# Machine Learning Problems

<table>
<thead>
<tr>
<th>Supervised Learning</th>
<th>Unsupervised Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete</td>
<td>Clustering</td>
</tr>
<tr>
<td>Classification or Categorization</td>
<td>Regression</td>
</tr>
<tr>
<td>Continuous</td>
<td>Dimensionality Reduction</td>
</tr>
</tbody>
</table>
Supervised learning

\[ y = f(x) \]

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

Slide credit: L. Lazebnik
Image Categorization

**Training**
- Training Images
- Image Features
- Classifier Training
- Trained Classifier

**Testing**
- Test Image
- Image Features
- Trained Classifier
- Prediction
- Outdoor
Example: Scene Categorization

• Is this a kitchen?
Image features

Training Images → Image Features → Classifier Training → Trained Classifier

Training Labels
Features

• Raw pixels

• Histograms

• GIST descriptors

• …
Classifiers

Training Images

Training

Image Features

Classifier Training

Training Labels

Trained Classifier
Learning a classifier

Given some set of features with corresponding labels, learn a function to predict the labels from the features
Recognition: Overview and History

Slides from Lana Lazebnik, Fei-Fei Li, Rob Fergus, Antonio Torralba, and Jean Ponce
How many visual object categories are there?

~10,000 to 30,000

Biederman 1987
Specific recognition tasks
Scene categorization or classification

- outdoor/indoor
- city/forest/factory/etc.

Svetlana Lazebnik
Image annotation / tagging / attributes

- street
- people
- building
- mountain
- tourism
- cloudy
- brick
- ...

Svetlana Lazebnik
Object detection

- find pedestrians
Image parsing / semantic segmentation
Scene understanding?
Recognition is all about modeling variability

Variability: Camera position
            Illumination
            Shape parameters

Within-class variations?
Within-class variations
History of ideas in recognition

• 1960s – early 1990s: the geometric era
Variability:
Camera position
Illumination

Shape: assumed known

Alignment

Roberts (1965); Lowe (1987); Faugeras & Hebert (1986); Grimson & Lozano-Perez (1986); Huttenlocher & Ullman (1987)
Recall: Alignment

• Alignment: fitting a model to a transformation between pairs of features (matches) in two images

Find transformation $T$ that minimizes

$$\sum_i \text{residual}(T(x_i), x'_i)$$
Recognition as an alignment problem: Block world


**Fig. 1.** A system for recognizing 3-d polyhedral scenes. a) L.G. Roberts. b) A blocks world scene. c) Detected edges using a 2x2 gradient operator. d) A 3-d polyhedral description of the scene, formed automatically from the single image. e) The 3-d scene displayed with a viewpoint different from the original image to demonstrate its accuracy and completeness. (b) - (e) are taken from [64] with permission MIT Press.)

Representing and recognizing object categories is harder...

ACRONYM (Brooks and Binford, 1981)
Binford (1971), Nevatia & Binford (1972), Marr & Nishihara (1978)
Recognition by components

Biederman (1987)

Primitives (geons)

- Cube: Straight Edge, Straight Axis, Constant
- Wedge: Straight Edge, Straight Axis, Expanded
- Pyramid: Straight Edge, Straight Axis, Expanded
- Cylinder: Curved Edge, Straight Axis, Constant
- Barrel: Curved Edge, Straight Axis, Exp & Cont
- Arch: Straight Edge, Curved Axis, Constant
- Cone: Curved Edge, Straight Axis, Expanded
- Expanded Cylinder: Curved Edge, Straight Axis, Expanded
- Handle: Curved Edge, Curved Axis, Constant
- Expanded Handle: Curved Edge, Curved Axis, Expanded

Objects


Svetlana Lazebnik
Generalized cylinders
Ponce et al. (1989)

General shape primitives?

Zisserman et al. (1995)

Forsyth (2000)
History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
Empirical models of image variability

Appearance-based techniques

Turk & Pentland (1991); Murase & Nayar (1995); etc.
Eigenfaces (Turk & Pentland, 1991)

<table>
<thead>
<tr>
<th>Experimental Condition</th>
<th>Correct/Unknown Recognition Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lighting</td>
</tr>
<tr>
<td>Forced classification</td>
<td>96/0</td>
</tr>
<tr>
<td>Forced 100% accuracy</td>
<td>100/19</td>
</tr>
<tr>
<td>Forced 20% unknown rate</td>
<td>100/20</td>
</tr>
</tbody>
</table>
Color Histograms

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- 1990s – present: sliding window approaches
Sliding window approaches
Sliding window approaches

- Turk and Pentland, 1991
- Belhumeur, Hespanha, & Kriegman, 1997
- Schneiderman & Kanade, 2004
- Viola and Jones, 2000
- Schneiderman & Kanade, 2004
- Argawal and Roth, 2002
- Poggio et al. 1993
History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
Local features for object instance recognition

Large-scale image search

Combining local features, indexing, and spatial constraints

Image credit: K. Grauman and B. Leibe
Large-scale image search
Combining local features, indexing, and spatial constraints

Philbin et al. ‘07
Large-scale image search
Combining local features, indexing, and spatial constraints
History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
Parts-and-shape models

- Model:
  - Object as a set of parts
  - Relative locations between parts
  - Appearance of part

Figure from [Fischler & Elschlager 73]
Constellation models

Pictorial structure model

Fischler and Elschlager(73), Felzenszwalb and Huttenlocher(00)

\[
\Pr(P_{\text{tor}}, P_{\text{arm}}, \ldots | \text{Im}) \propto \prod_{i,j} \Pr(P_i | P_j) \prod_{i} \Pr(\text{Im}(P_i))
\]
Discriminatively trained part-based models

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features
Bag-of-features models
Bag-of-features models
Objects as texture

- All of these are treated as being the same

- No distinction between foreground and background: scene recognition?
Origin 1: Texture recognition

- Texture is characterized by the repetition of basic elements or *textons*
- For stochastic textures, it is the identity of the textons, not their spatial arrangement, that matters

Origin 1: Texture recognition

Origin 2: Bag-of-words models

Origin 2: Bag-of-words models

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

US Presidential Speeches Tag Cloud
http://chir.ag/phernalia/preztags/
Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary  
  Salton & McGill (1983)
Origin 2: Bag-of-words models


US Presidential Speeches Tag Cloud
http://chir.ag/phernalia/preztags/