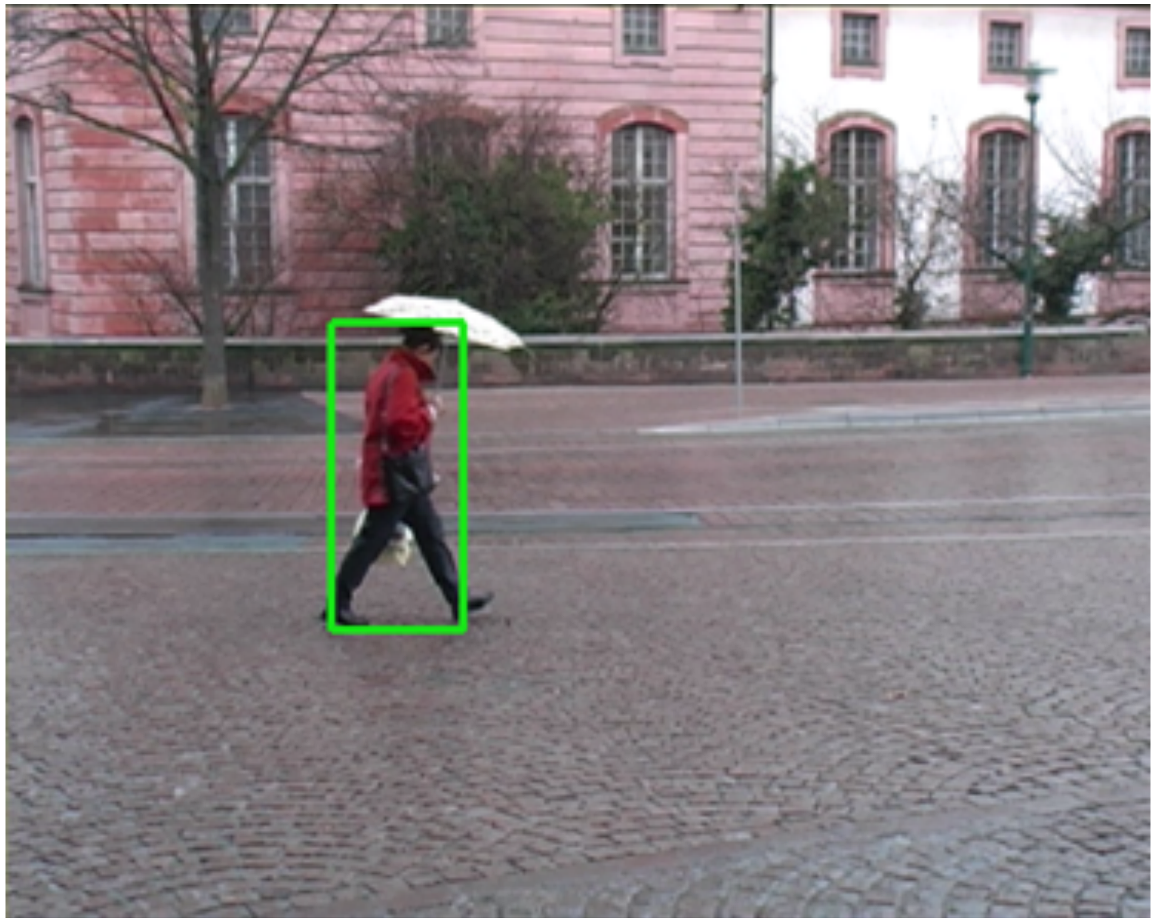
$$U_1(A) = |A| \cdot Entropy(\{c_i\})$$

- Offset uncertainty:

$$U_2(A) = \sum_{i:c_i=1} (\mathbf{d}_i - \mathbf{d}_A)^2$$

- Interleaved: Type of uncertainty is randomly selected for each node

Hough Forests

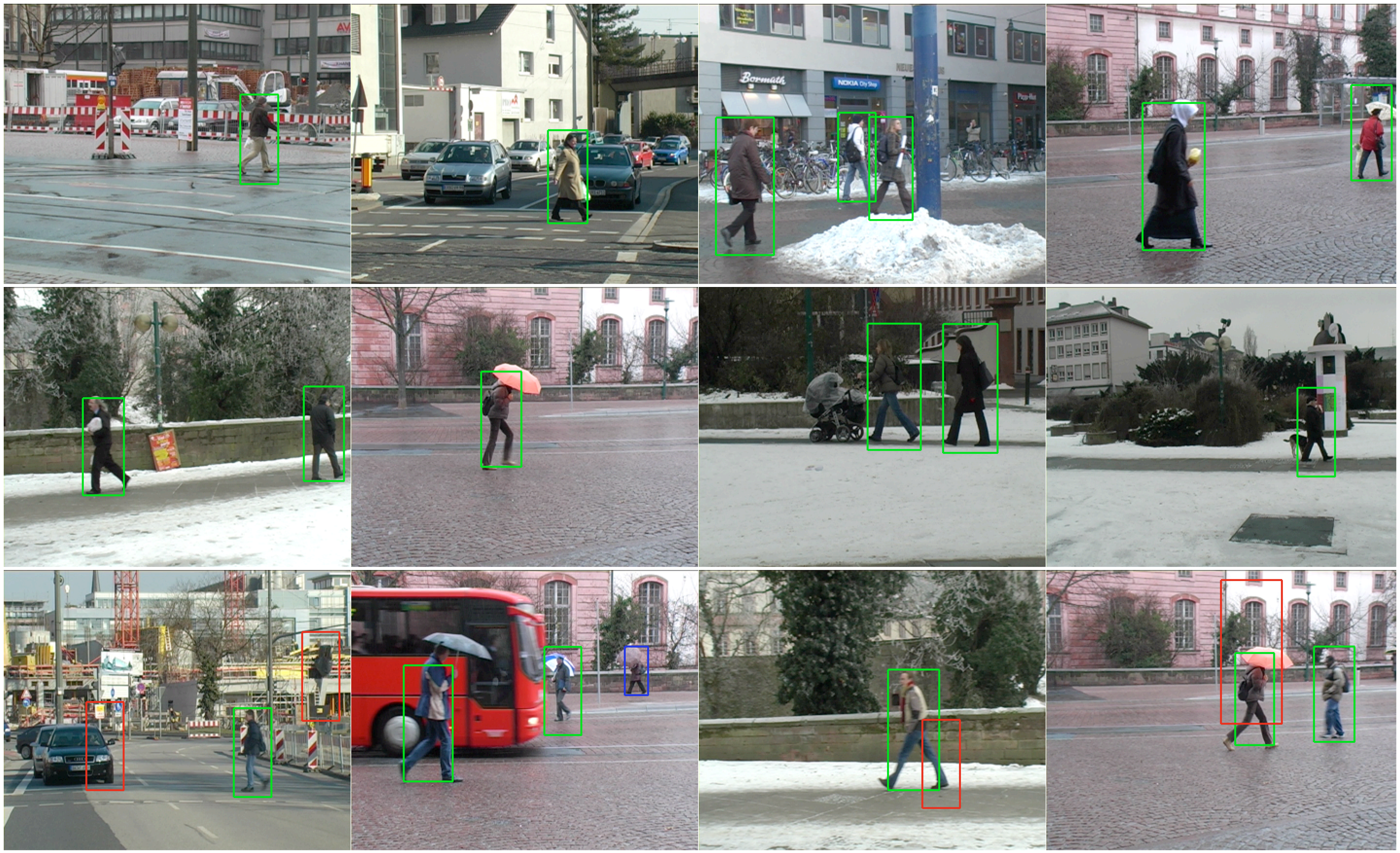


In the test image all features cast votes about the location of the bounding box

"Class-Specific Hough Forests for Object Detection"
Juergen Gall and Victor Lempitsky
CVPR 2009

Hough Forests

Pedestrians (TUD)



"Class-Specific Hough Forests for Object Detection"
Juergen Gall and Victor Lempitsky
CVPR 2009

Thank you!