Non-Parametric Methods: nearest neighbor rule

CS 189

Alexei Efros

Fall 2015

Key idea: Just store all training examples $\langle x_i, f(x_i) \rangle$

Nearest neighbor:

- Given query instance x_q, first locate nearest training example x_n, then estimate $\hat{f}(x_q) \leftarrow f(x_n)$

Key idea: Just store all training examples $\langle x_i, f(x_i) \rangle$

Nearest neighbor:

- Given query instance x_q, first locate nearest training example x_n, then estimate $\hat{f}(x_q) \leftarrow f(x_n)$

***k*-Nearest neighbor:**

- Given x_q, take vote among its k nearest neighbors (if discrete-valued target function)

Key idea: Just store all training examples $\langle x_i, f(x_i)\rangle$

Nearest neighbor:

- Given query instance x_q, first locate nearest training example x_n, then estimate $\hat{f}(x_q) \leftarrow f(x_n)$

k-Nearest neighbor:

- Given x_q, take vote among its k nearest neighbors (if discrete-valued target function)
- Take mean of f values of k nearest neighbors (if real-valued)

$$\hat{f}(x_q) \leftarrow \frac{1}{k}\sum_{i=1}^{k} f(x_i)$$

Nearest-Neighbor Rule

- Also known as:
  - instance-based learning
  - memory-based learning
  - exemplar-based learning
  - lazy learning

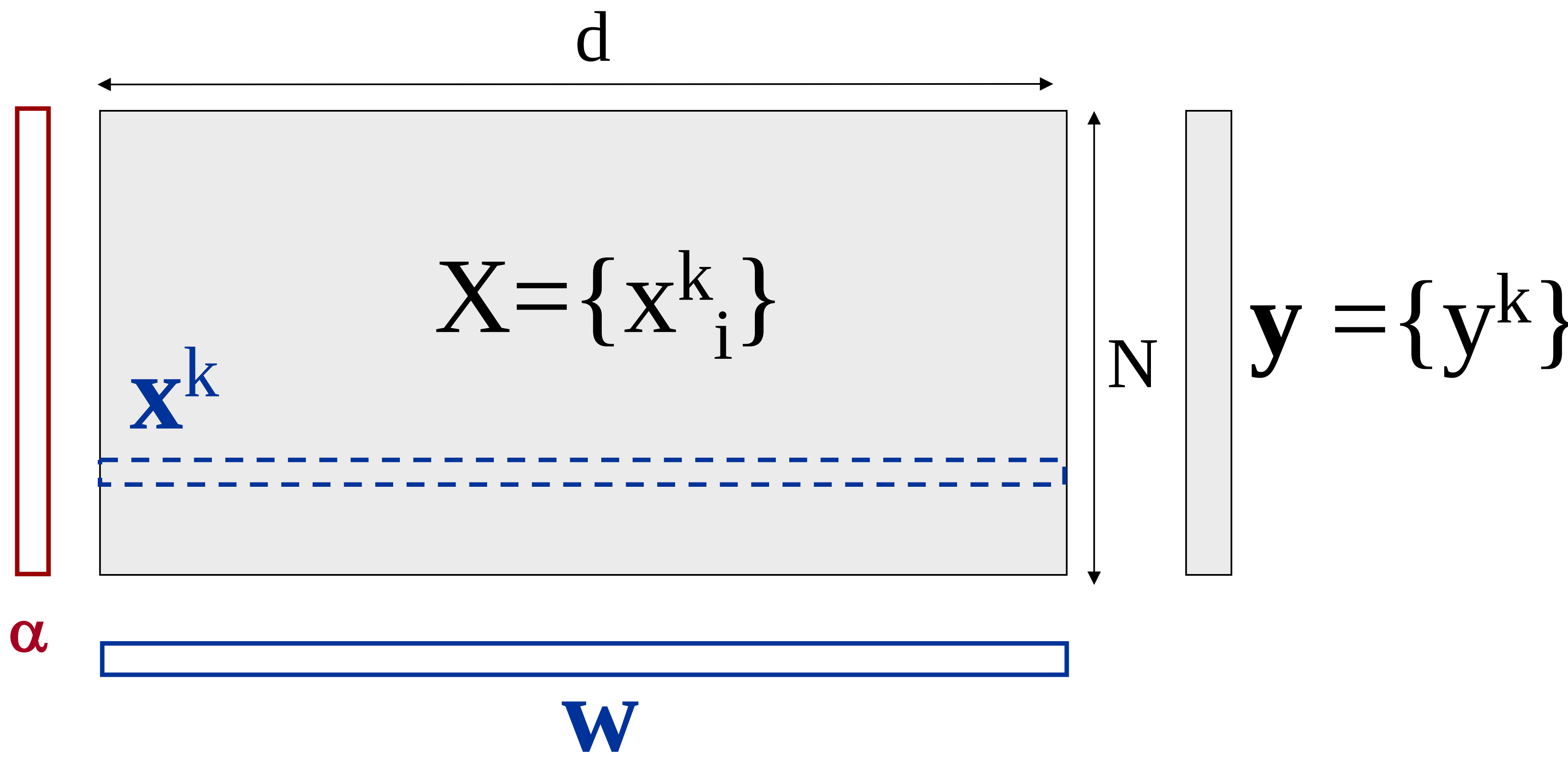
- What do you do at training time?
  - Nothing! The only O(0) algorithm in this class!

- A type of non-parametric method

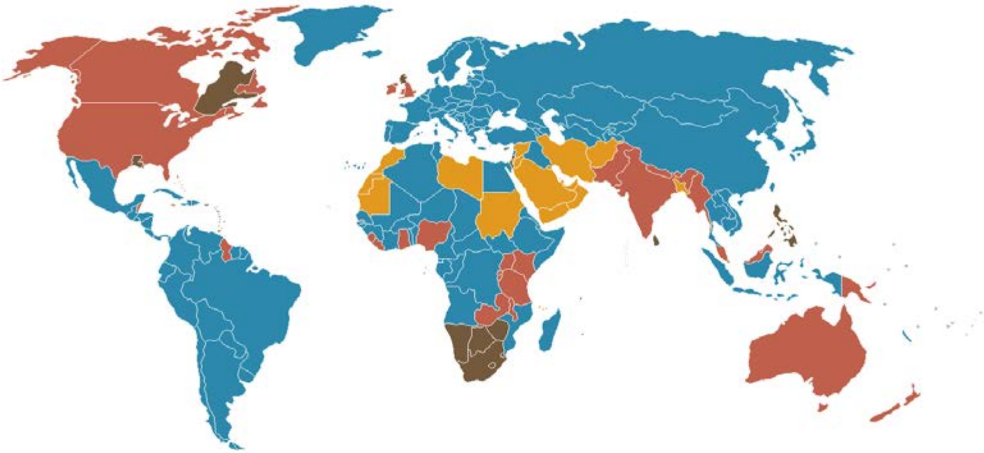
Parametric vs. Non-Parametric

- Parametric: # parameters is independent of N
- Non-parametric: # parameters grows with N

- Studied so far:
  - Logistic Regression?
  - Parzen Windows?
  - Linear Discriminate Analysis?
  - k-Nearest Neighbor
  - Kernel SVM?

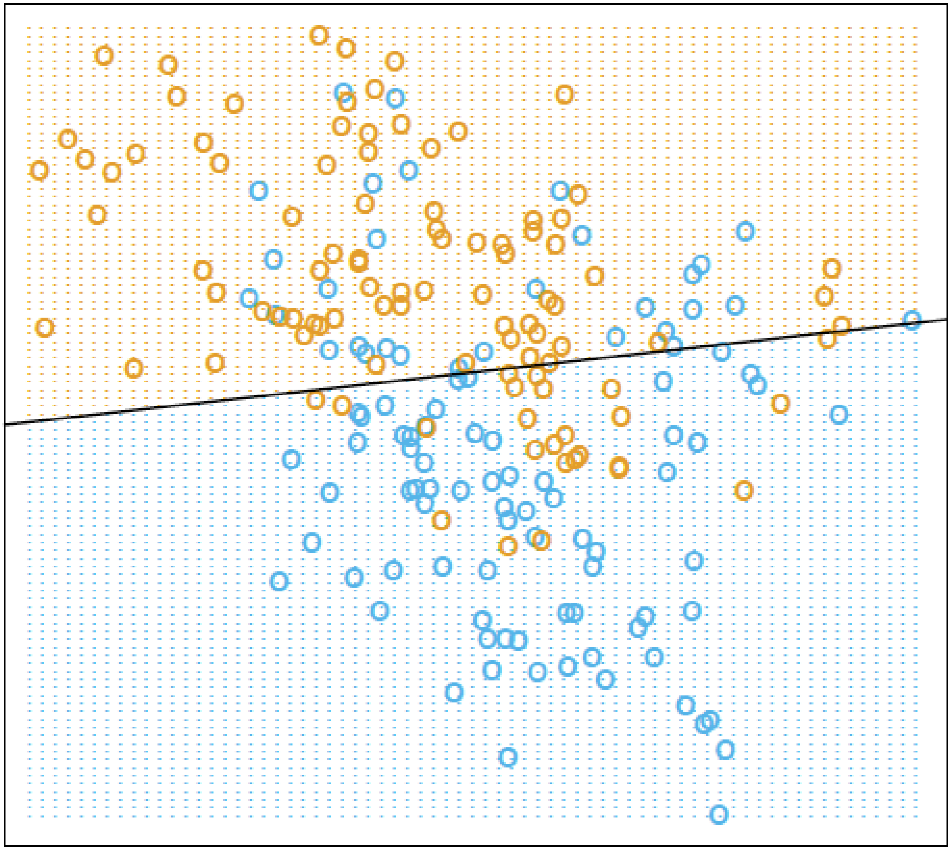

d
X={x^k_i}
x^k
N
y ={y^k}
α
w

Analogy: Legal Systems

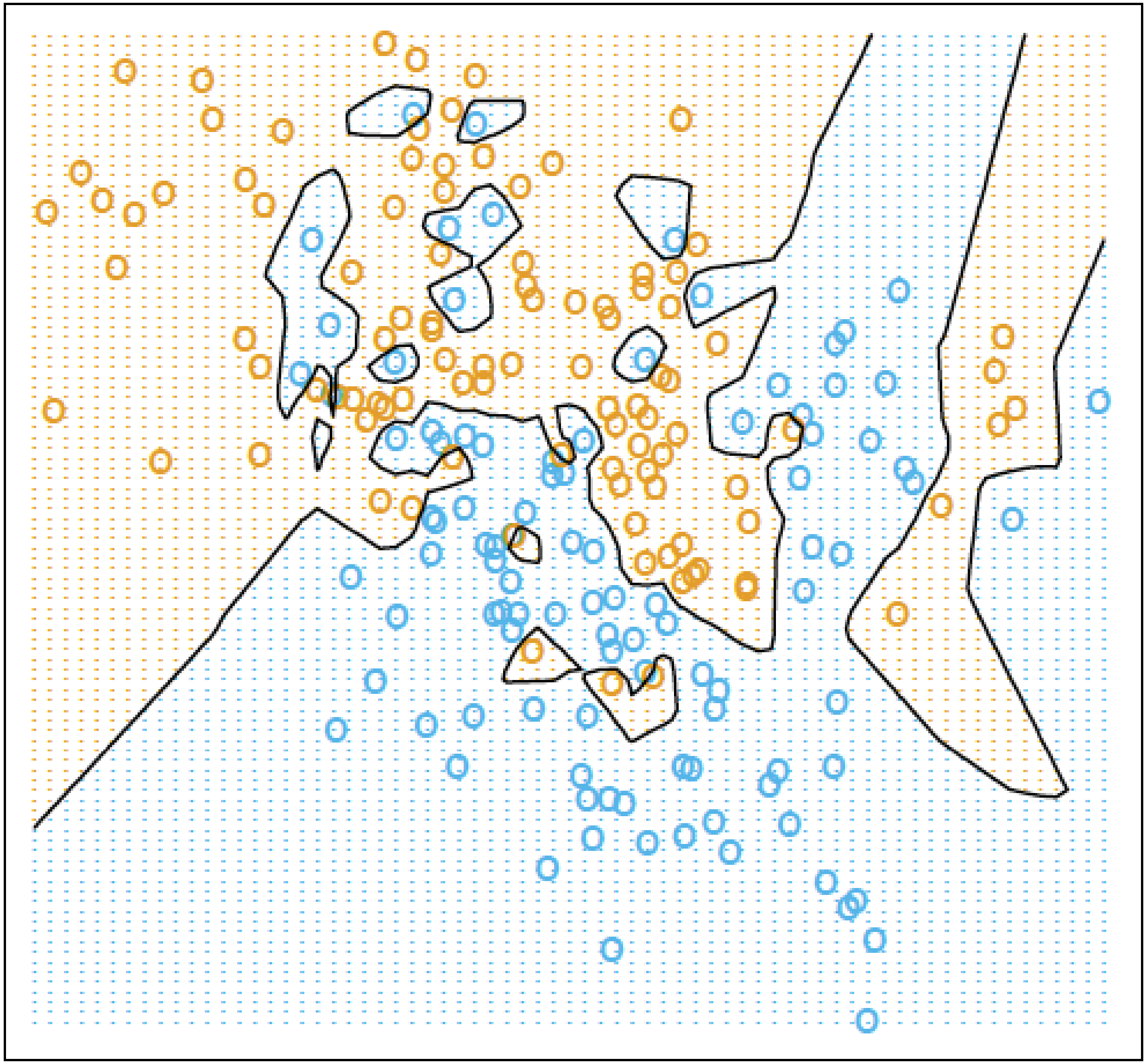


- Civil Law
  - Codified / Continental Law
  - “organized laws that attempt to cover exhaustively the various legal domains”
- Common Law
  - Case / Precedent Law

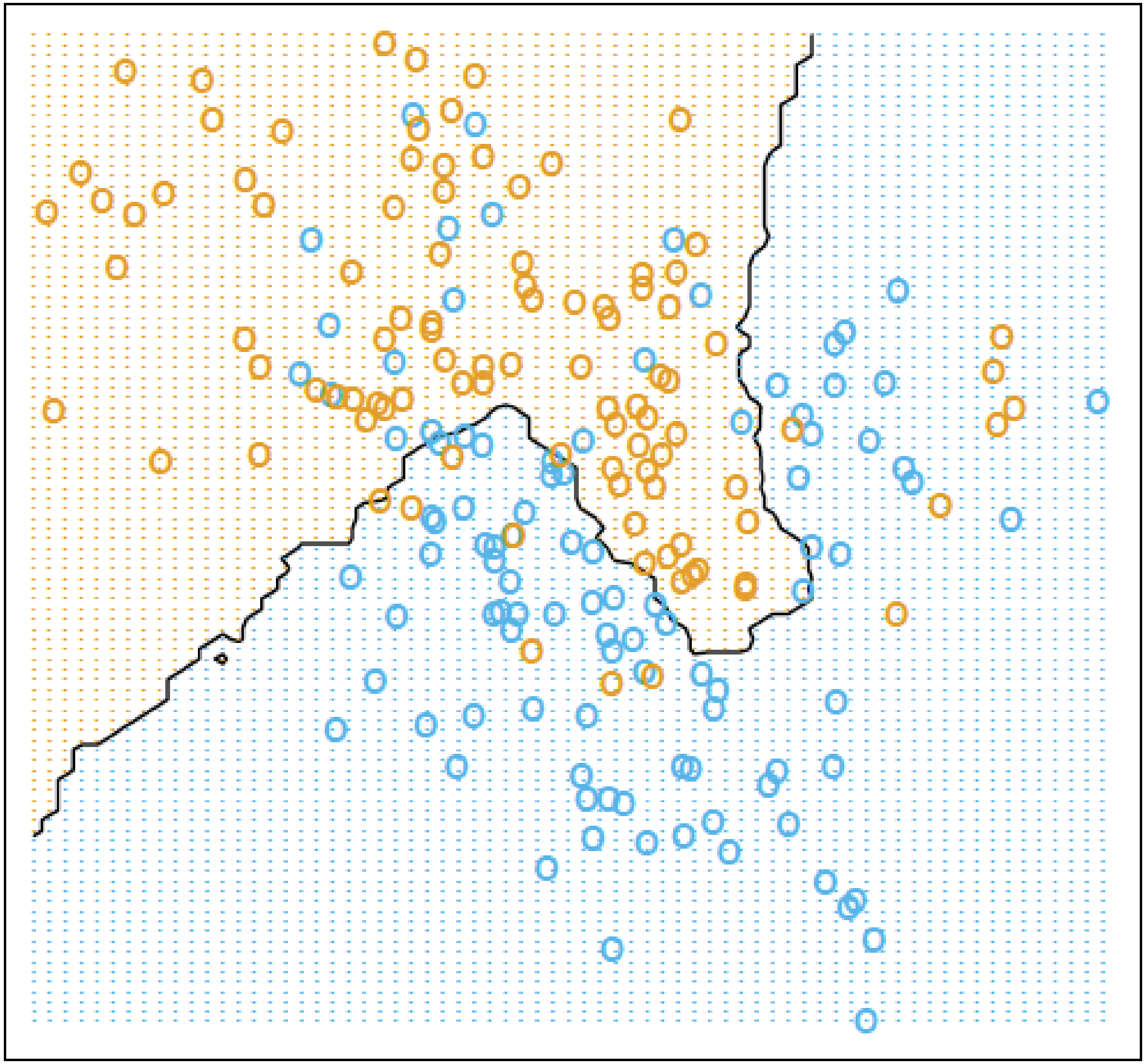
Linear Classifier



1-Nearest Neighbor



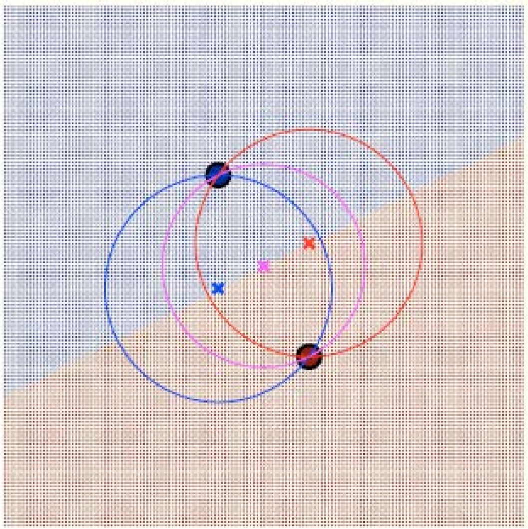
15-Nearest Neighbor



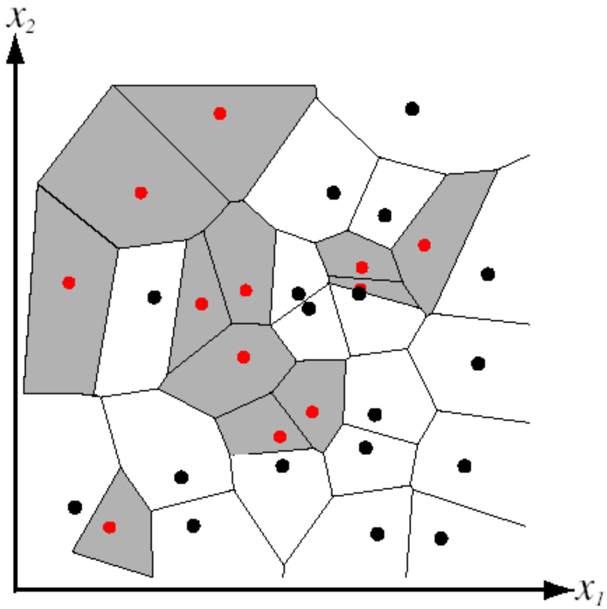
What’s the decision region for NN?

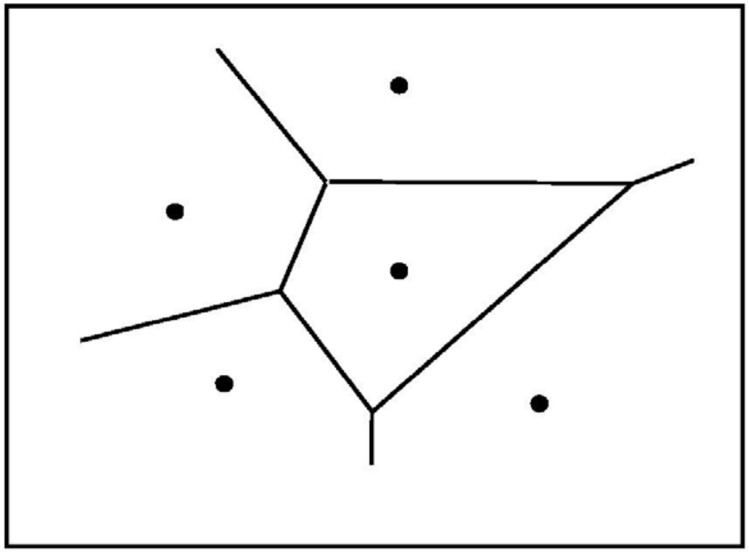
- closest post-office

- all homes with the same post-office (ZIP code)



Voronoi Diagram



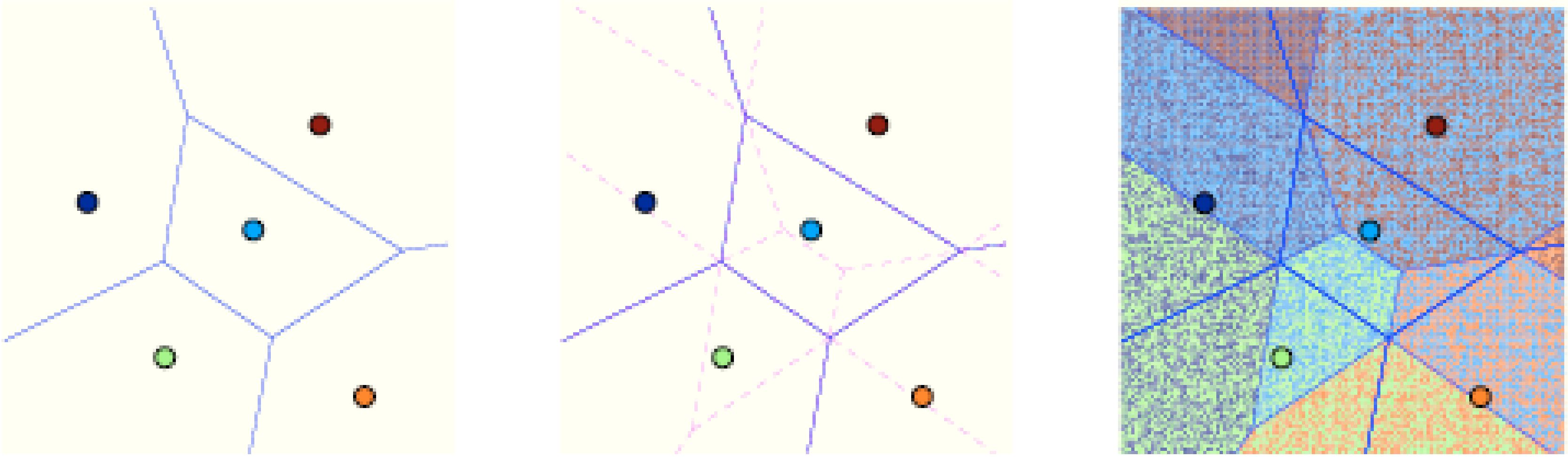


S: Training set

Voronoi cell of $\mathbf{x} \in S$:
All points closer to $\mathbf{x}$ than to any other instance in S

Region of class C:
Union of Voronoi cells of instances of C in S

1 vs. 2 nearest neighbors



(left) Voronoi tesselation for five exemplars. **(middle)** Taking the two nearest exemplars into account leads to a further subdivision of each Voronoi cell. **(right)** The shading indicates which exemplars contribute to which cell.

Behavior at the limit

- Cover and Hart (1967)

$\epsilon^*(\mathbf{x})$: Error of optimal prediction
$\epsilon_{NN}(\mathbf{x})$: Error of nearest neighbor

Theorem: $\lim_{n\rightarrow\infty} \epsilon_{NN} \leq 2\epsilon^*$

- For k-NN:

Theorem: $\lim_{n\rightarrow\infty,\, k\rightarrow\infty,\, k/n\rightarrow 0} \epsilon_{kNN} = \epsilon^*$

Advantages and Disadvantages

Advantages:

- Training is very fast
- Learn complex target functions easily
- Don't lose information

Disadvantages:

- Slow at query time
- Lots of storage
- Easily fooled by irrelevant attributes

“Everything is Miscellaneous”

Weinberger, “**Everything Is Miscellaneous: The Power of the New Digital Disorder**”, 2008

Classes/Categories are Overrated

YAHOO! vs. Google

Delayed/Lazy categorization

Tags: Sky, Water, Beach, Sunny, …
Time: 1pm, August, 2006, …
Location: Italy, Greece, Hawaii …
Photographer: Flickrbug21, Traveller2

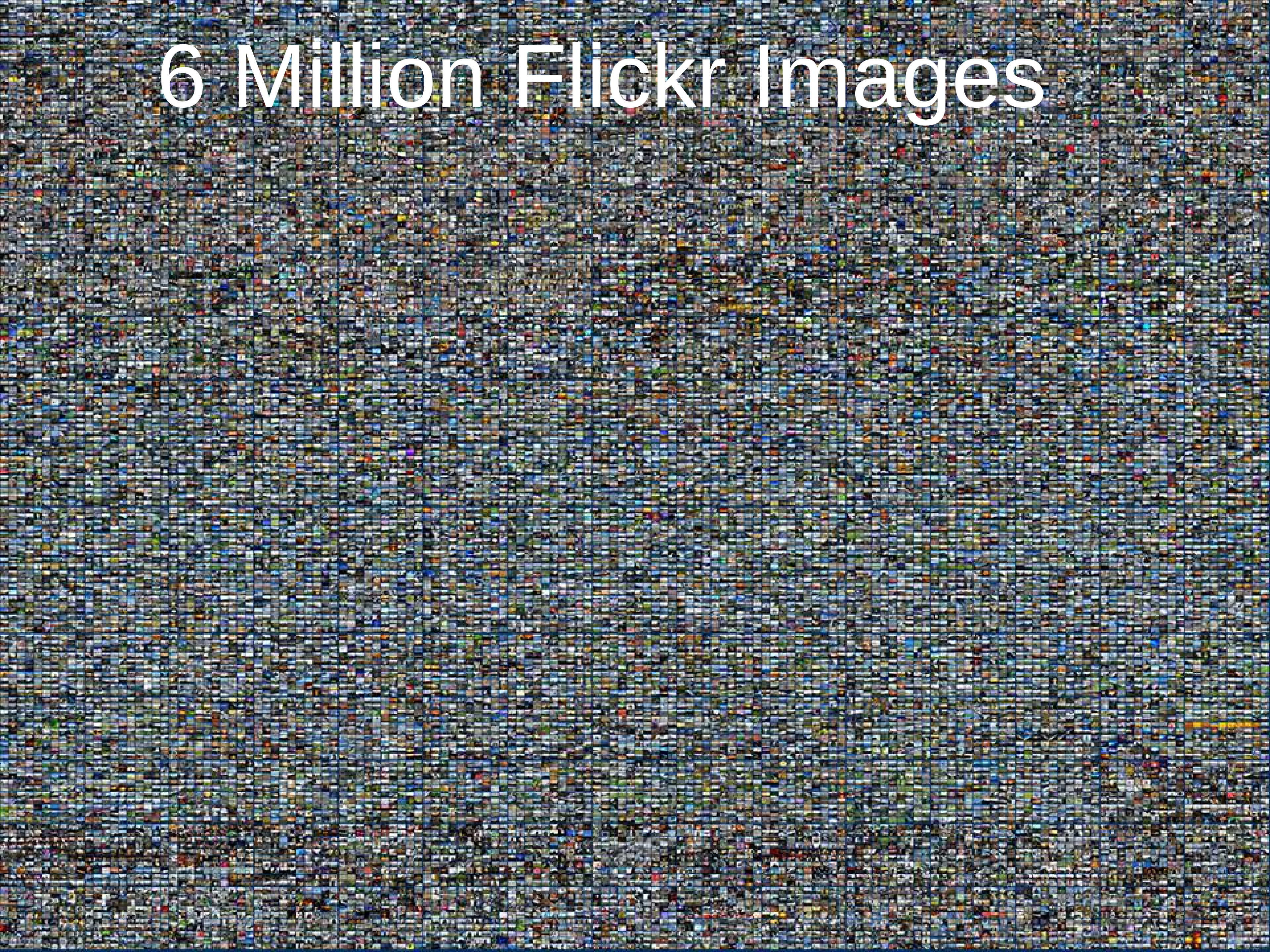
Label Transfer

im2GPS
(using 6 million GPS-tagged Flickr images)

Query Photograph

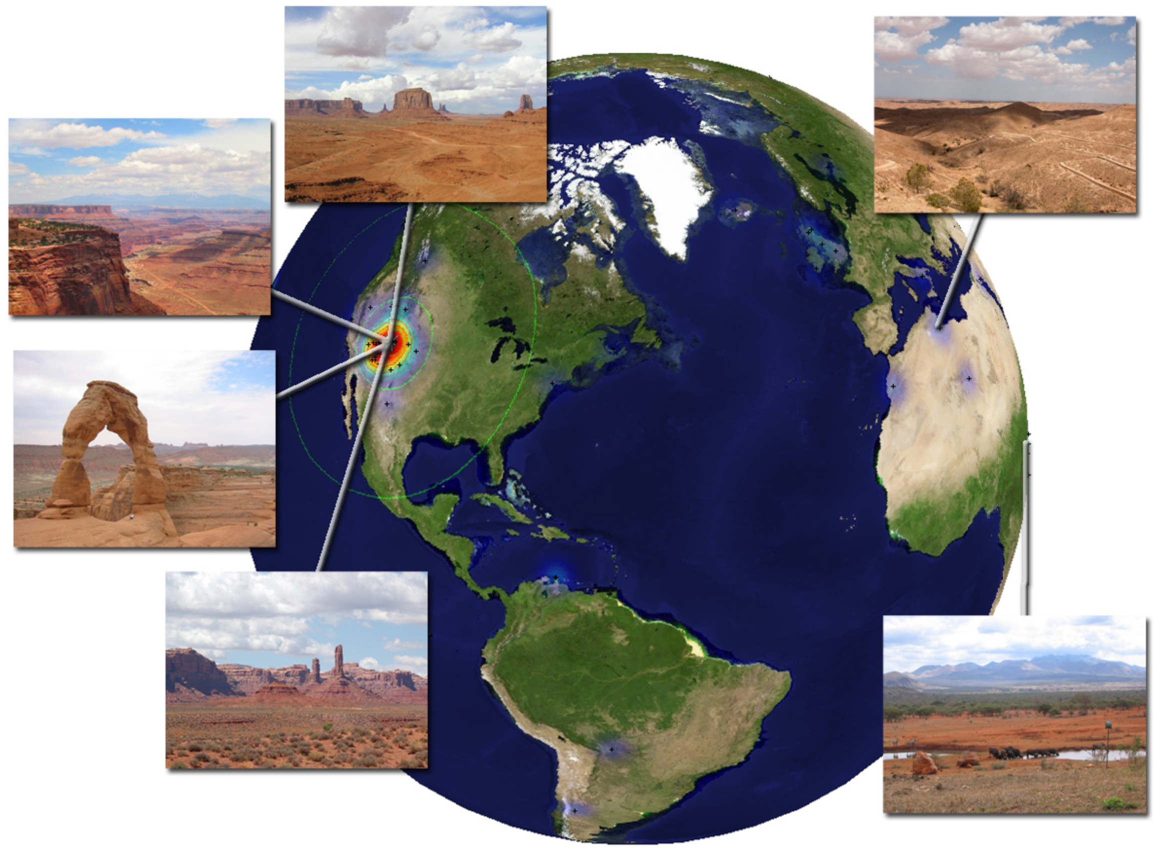
Im2gps [Hays & Efros, CVPR'08]

6 Million Flickr Images



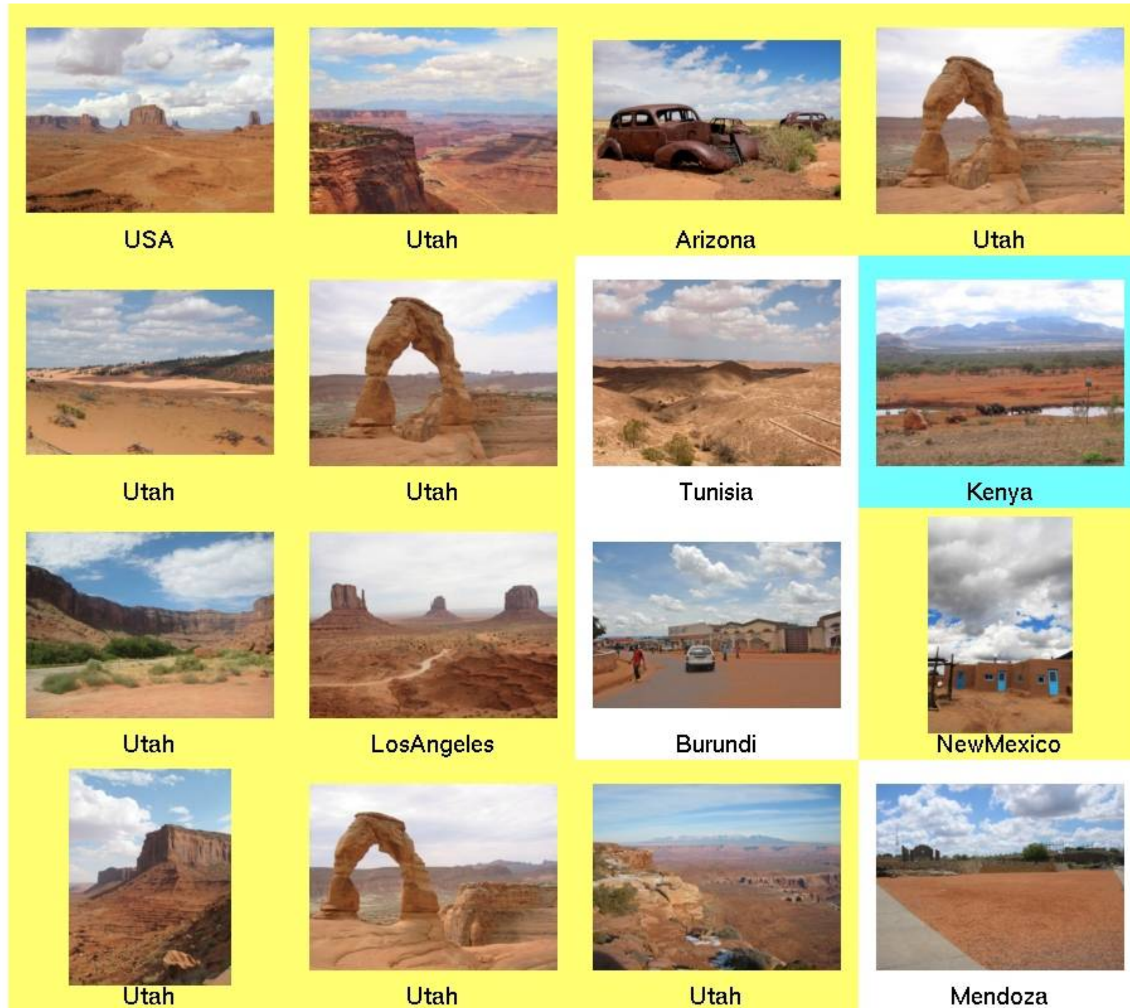
im2GPS
(using 6 million GPS-tagged Flickr images)

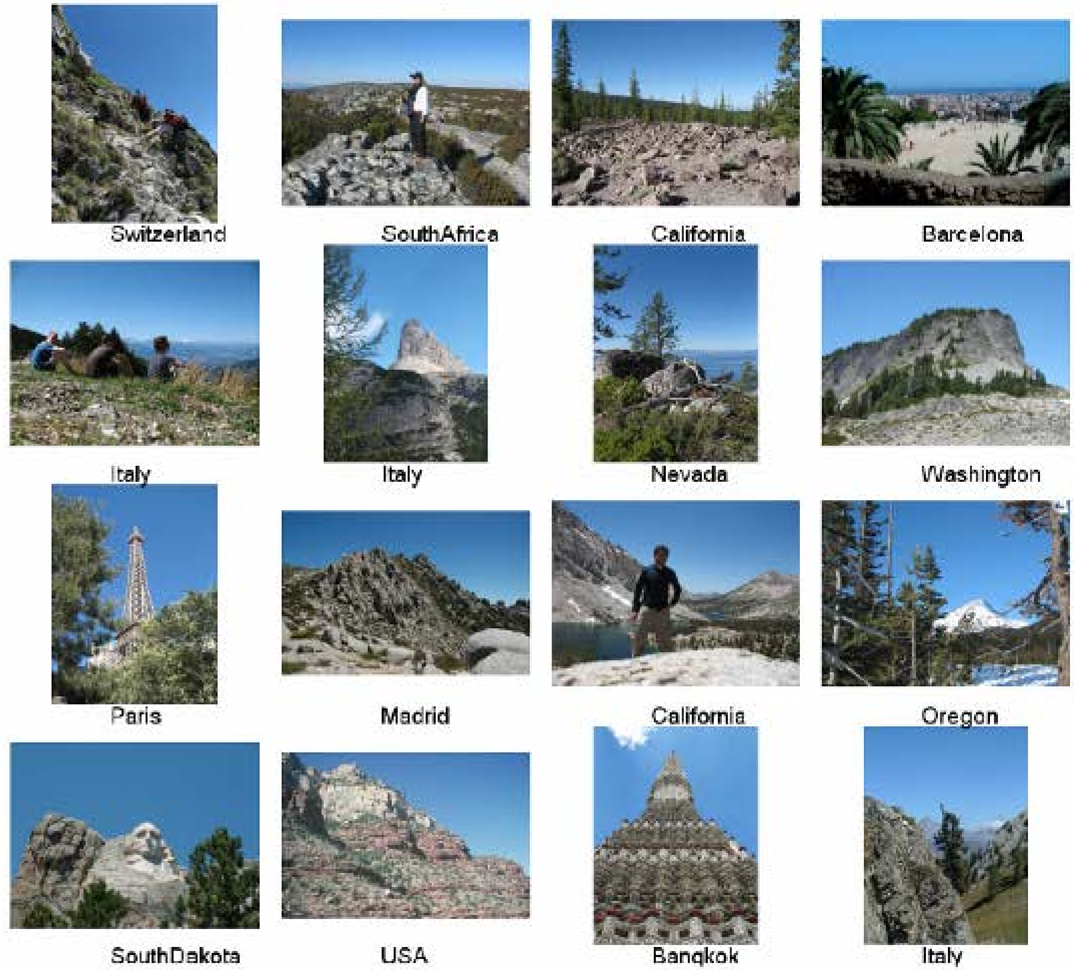
Query Photograph



Visually Similar Scenes

Im2gps [Hays & Efros, CVPR'08]


USA
Utah
Arizona
Utah
Utah
Utah
Tunisia
Kenya
Utah
LosAngeles
Burundi
NewMexico
Utah
Utah
Utah
Mendoza

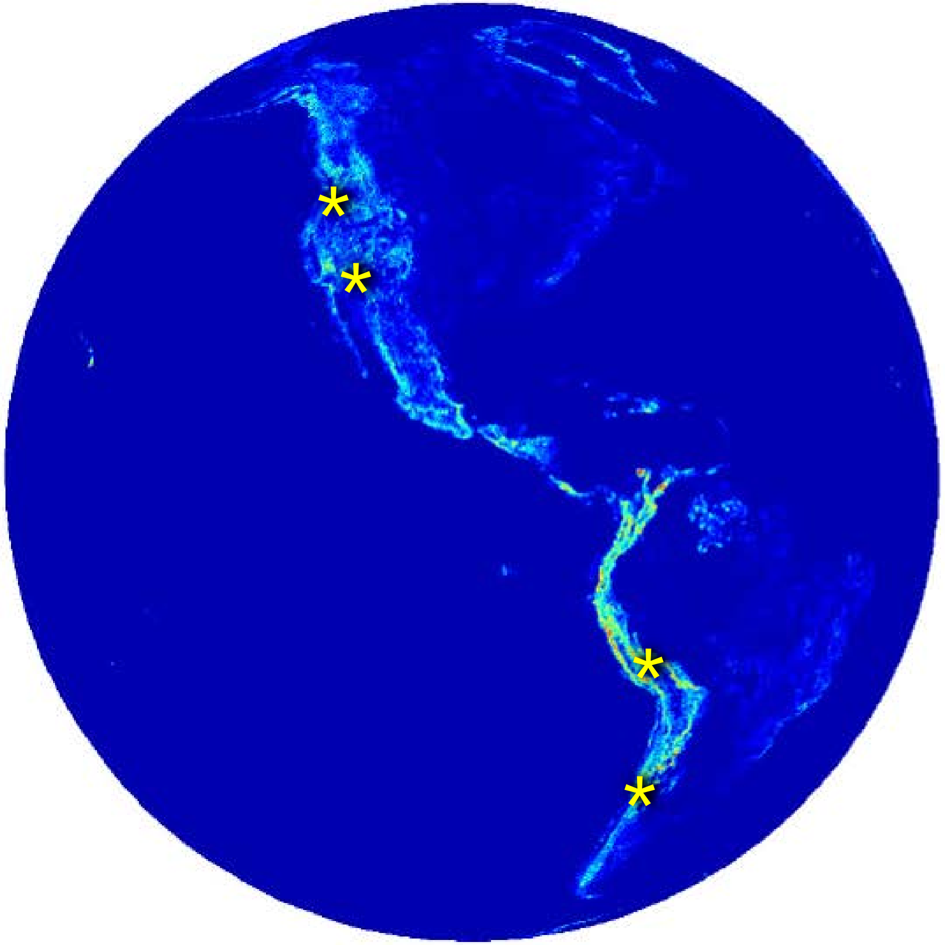


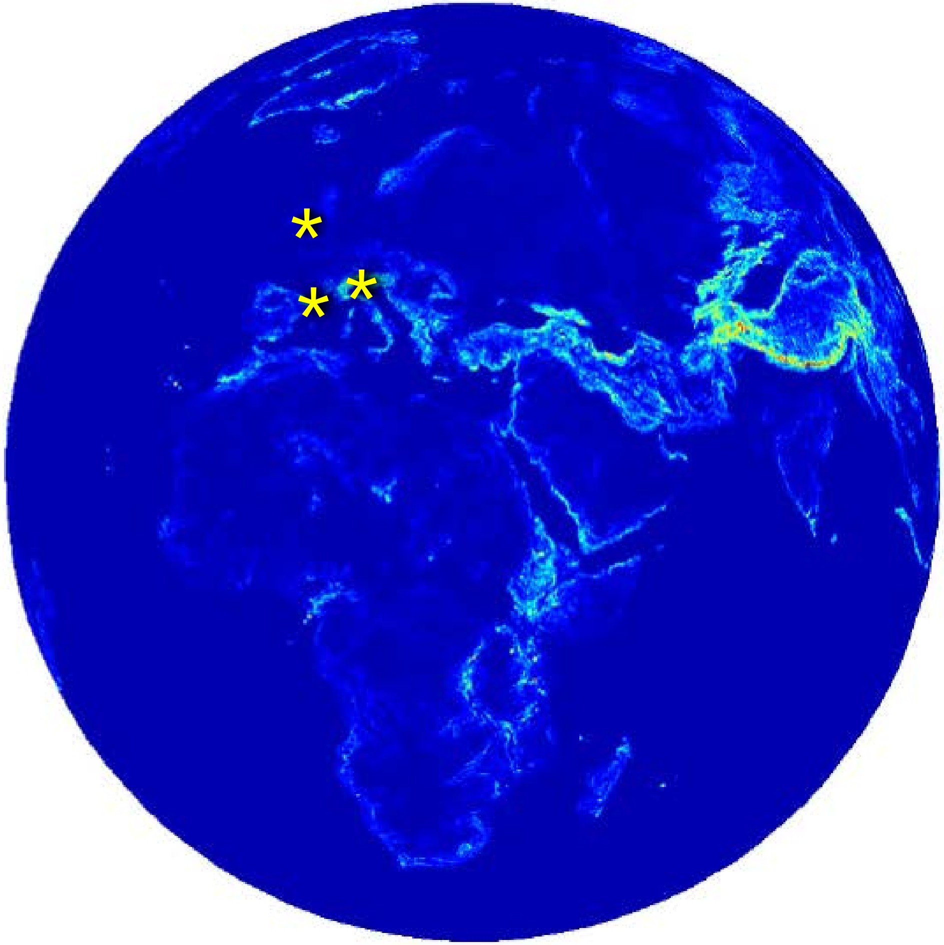
Switzerland
SouthAfrica
California
Barcelona
Italy
Italy
Nevada
Washington
Paris
Madrid
California
Oregon
SouthDakota
USA
Bangkok
Italy

Lazy label transfer

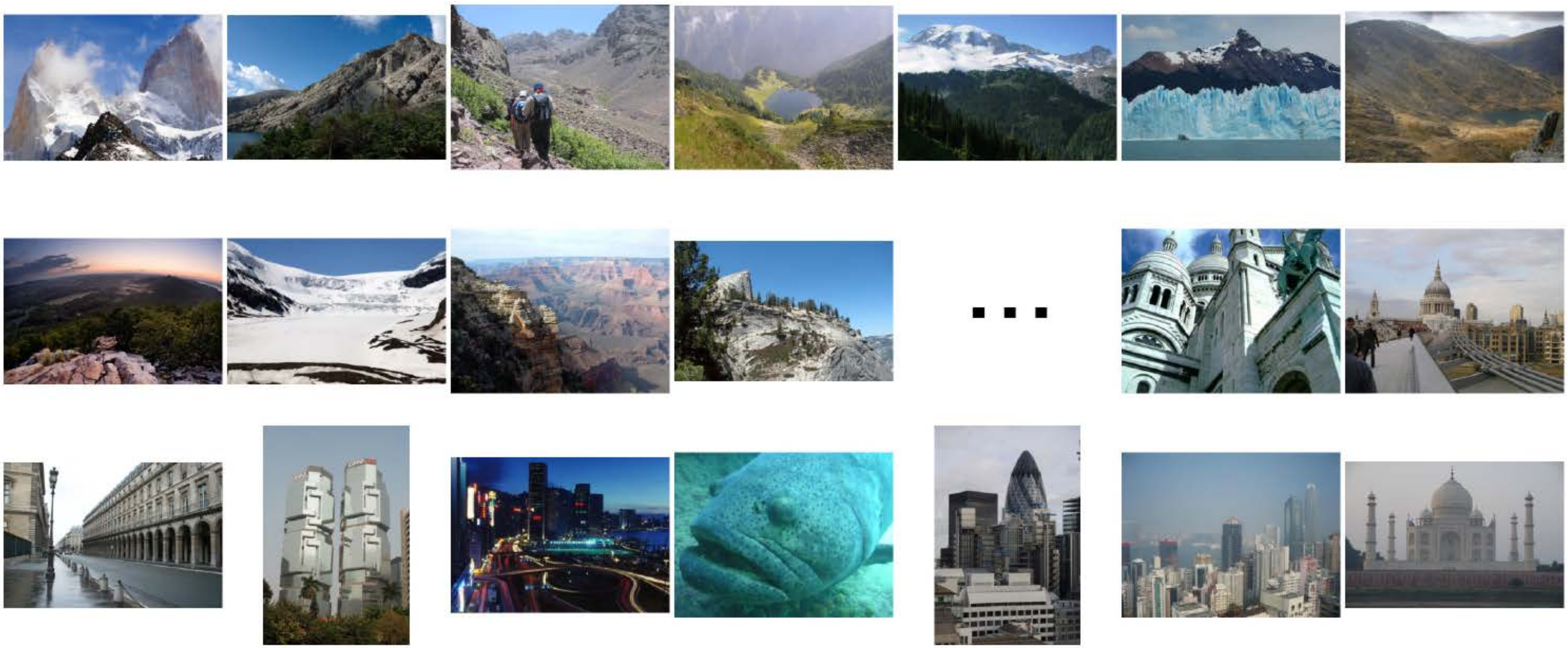

Argentina
Andorra
Andorra
Iceland
Idaho
Switzerland
Argentina
Bolivia
Nevada
Hawaii
Hawaii
Egypt
China
Arizona
Peru
Oregon

Elevation gradient = 112 m / km





Elevation gradient magnitude ranking



Population density map

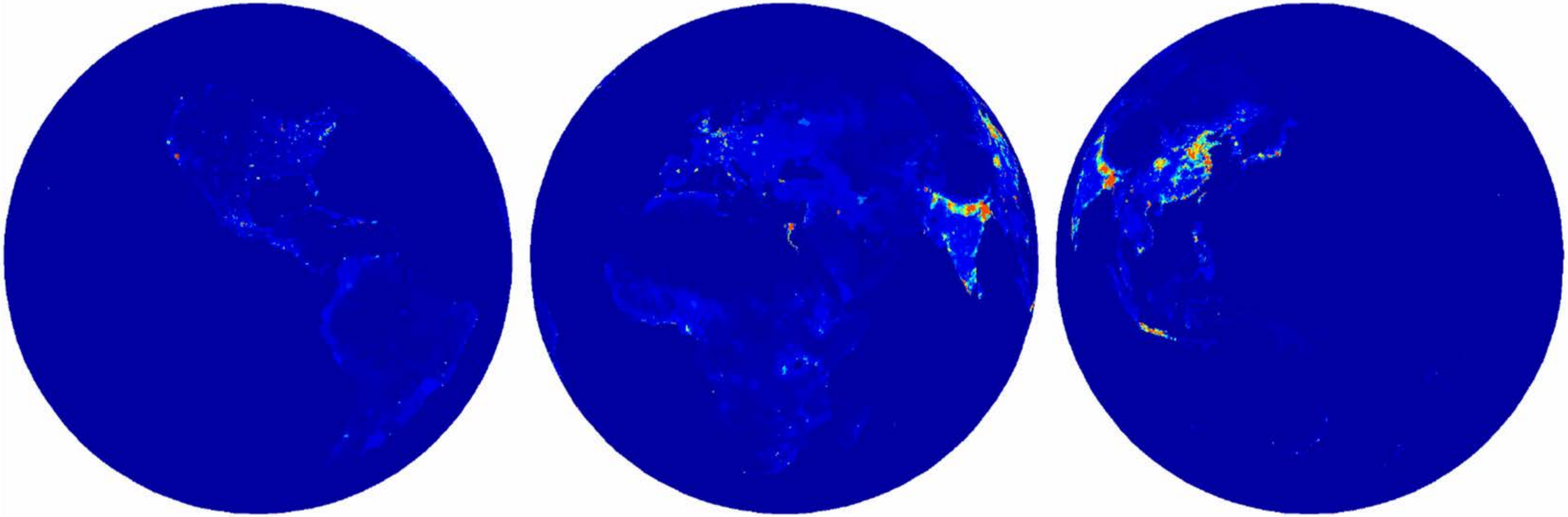
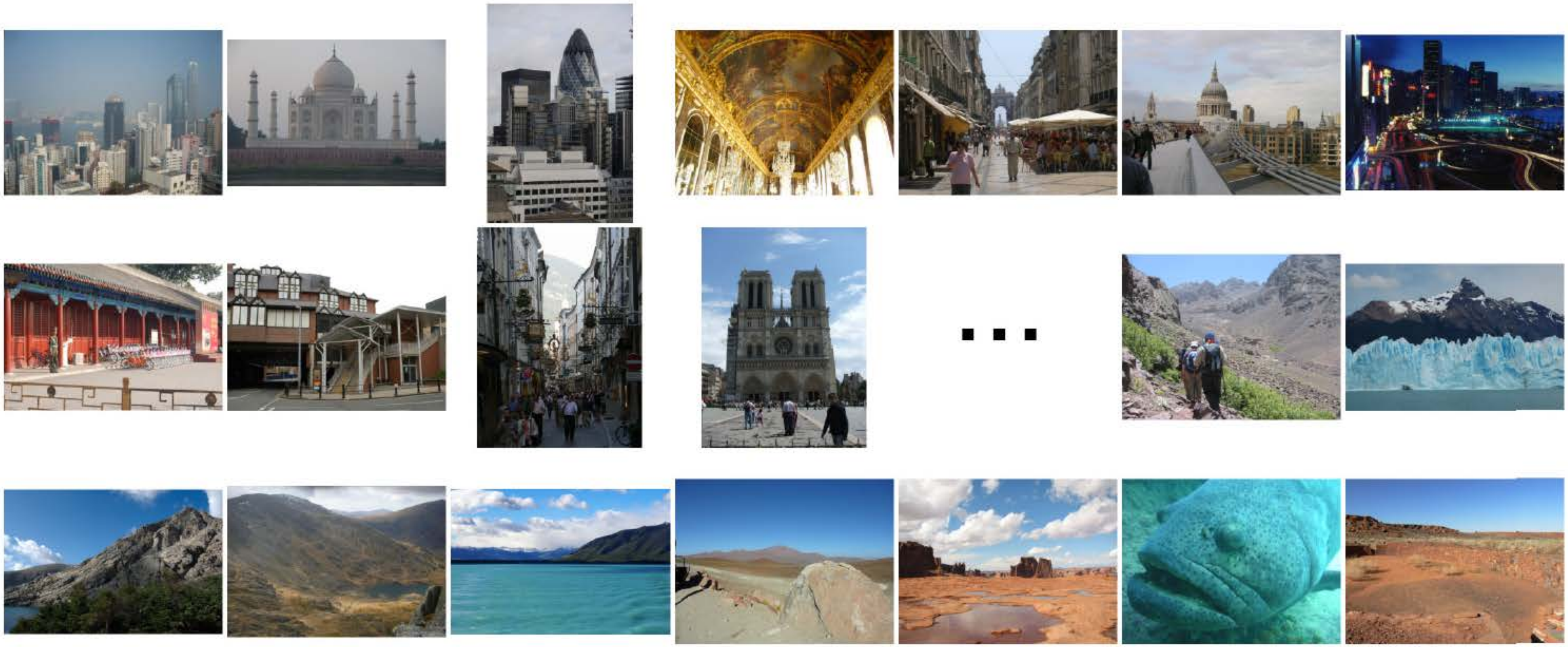


Figure 2. Global population density map.

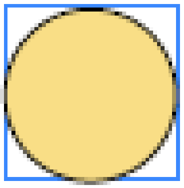
Population density ranking



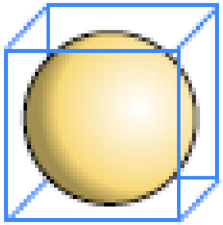
Curse of Dimensionality

- Nearest Neighbor is easily mislead in high dimensions
- Easy problems in low-D are hard in High-D
  - “If we could see in high dimension, there would be no need for Machine Learning”
- Low-D intuitions don’t apply to High D
  - Everything is far from everything else
- Examples:
  - Points on hyper-grid
  - Hypersphere
  - Hypercube vs. hypersphere
  - High-dimensional Gaussian

Hypersphere / Hypercube ratio



d = 2



d = 3