

Object Recognition

Object Recognition

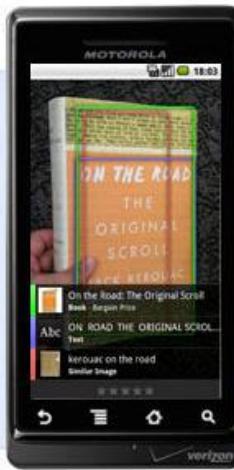
Introduction

Google Goggles



Google Goggles

Use pictures to search the web. [▶ Watch a video](#)



Get Google Goggles

Android (1.6+ required)

Download from Android Market.

[Send Goggles to Android phone](#)

New! iPhone (iOS 4.0 required)



Text



Landmarks



Books



Contact Info



Artwork



Wine



Logos



Slide credit: Kristen Grauman

Visual Recognition

- ▶ Design algorithms that have the capability to:
 - Classify images or videos
 - Detect and localize objects
 - Estimate semantic and geometrical attributes
 - Classify human activities and events

Does this image contain a building?

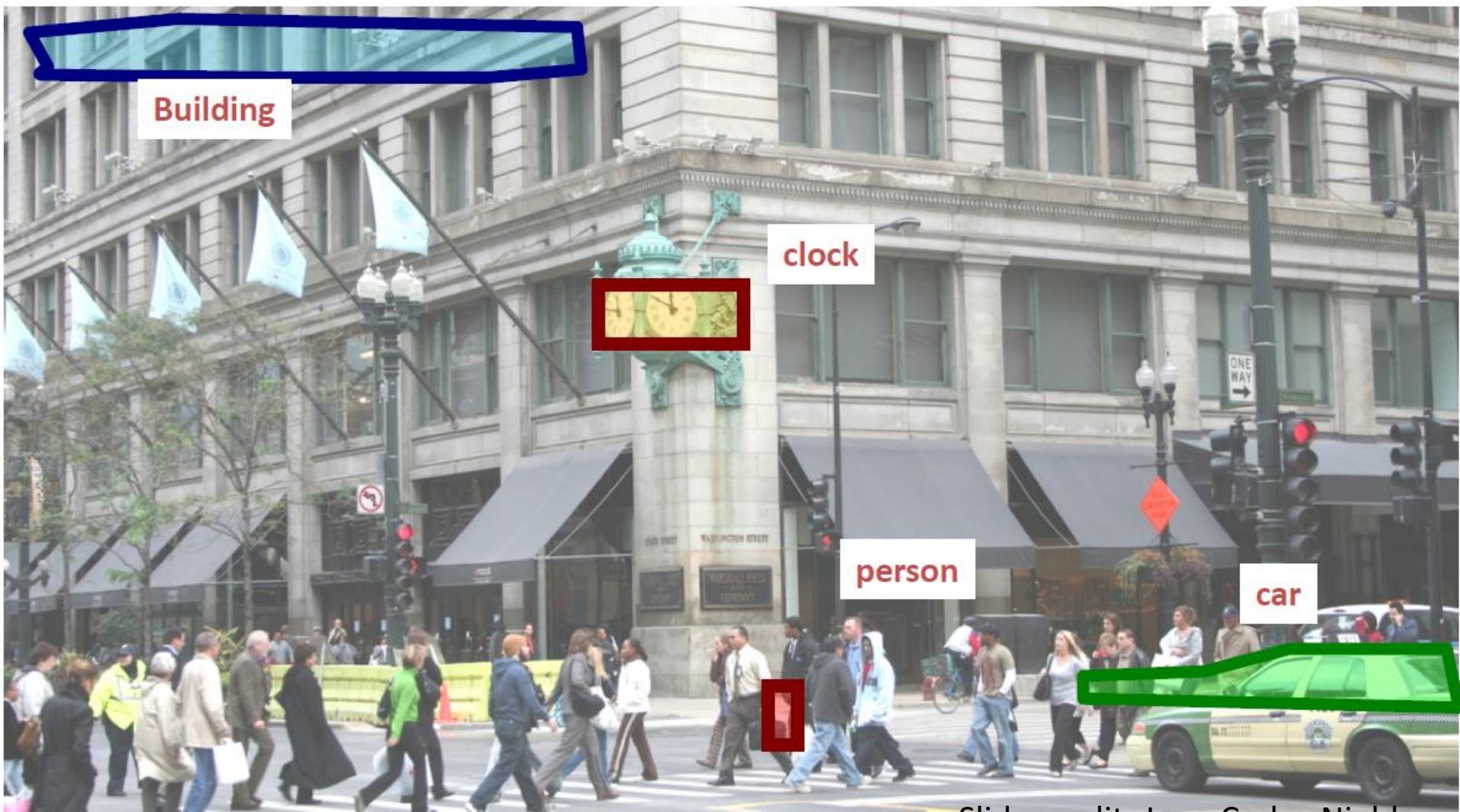


Slide credit: Juan Carlos Niebles

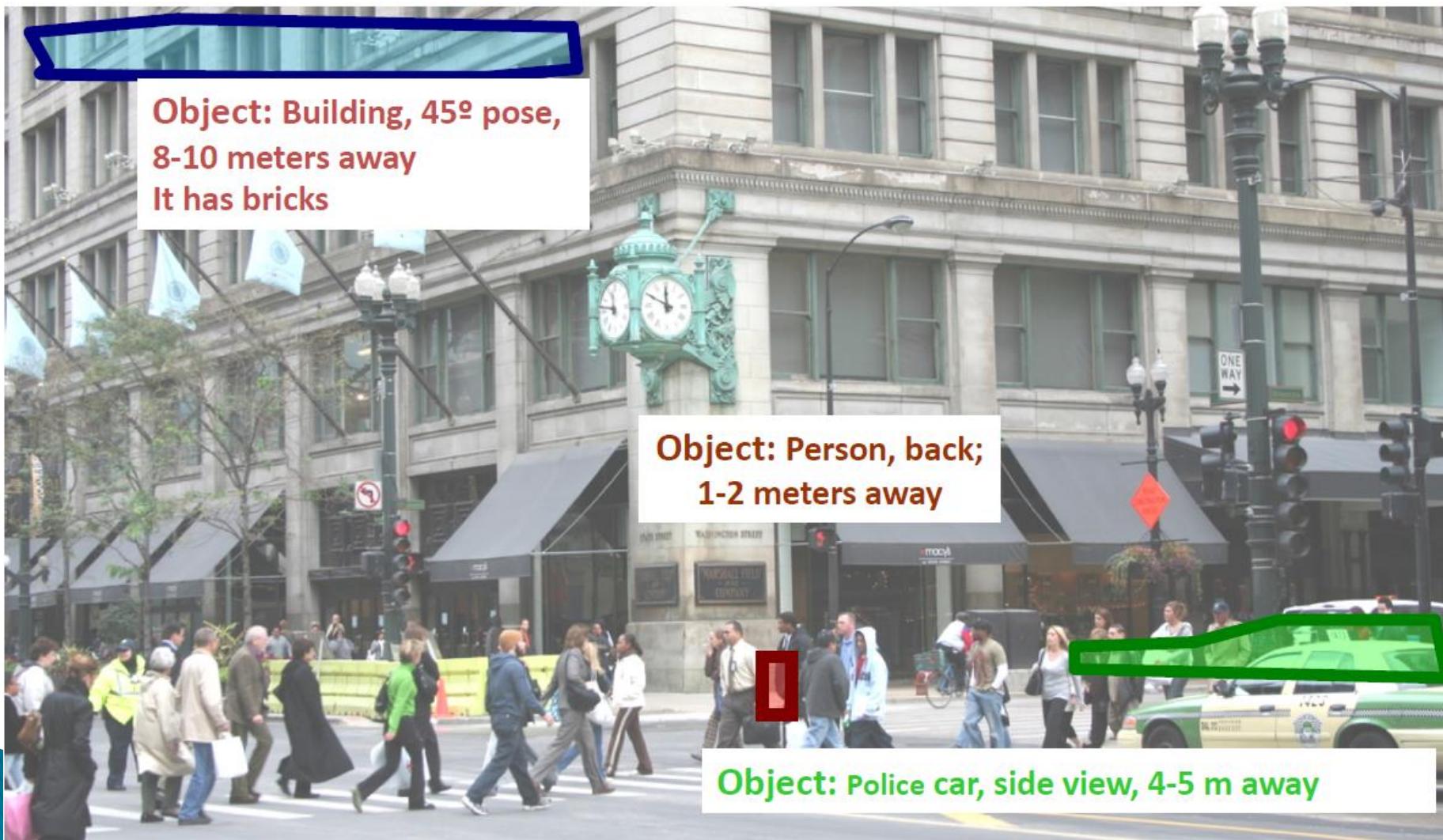
Does this image contain a car?



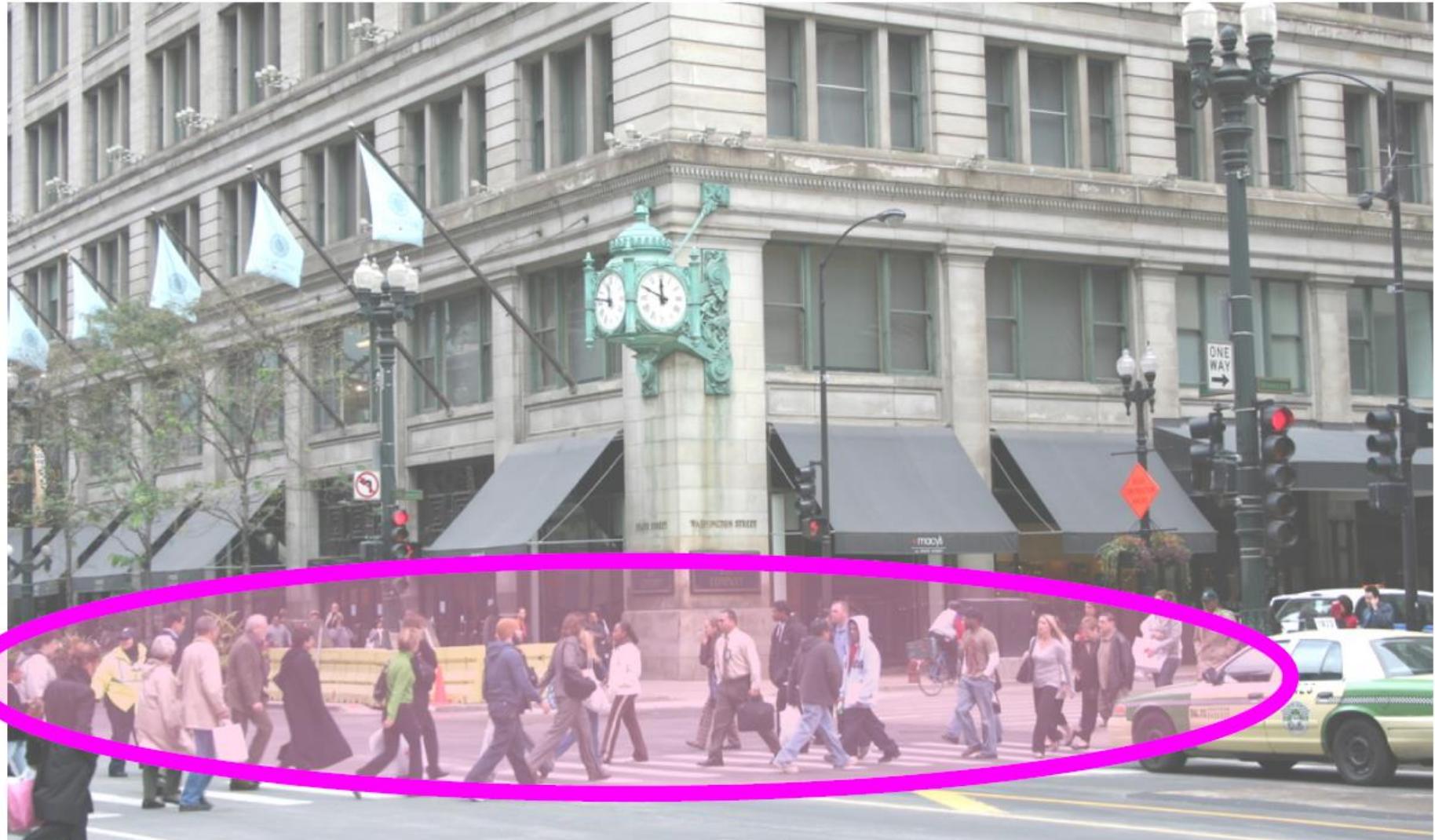
Which object does this image contain?



Estimating object semantic & geometric attributes



What are these people doing?



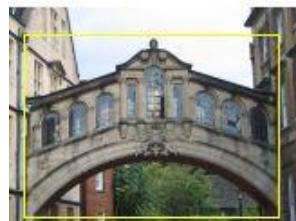
Slide credit: Juan Carlos Niebles

How many object categories are there?

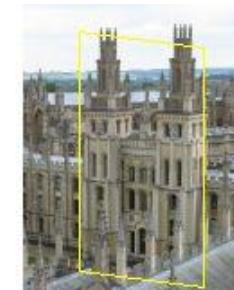
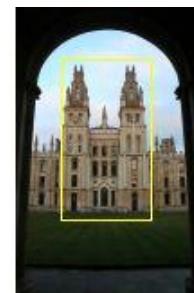


~10,000 to 30,000

Why is it difficult?



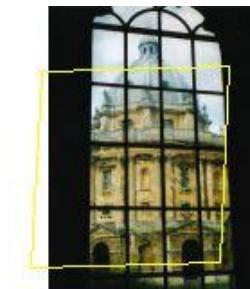
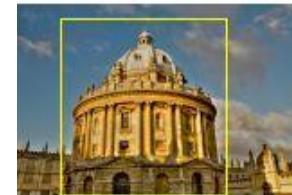
Scale



Viewpoint



Lighting



Occlusion

Challenges: viewpoint variation



Michelangelo 1475-1564



Scale



Slide credit: Juan Carlos Niebles

illumination

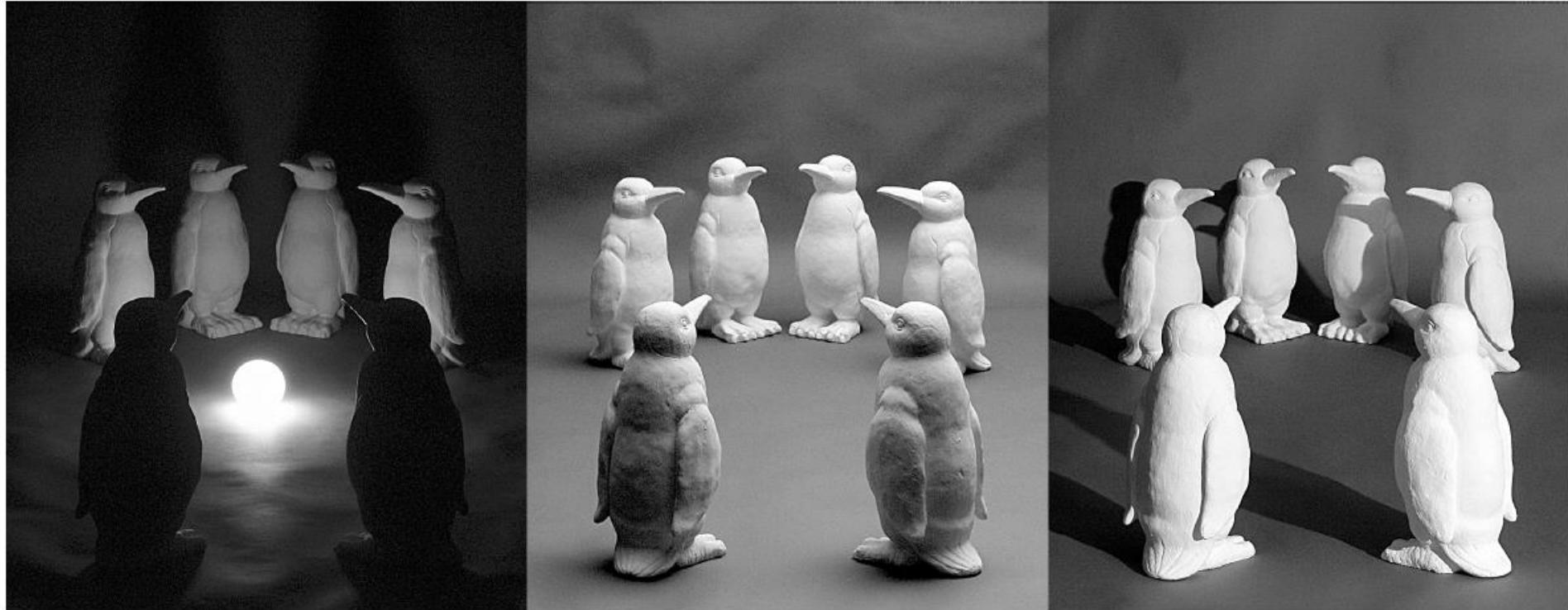


image credit: J. Koenderink

Deformation



Slide credit: Juan Carlos Niebles

Occlusion

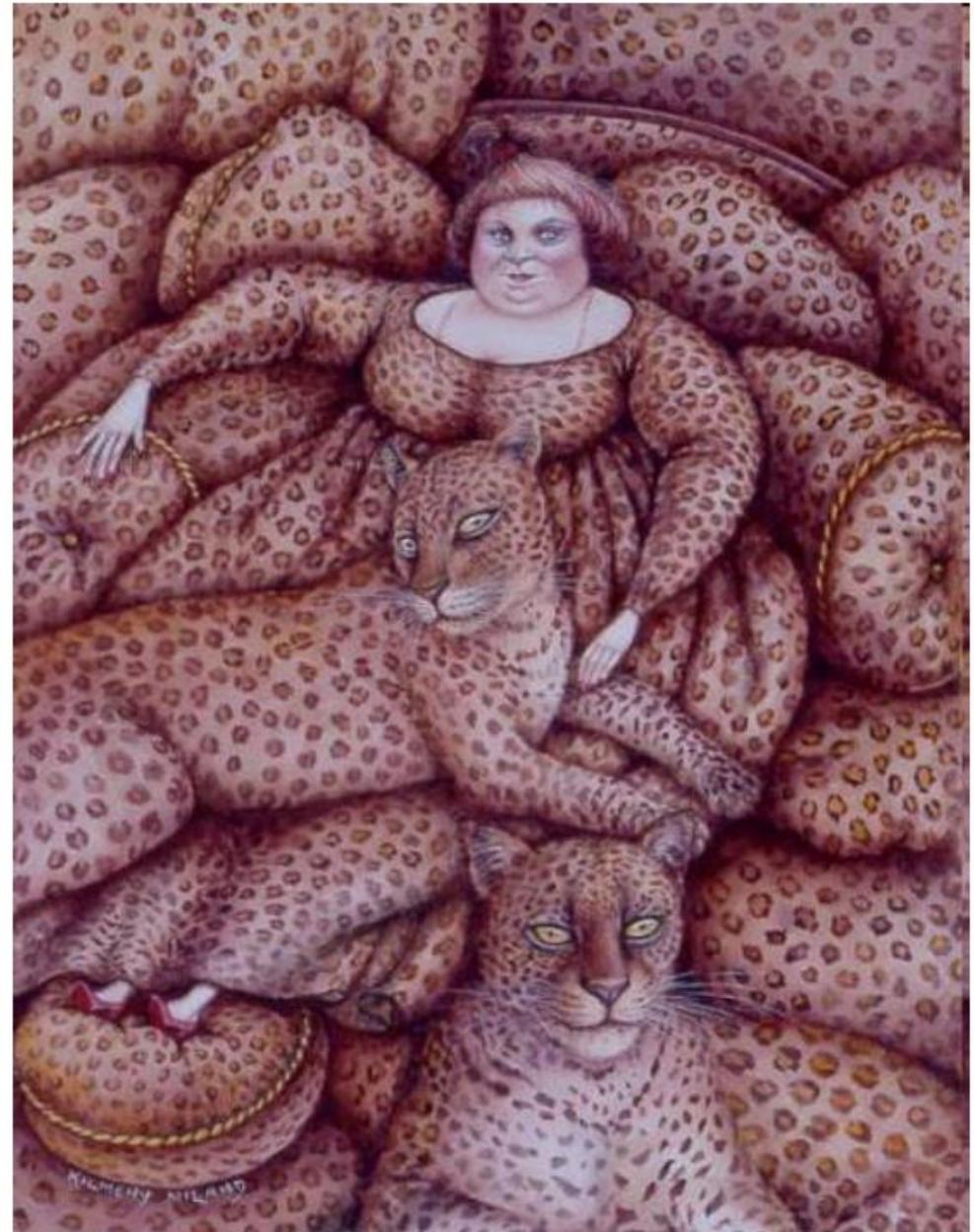
Magritte, 1957



Slide credit: Juan Carlos Niebles

Background

Kilmeny Niland. 1995

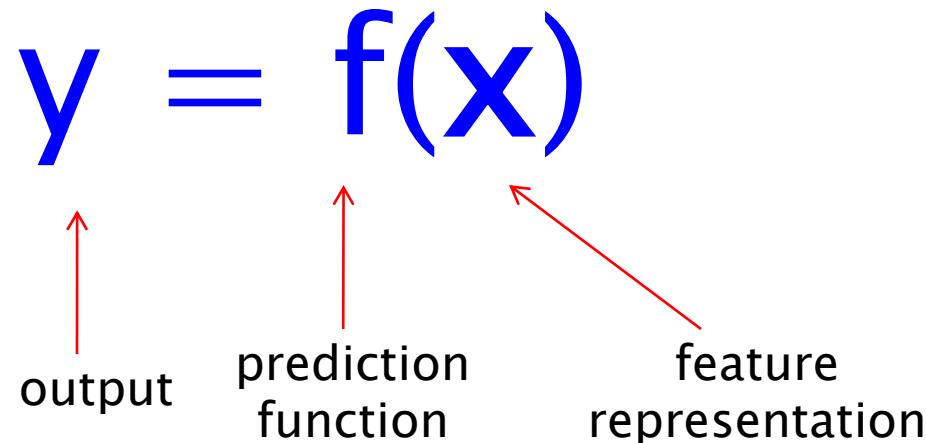


Intra-class variation



Slide credit: Juan Carlos Niebles

Machine Learning Framework



- ▶ **Training:** given a *training set* of labeled examples $\{(x_1, y_1), \dots, (x_N, y_N)\}$, 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 $y = f(x)$

Steps

Training

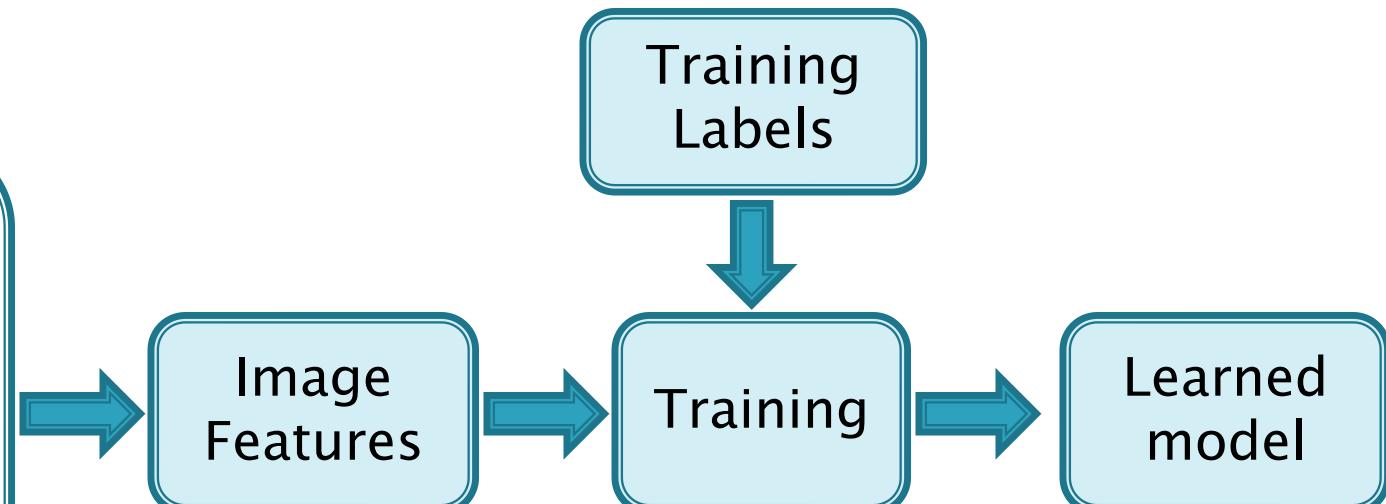
Training Images



Testing



Test Image



Learned
model

Image
Features

Prediction

Training
Labels

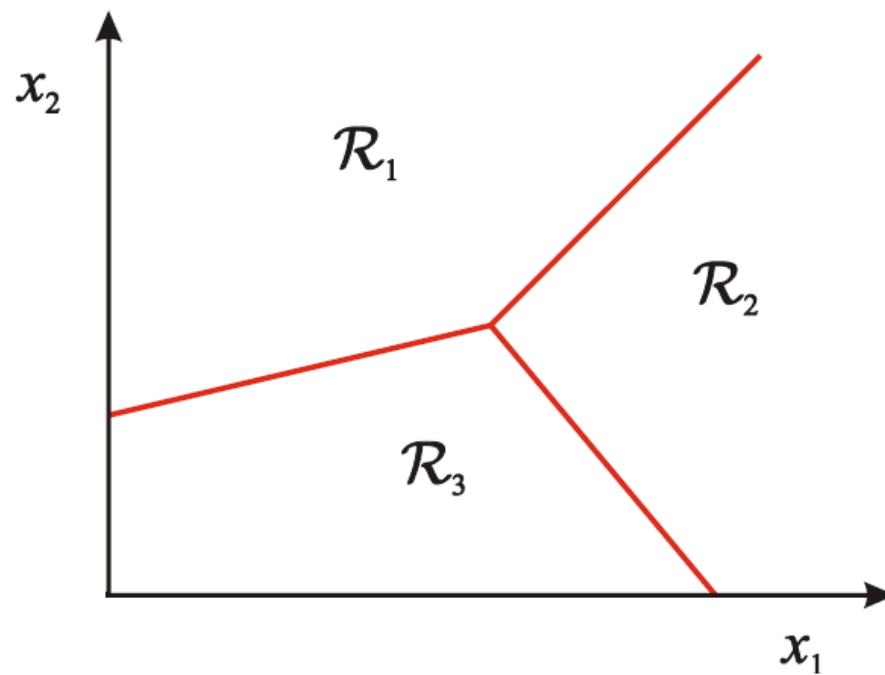
Training

Learned
model

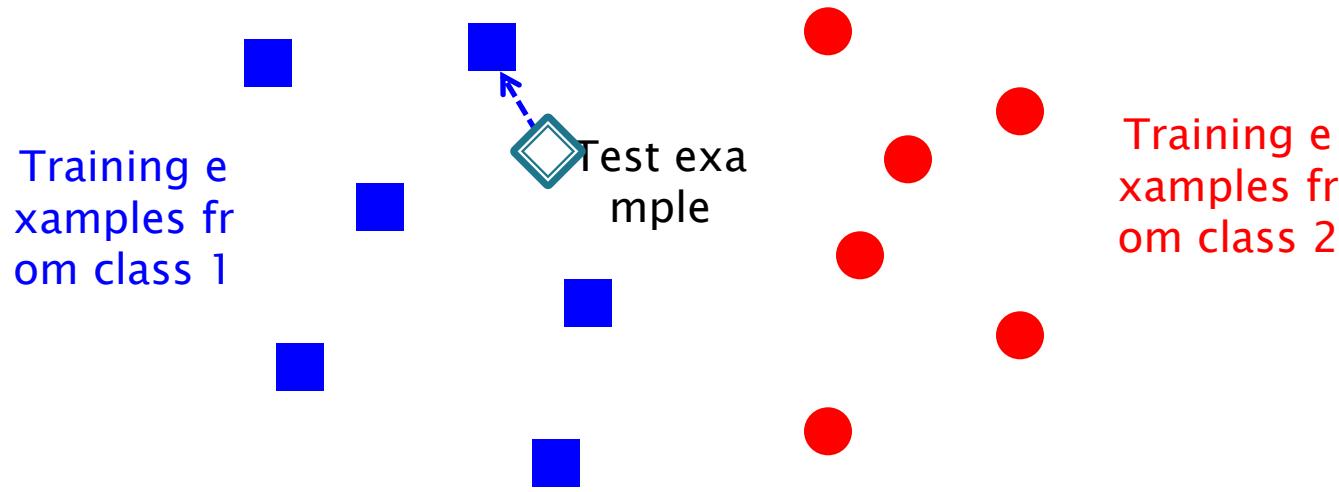


Classification

- Assign input vector to one of two or more classes
- Any decision rule divides input space into *decision regions* separated by *decision boundaries*



Nearest neighbor

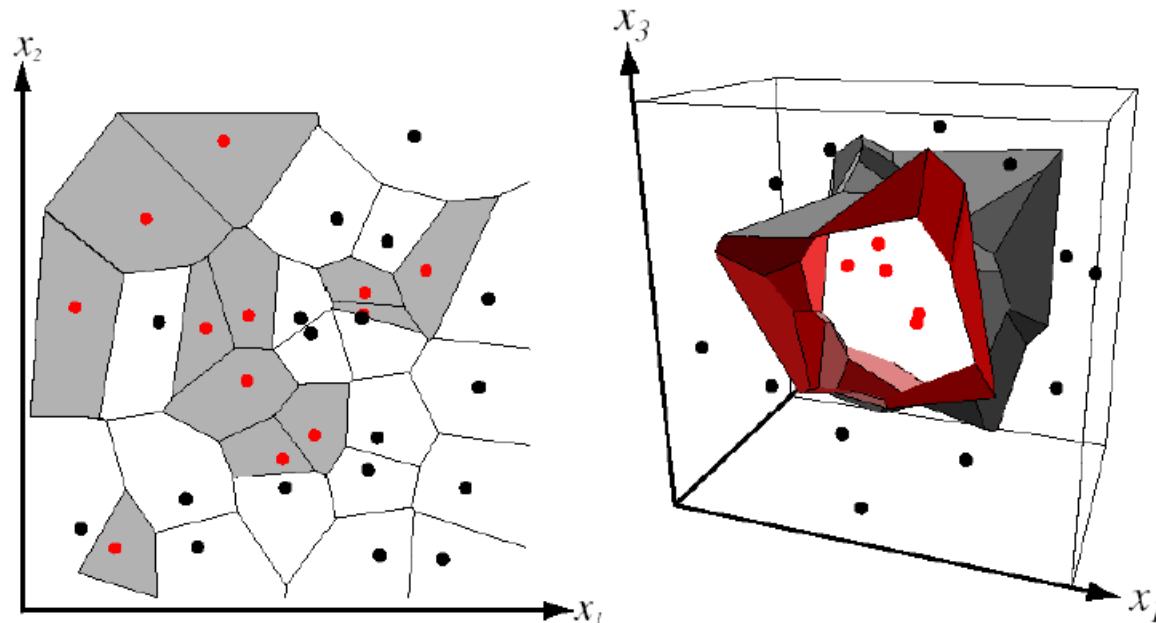


$f(x)$ = label of the training example nearest to x

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

Nearest Neighbor Classifier

- Assign label of nearest training data point to each test data point



from Duda *et al.*

partitioning of feature space
for two-category 2D and 3D data

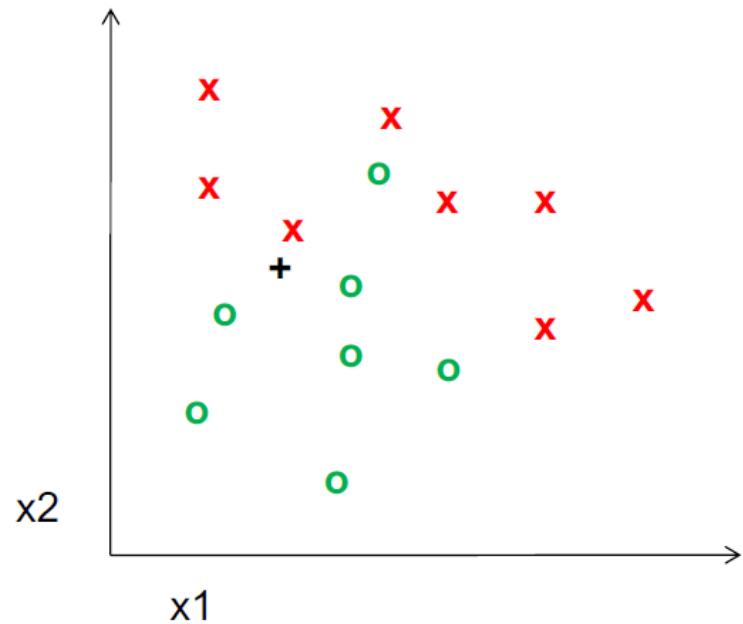
K-Nearest Neighbor

- For a new point, find the k closest points from training data
- Vote for class label with labels of the k points

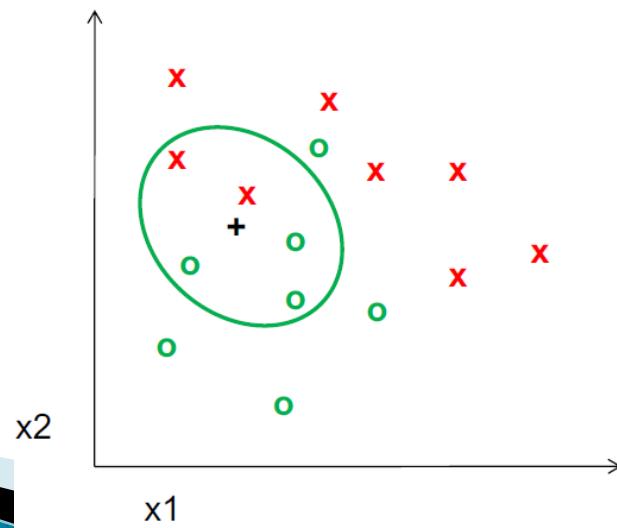
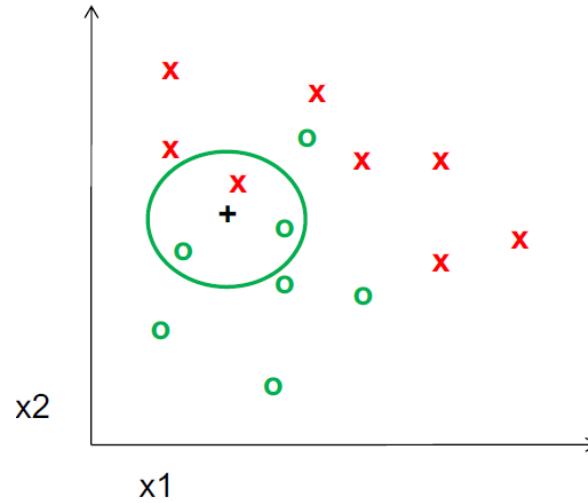
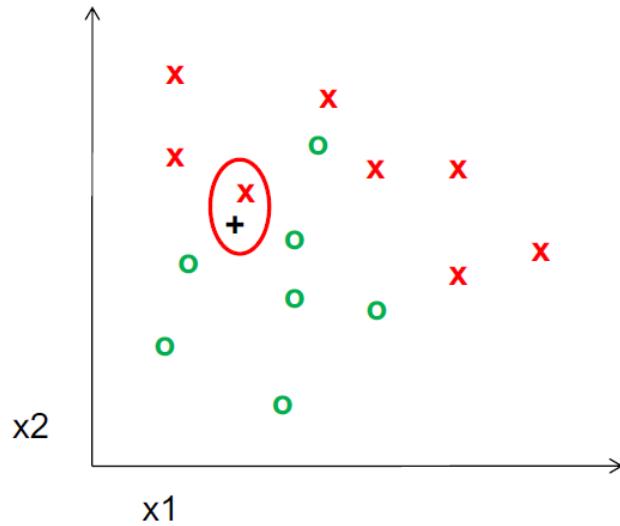
Distance measure - Euclidean

$$Dist(X^n, X^m) = \sqrt{\sum_{i=1}^D (X_i^n - X_i^m)^2}$$

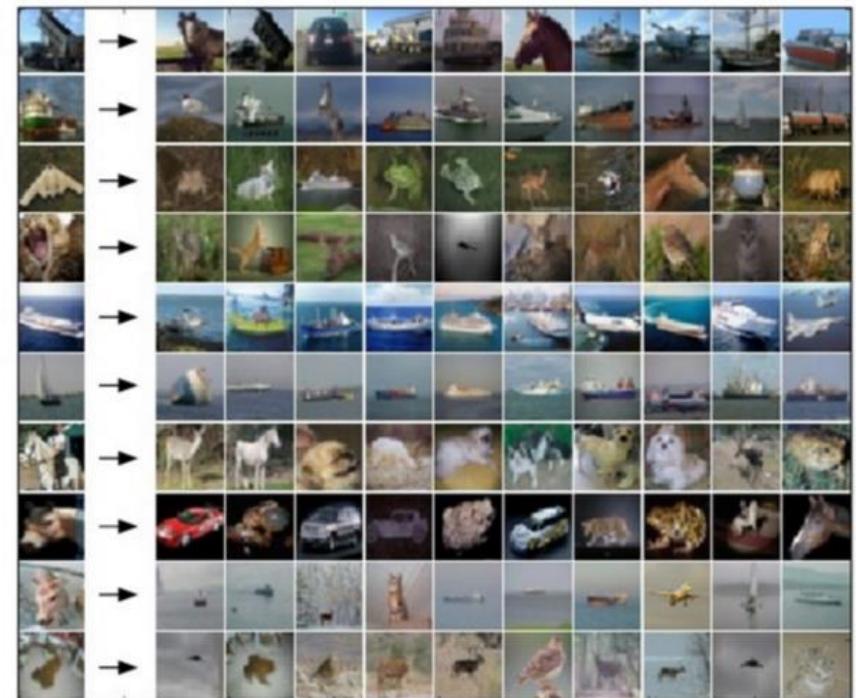
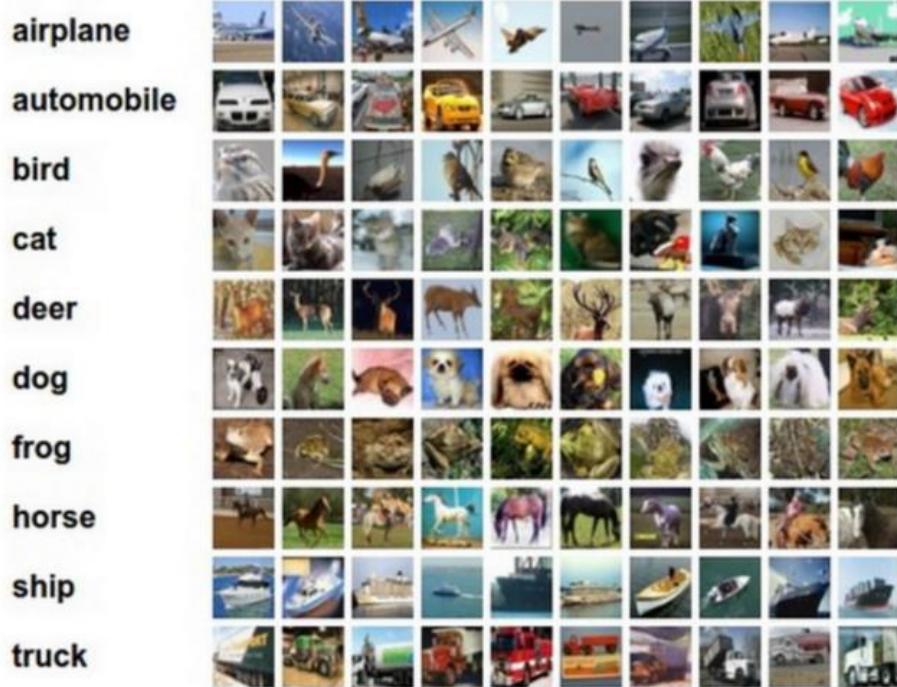
Where X^n and X^m are the n-th and m-th data points



K-Nearest Neighbor



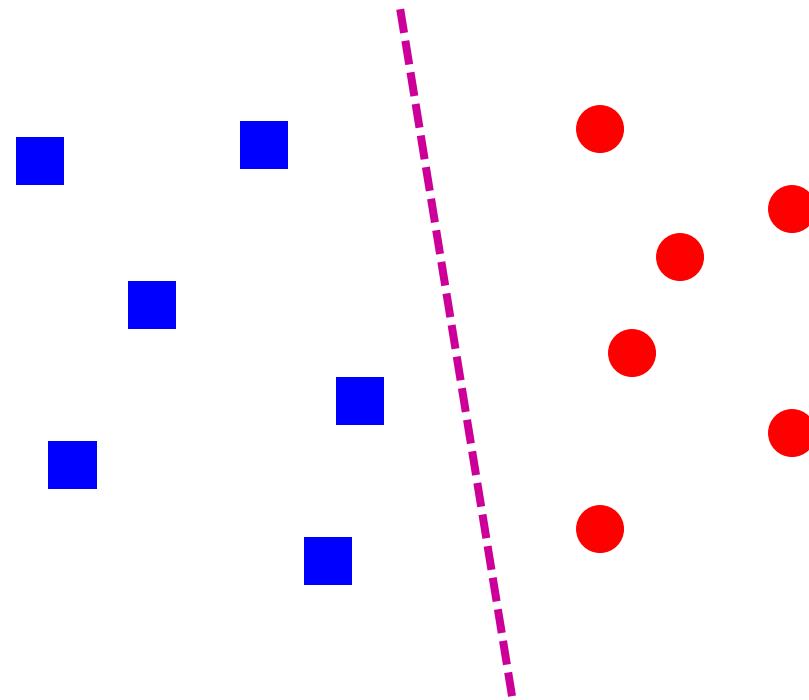
K-nearest neighbor classifier



Left: Example images from the [CIFAR-10 dataset](#). Right: first column shows a few test images and next to each we show the top 10 nearest neighbors in the training set according to pixel-wise difference.

Credit: Andrej Karpathy

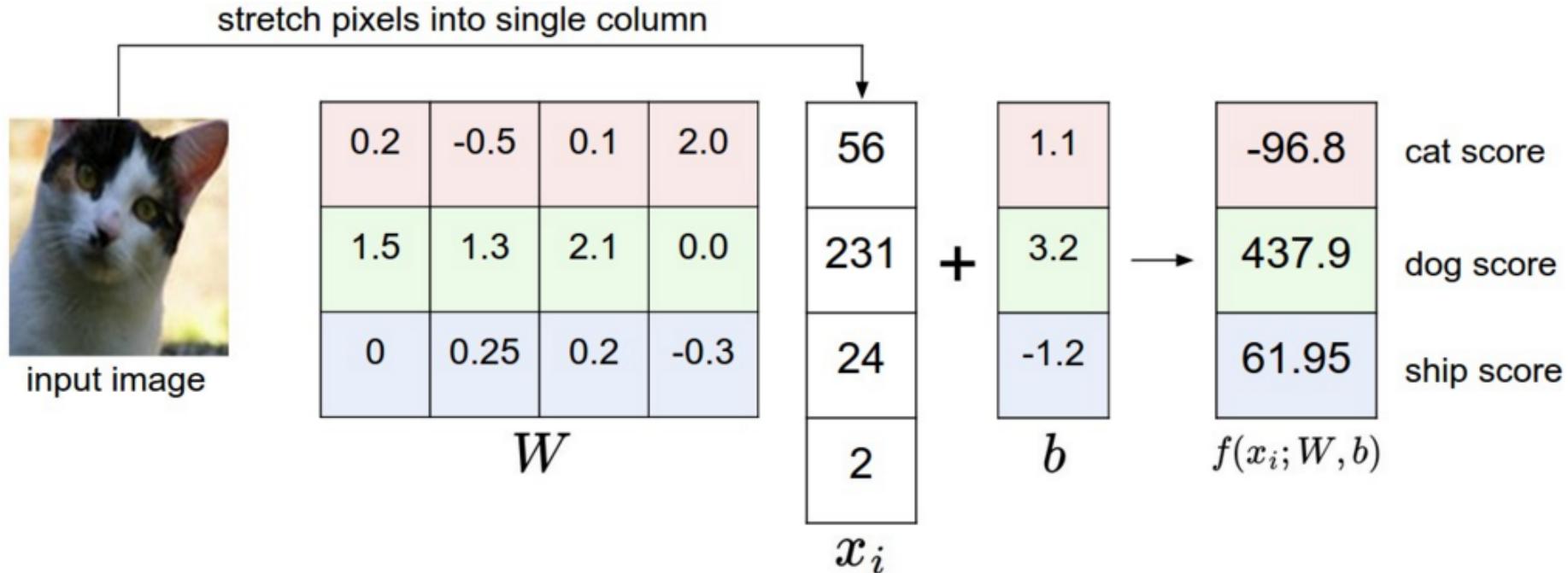
Linear Classifiers



- Find a *linear function* to separate the classes:

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

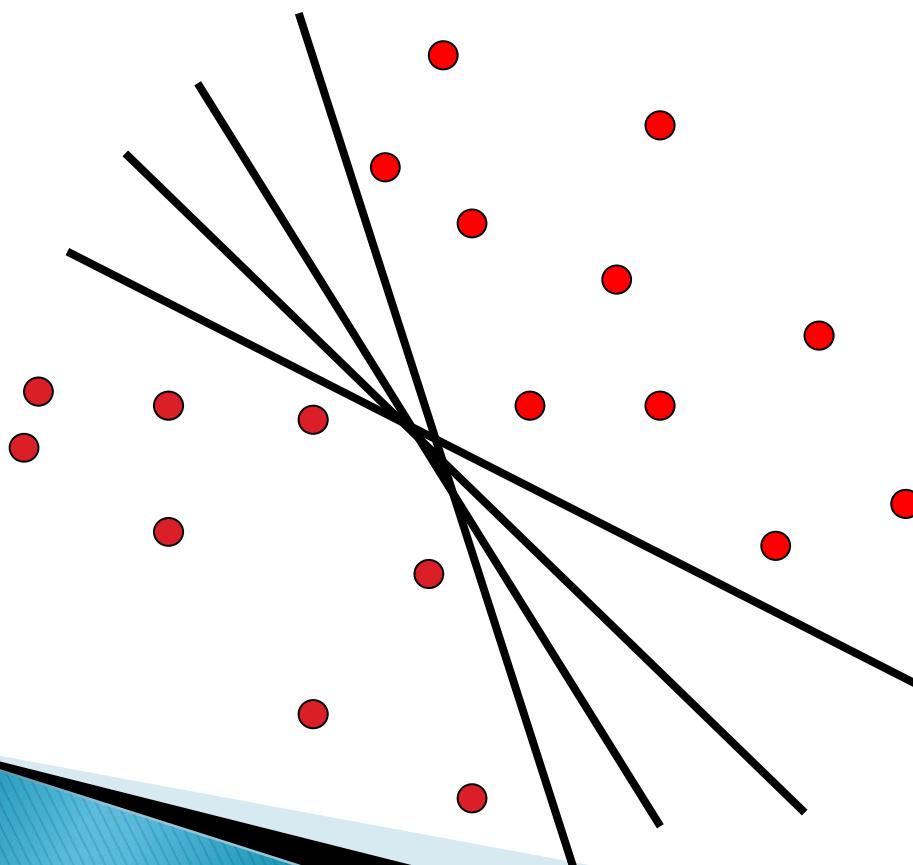
Visualizing linear classifiers



Source: Andrej Karpathy

Linear classifiers

- When the data is linearly separable, there may be more than one separator (hyperplane)

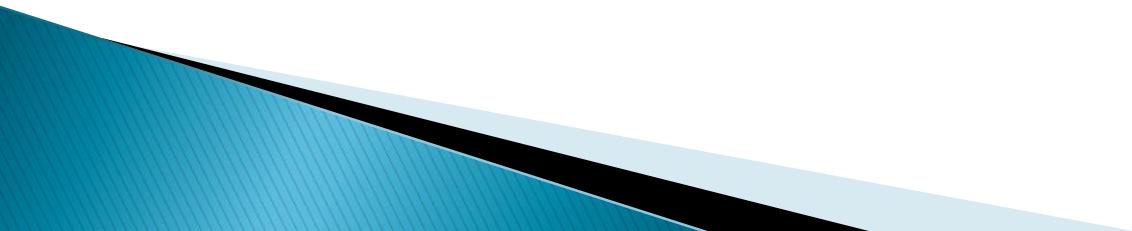


Which separator
is best?

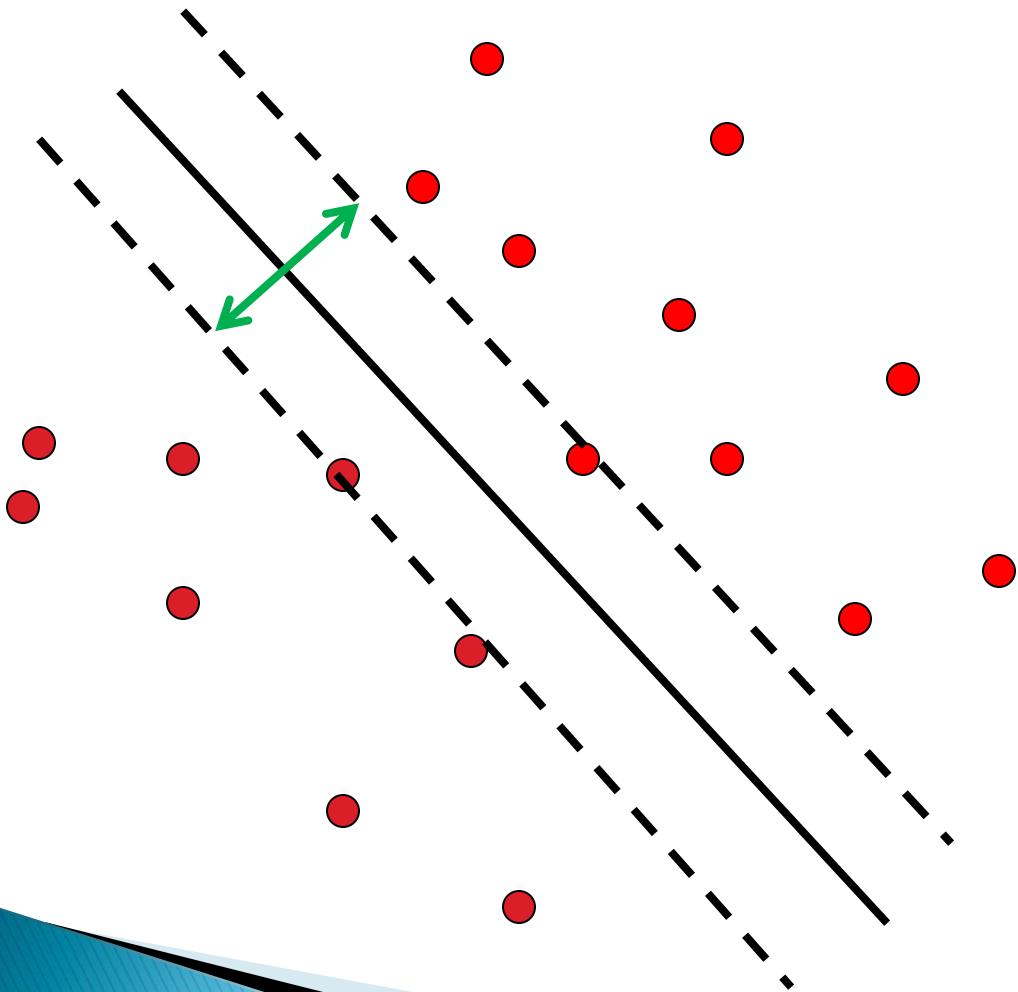
Nearest neighbor vs. linear classifiers

- NN pros:
 - Simple to implement
 - Decision boundaries not necessarily linear
 - Works for any number of classes
 - *Nonparametric* method
- NN cons:
 - Need good distance function
 - Slow at test time
- Linear pros:
 - Low-dimensional *parametric* representation
 - Very fast at test time
- Linear cons:
 - Works for two classes
 - How to train the linear function?
 - What if data is not linearly separable?

Support Vector Machines (SVMs)

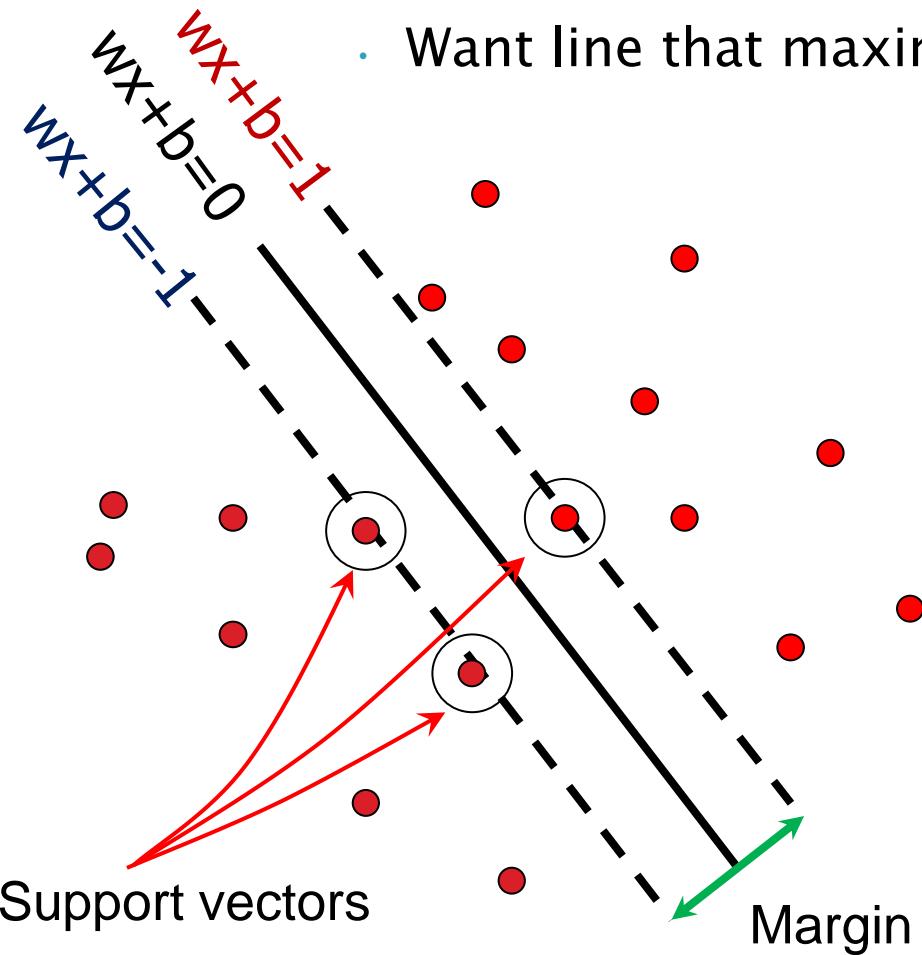


Support Vector Machines (SVMs)



- ▶ Discriminative classifier based on *optimal separating line* (for 2d case)
- ▶ Maximize the *margin* between the positive and negative training examples

Support vector machines



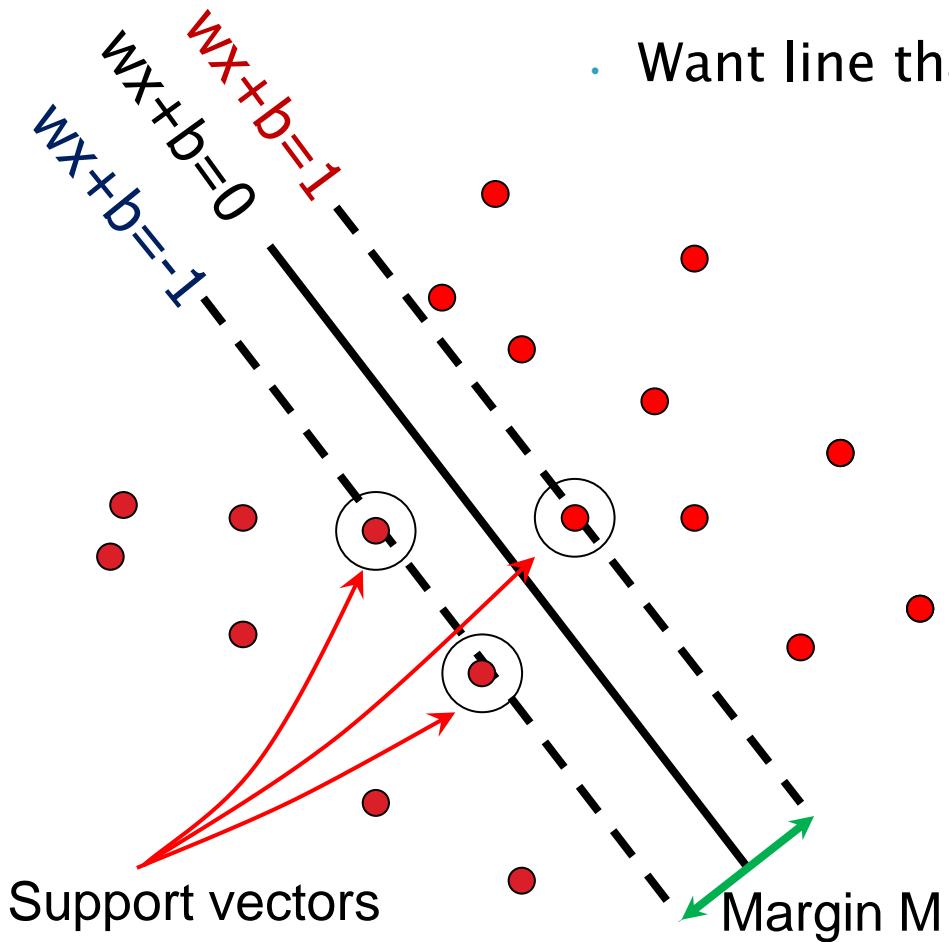
- Want line that maximizes the margin.

\mathbf{x}_i positive ($y_i = 1$): $\mathbf{x}_i \cdot \mathbf{w} + b \geq 1$

\mathbf{x}_i negative ($y_i = -1$): $\mathbf{x}_i \cdot \mathbf{w} + b \leq -1$

For support vectors, $\mathbf{x}_i \cdot \mathbf{w} + b = \pm 1$

Support vector machines



- Want line that maximizes the margin.

$$\mathbf{x}_i \text{ positive } (y_i = 1): \quad \mathbf{x}_i \cdot \mathbf{w} + b \geq 1$$

$$\mathbf{x}_i \text{ negative } (y_i = -1): \quad \mathbf{x}_i \cdot \mathbf{w} + b \leq -1$$

- For support, vectors, $\mathbf{x}_i \cdot \mathbf{w} + b = \pm 1$

- Distance between point and line:
$$\frac{|\mathbf{x}_i \cdot \mathbf{w} + b|}{\|\mathbf{w}\|}$$

- For support vectors:

$$\frac{\mathbf{w}^T \mathbf{x} + b}{\|\mathbf{w}\|} = \frac{\pm 1}{\|\mathbf{w}\|} \quad M = \left| \frac{1}{\|\mathbf{w}\|} - \frac{-1}{\|\mathbf{w}\|} \right| = \frac{2}{\|\mathbf{w}\|}$$

Finding the maximum margin line

Maximize margin $2/\|\mathbf{w}\|$

Correctly classify all training data points:

$$\mathbf{x}_i \text{ positive } (y_i = 1): \quad \mathbf{x}_i \cdot \mathbf{w} + b \geq 1$$

$$\mathbf{x}_i \text{ negative } (y_i = -1): \quad \mathbf{x}_i \cdot \mathbf{w} + b \leq -1$$

- ▶ *Quadratic optimization problem:*

$$\text{Minimize } \frac{1}{2} \mathbf{w}^T \mathbf{w}$$

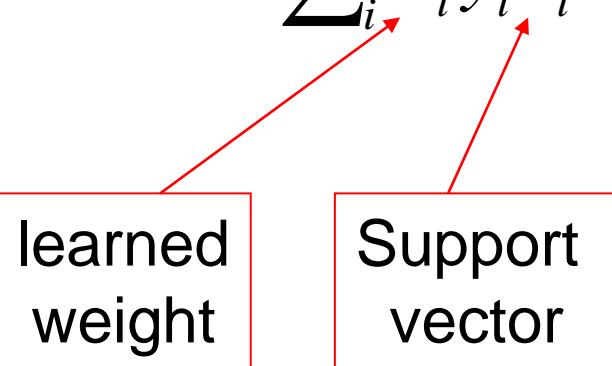
$$\text{Subject to } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1$$

Finding the maximum margin line

- . Solution: $\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$

learned
weight

Support
vector



Finding the maximum margin line

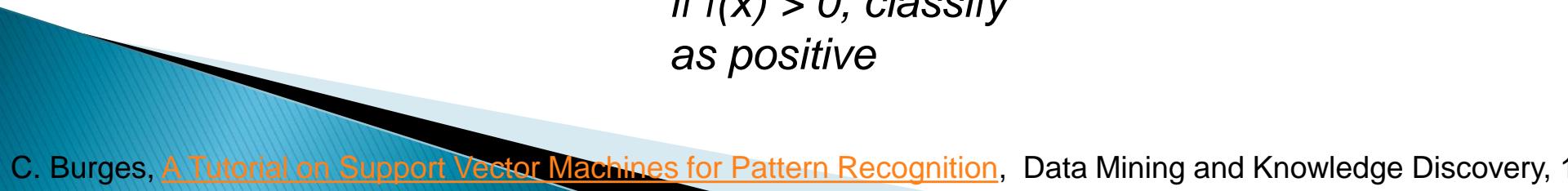
- Solution: $\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$
 $b = y_i - \mathbf{w} \cdot \mathbf{x}_i$ (for any support vector)

$$\mathbf{w} \cdot \mathbf{x} + b = \sum_i \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b$$

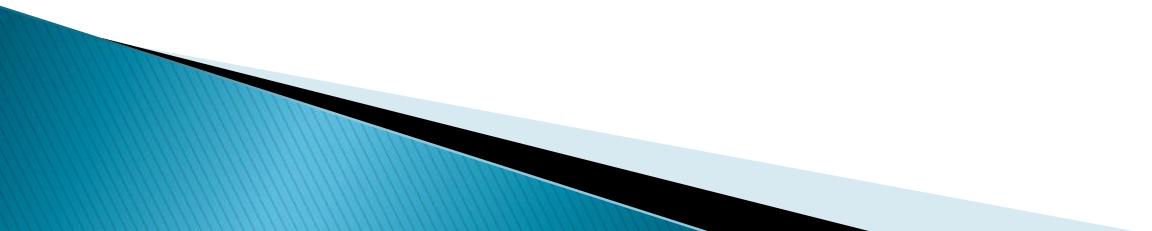
- Classification function:

$$\begin{aligned}f(x) &= \text{sign}(\mathbf{w} \cdot \mathbf{x} + b) \\&= \text{sign}\left(\sum_i \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b\right)\end{aligned}$$

*If $f(x) < 0$, classify
as negative,
if $f(x) > 0$, classify
as positive*



Bag of Features

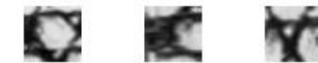
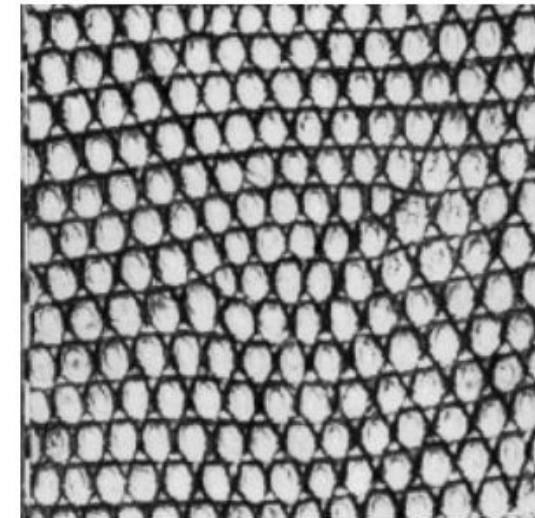
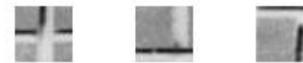
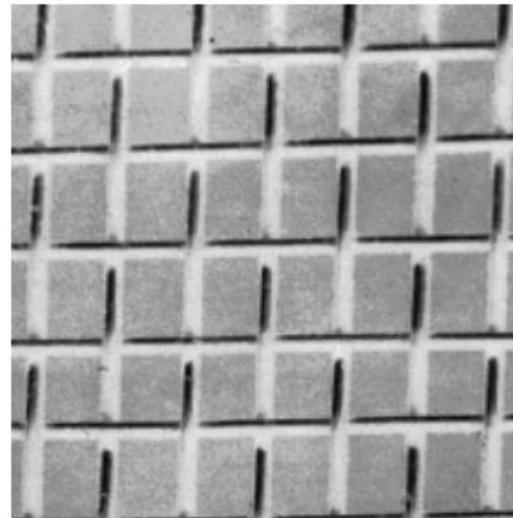
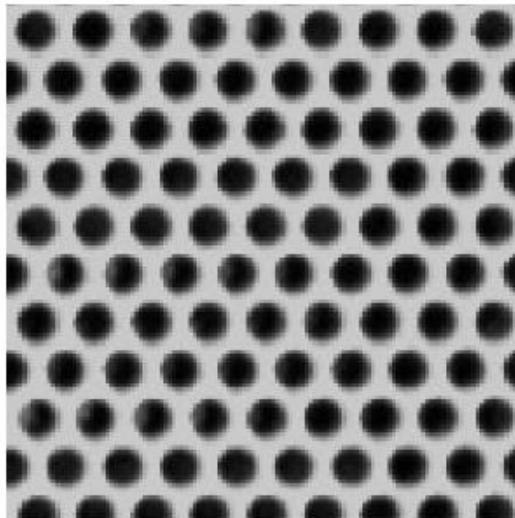


Bag of Features



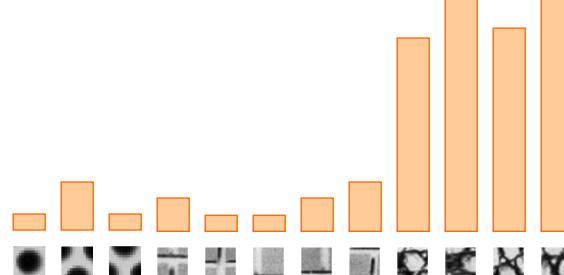
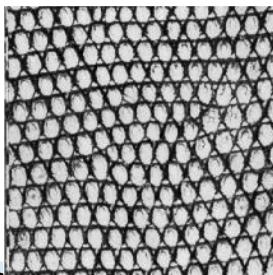
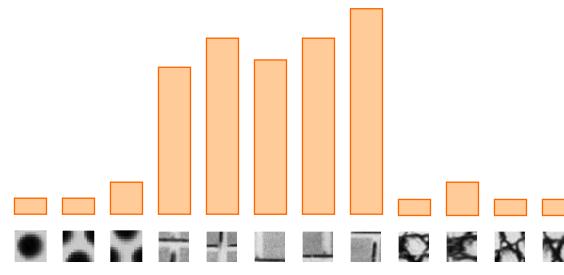
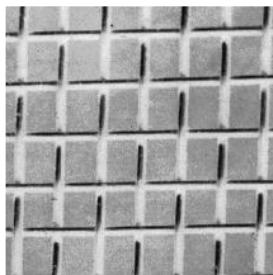
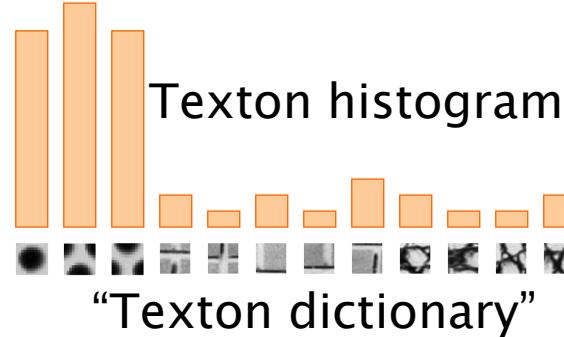
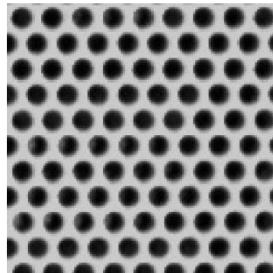
Texture models

- ▶ Texture is characterized by the repetition of basic elements or ***textons***



Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001; Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

Texture models



Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001; Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

Bags of words

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

Bags of words

2007-01-23: State of the Union Address

George W. Bush (2001-)

abandon accountable affordable afghanistan africa aided ally anbar armed army **baghdad** bless **challenges** chamber chaos
choices civilians coalition commanders **commitment** confident confront congressman constitution corps debates deduction
deficit deliver **democratic** deploy dikembe diplomacy disruptions earmarks **economy** einstein **elections** eliminates
expand **extremists** failing faithful families **freedom** fuel **funding** god haven ideology immigration impose
insurgents **iran** **iraq** islam julie lebanon love madam marine math medicare moderation neighborhoods nuclear offensive
palestinian payroll province pursuing **qaeda** radical regimes resolve retreat **rieman** sacrifices science sectarian senate
september **shia** stays strength students succeed **sunni** **tax** territories **terrorists** threats uphold victory
violence violent **war** washington weapons wesley

Bags of words

2007-01-23: State of the Union Address

George W. Bush (2001-)

abandon
choices
deficit
expand
insurgents
palestinians
september
violence

1962-10-22: Soviet Missiles in Cuba

John F. Kennedy (1961-63)

abandon achieving adversaries aggression agricultural appropriate armaments arms assessments atlantic ballistic berlin buildup burdens cargo college commitment communist constitution consumers cooperation crisis cuba dangers declined defensive deficit depended disarmament divisions domination doubled economic education elimination emergence endangered equals europe expand exports fact false family forum freedom fulfill gromyko halt hazards hemisphere hospitals ideals independent industries inflation labor latin limiting minister missiles modernization neglect nuclear oas obligation observer offensive peril pledged predicted purchasing quarantine quote recession rejection republics retaliatory safeguard sites solution soviet space spur stability standby strength surveillance tax territory treaty undertakings unemployment war warhead weapons welfare western widen withdraw

Bags of words

2007-01-23: State of the Union Address

George W. Bush (2001-)

abandon
choices
deficit
expand
insurgents
palestinians
september
violence

1962-10-22: Soviet Missiles in Cuba

John F. Kennedy (1961-63)

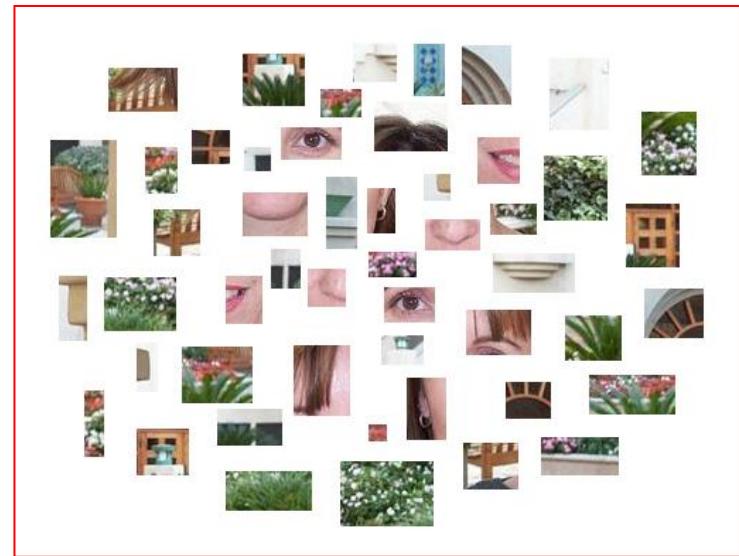
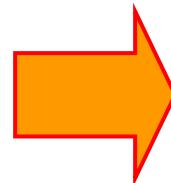
abandon
build
declined
eliminate
halt
modern
recession
surveil

1941-12-08: Request for a Declaration of War

Franklin D. Roosevelt (1933-45)

abandoning acknowledge aggression aggressors airplanes armaments **armed army** assault assembly authorizations bombing
britain british cheerfully claiming constitution curtail december defeats defending delays democratic dictators disclose
economic empire endanger **facts** false forgotten fortunes france **freedom** fulfilled fullness fundamental gangsters
german germany god guam harbor hawaii **hemisphere** hint hitler hostilities immune improving indies innumerable
invasion **islands** isolate **japanese** labor metals midst midway **navy** nazis obligation offensive
officially **pacific** partisanship patriotism pearl peril perpetrated perpetual philippine preservation privilege reject
repaired **resisting** retain revealing rumors seas soldiers speaks speedy stamina **strength** sunday sunk supremacy tanks taxes
treachery true tyranny undertaken victory **war** wartime washington

Bag of Features for Object Recognition

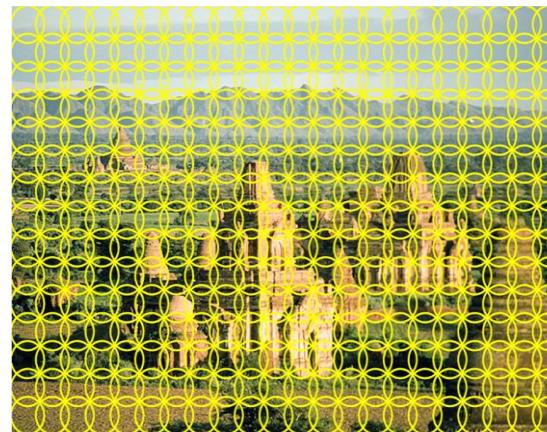
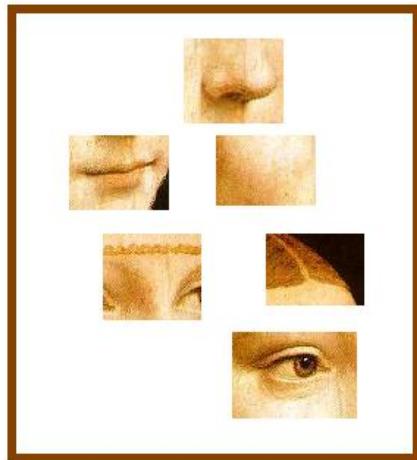


Csurka et al. (2004), Willamowski et al. (2005), Grauman & Darrell (2005), Sivic et al. (2003, 2005)

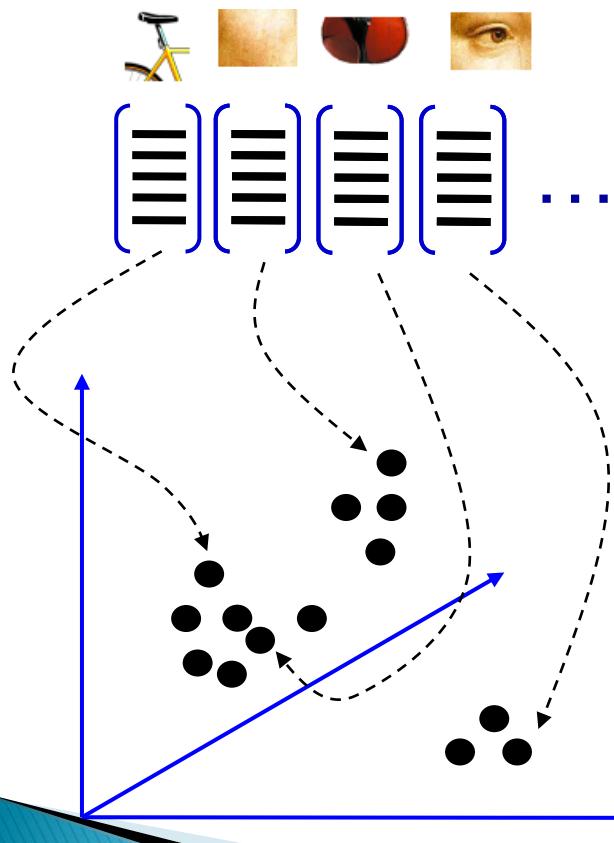
Overview of Bag of Features

- ▶ Extract local features.
- ▶ Learn visual vocabulary.
- ▶ Quantize local features.
- ▶ Represent images by frequencies of visual words.

Step 1: Extract Local Features

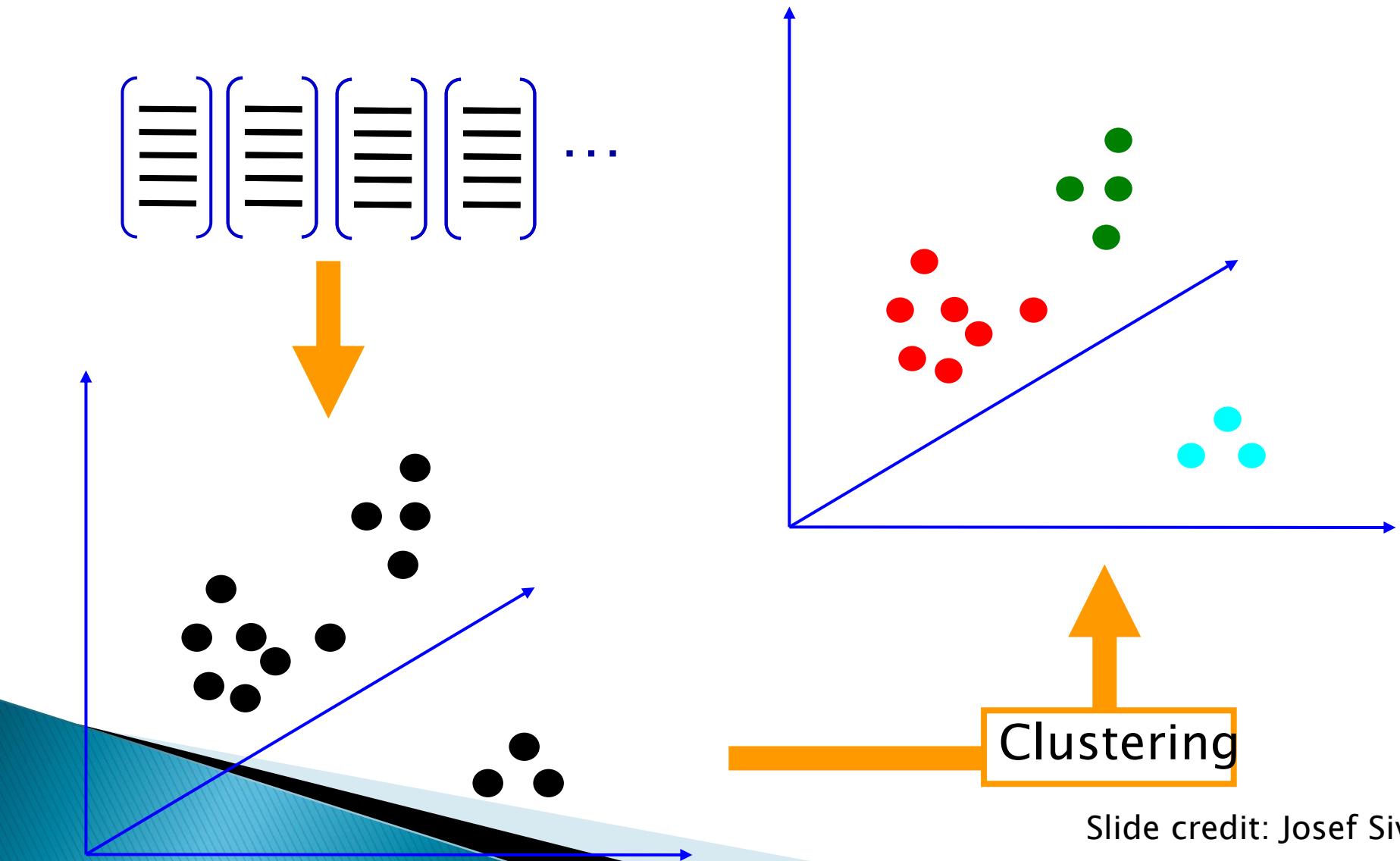


Step 2: Learn Visual Vocabulary



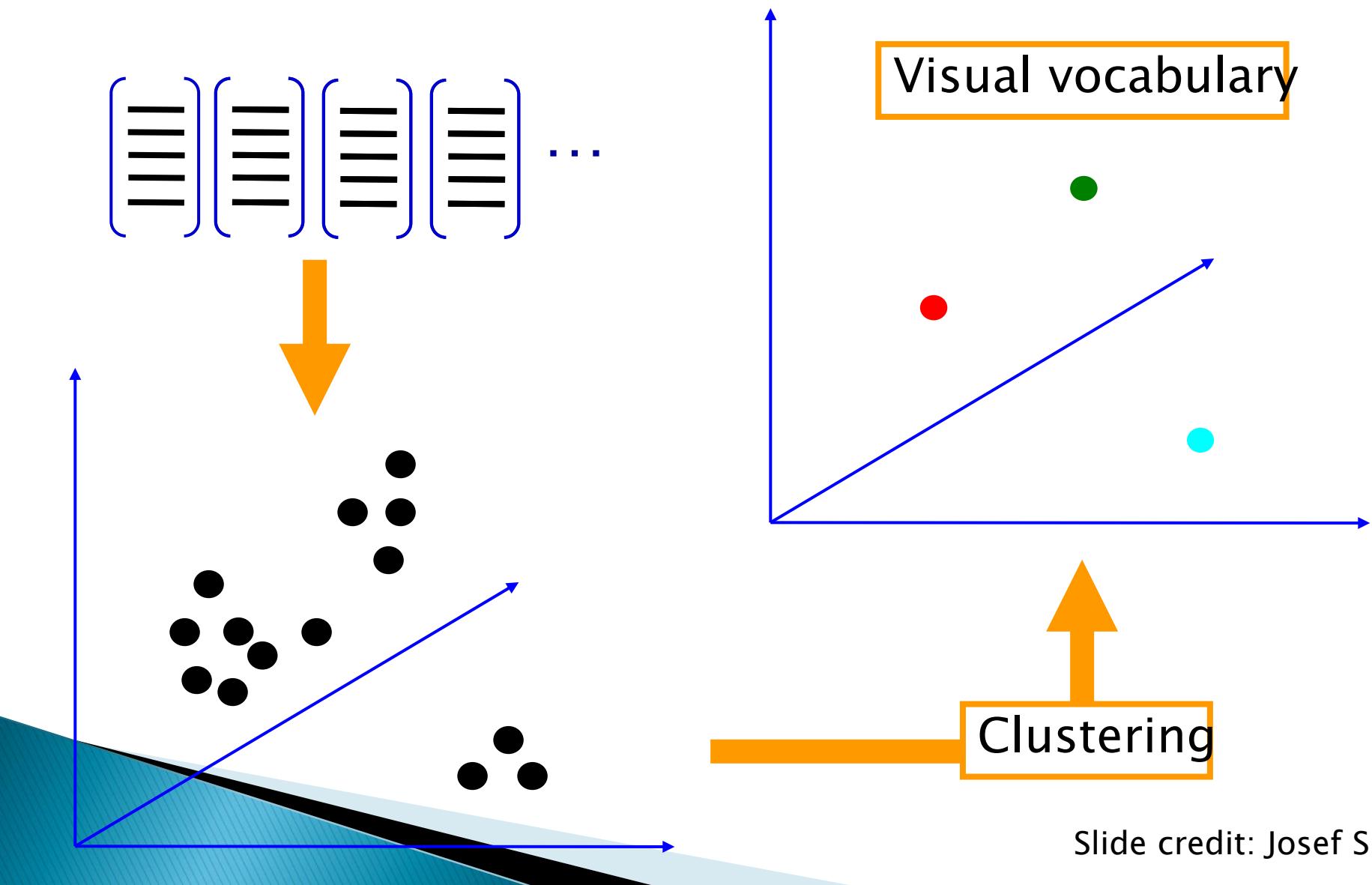
Extracted descriptors
from the training set

Step 2: Learn Visual Vocabulary



Slide credit: Josef Sivic

Step 2: Learn Visual Vocabulary



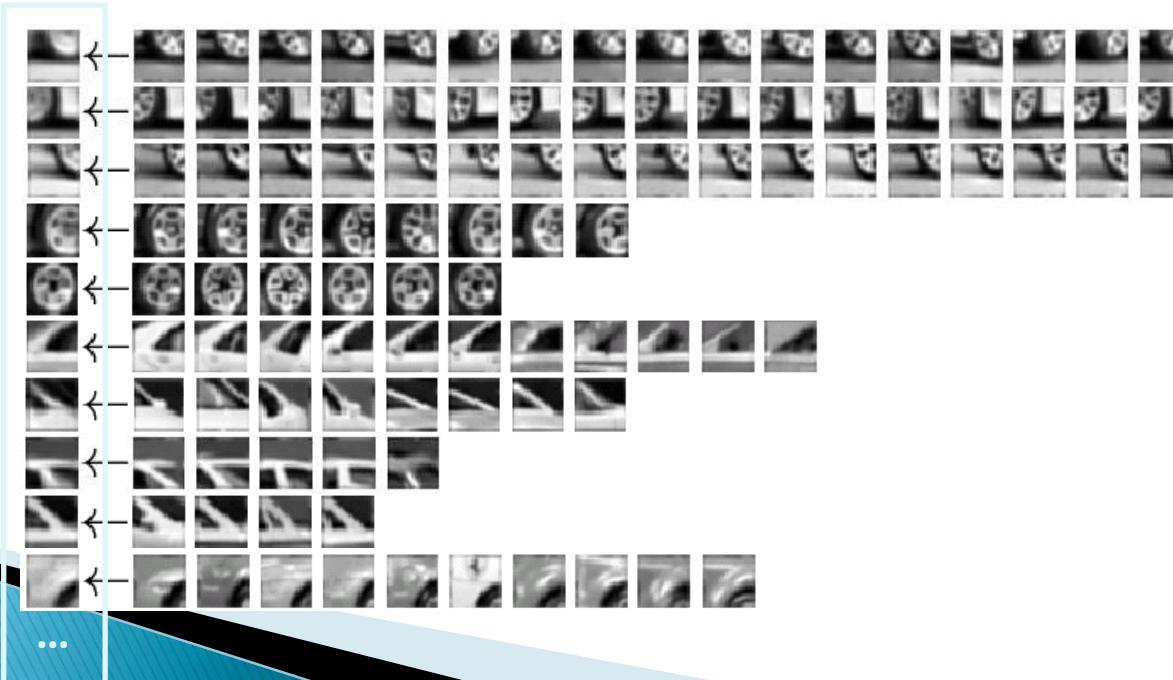
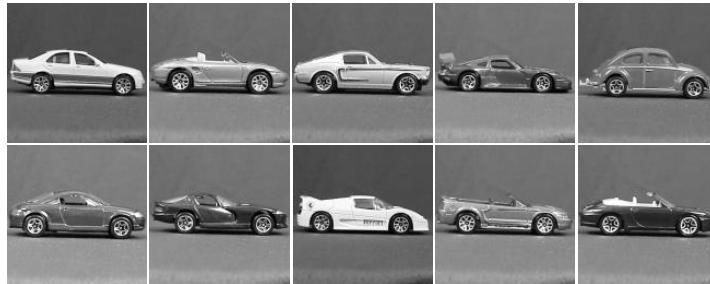
Recall: K-means clustering

- Want to minimize sum of squared Euclidean distances between features \mathbf{x}_i and their nearest cluster centers \mathbf{m}_k

$$D(X, M) = \sum_{\text{cluster } k} \sum_{\substack{\text{point } i \text{ in} \\ \text{cluster } k}} (\mathbf{x}_i - \mathbf{m}_k)^2$$

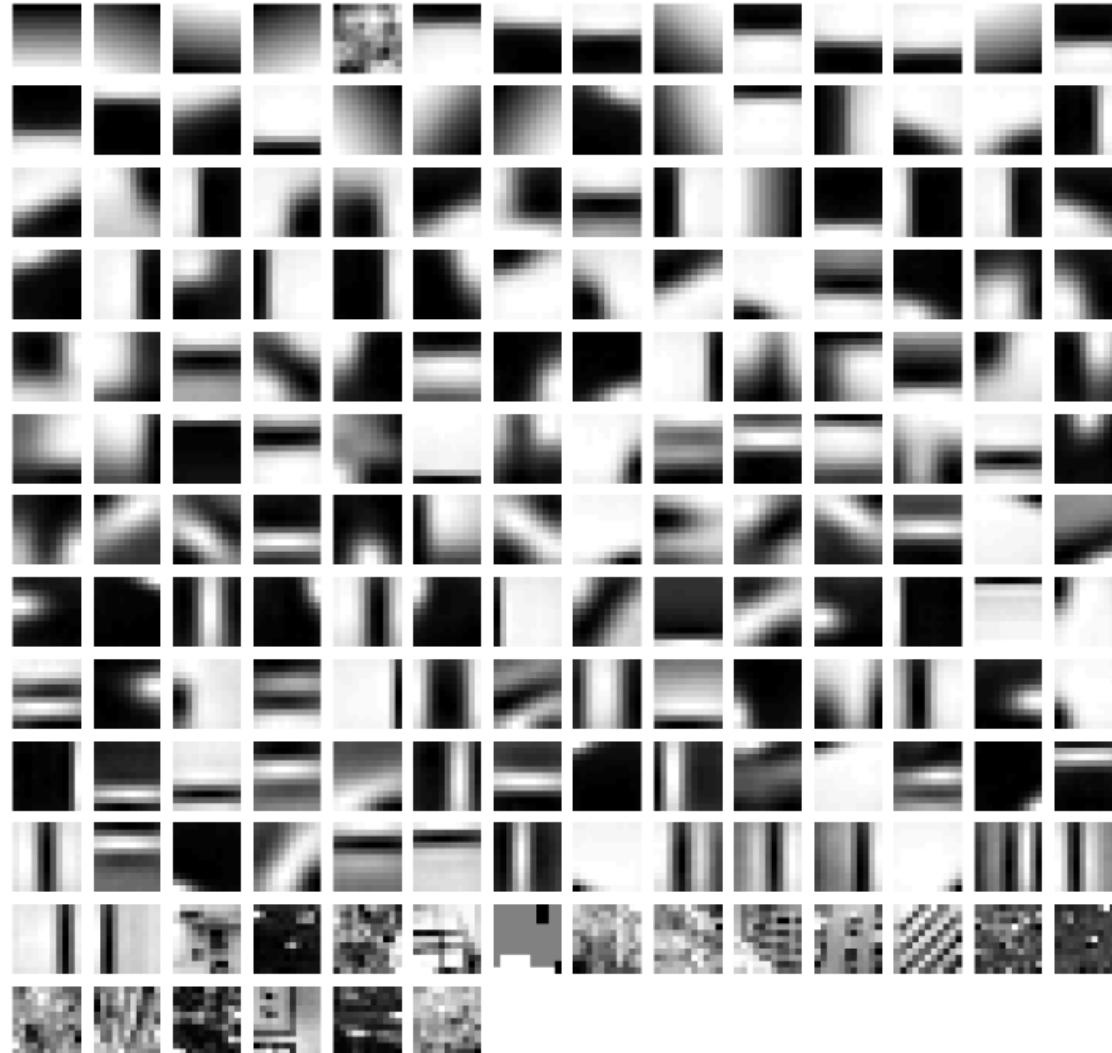
- Algorithm:
 - Randomly initialize K cluster centers
 - Iterate until convergence:
 - Assign each feature to the nearest center
 - Recompute each cluster center as the mean of all features assigned to it

Visual Vocabularies



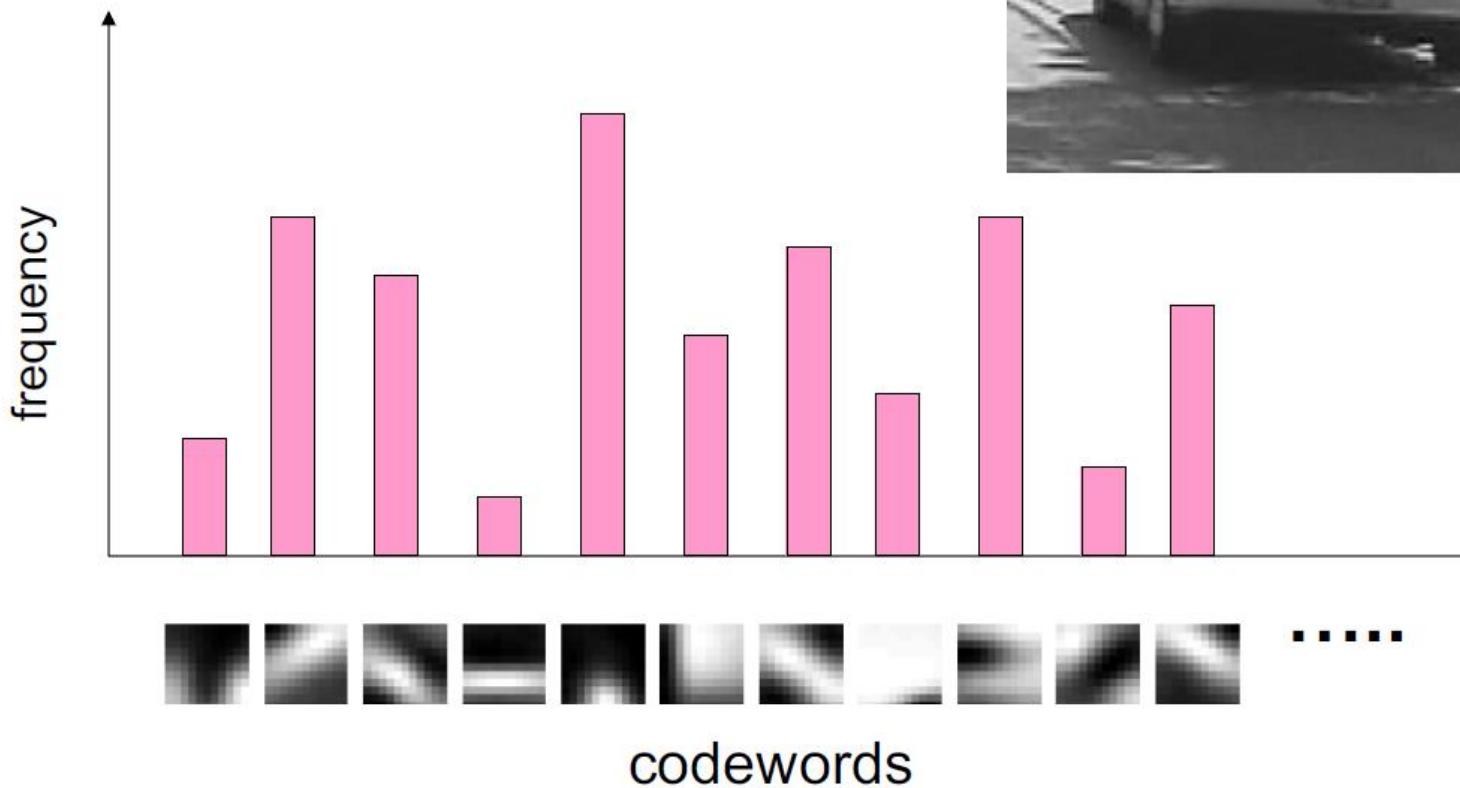
Source: B. Leibe

Example Visual Vocabularies

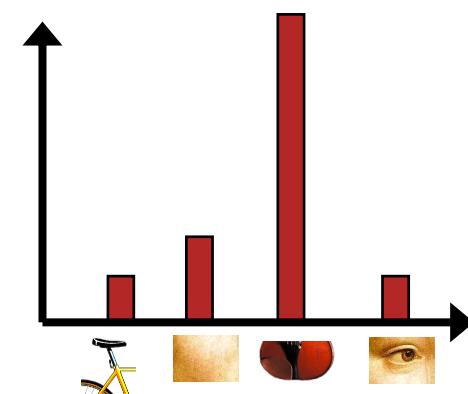
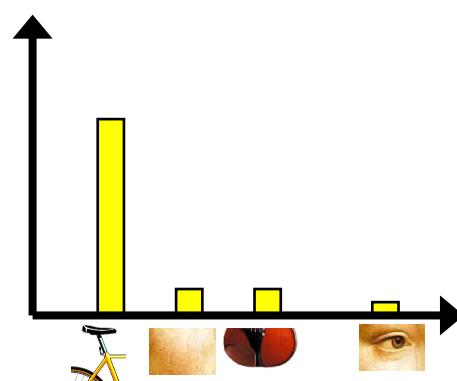
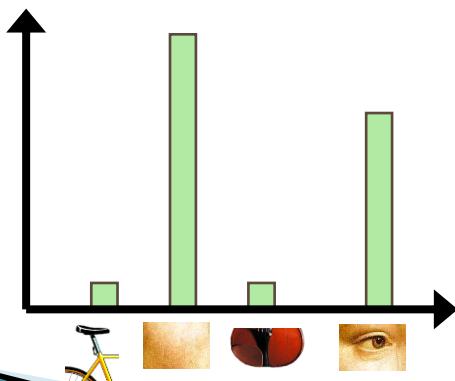
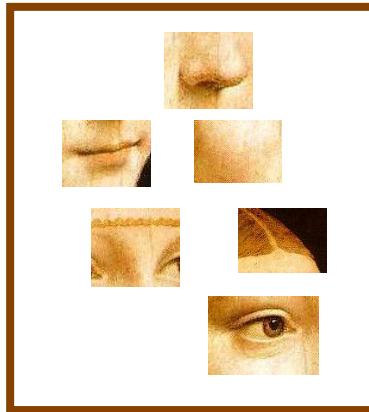


Fei-Fei et al. 2005

Step 3: Image Representation



Step 3: Image Representation



Spatial Pyramids

