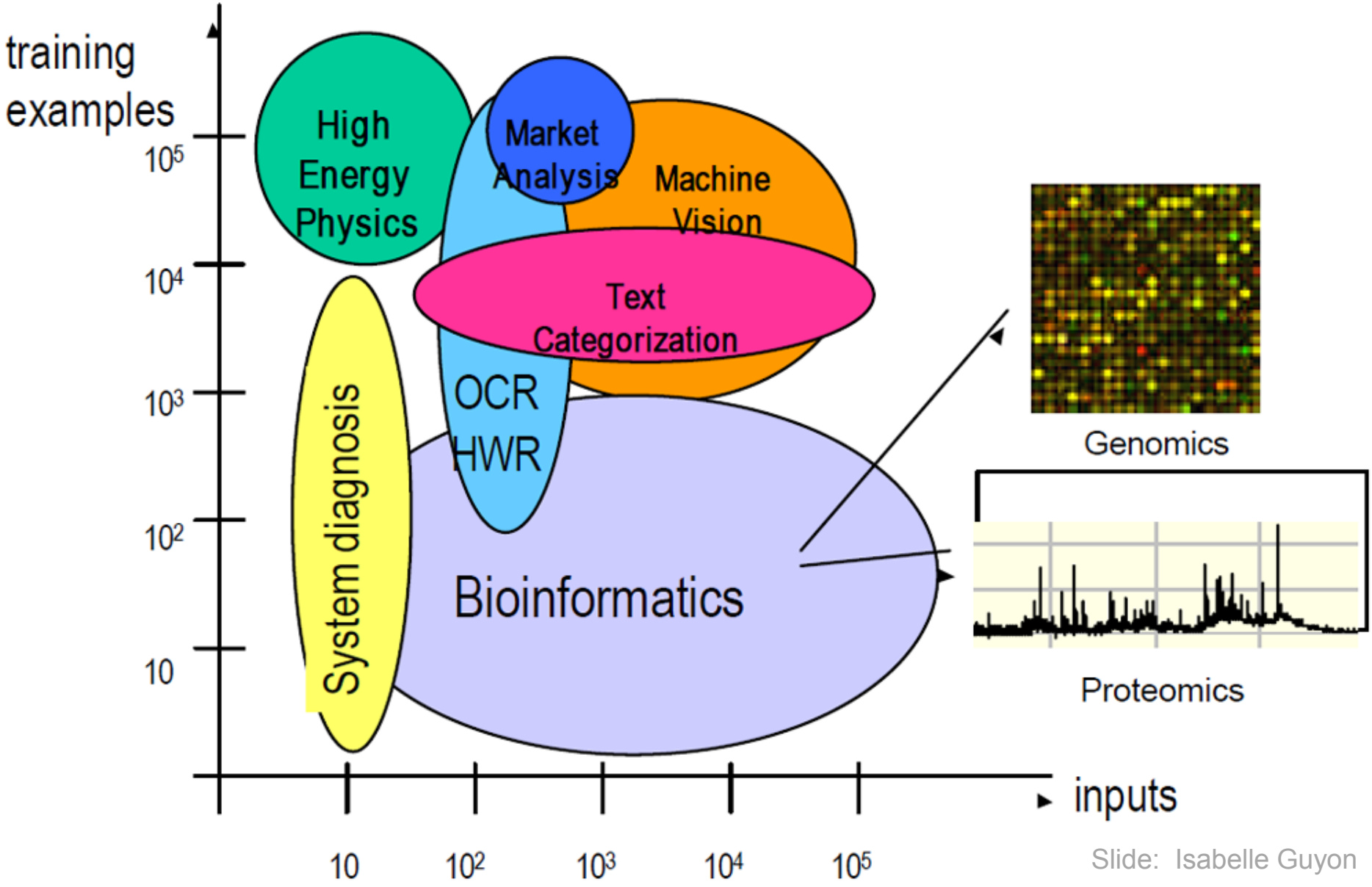
Machine learning: Overview

- Core of ML: Making predictions or decisions from Data.

Impact of Machine Learning

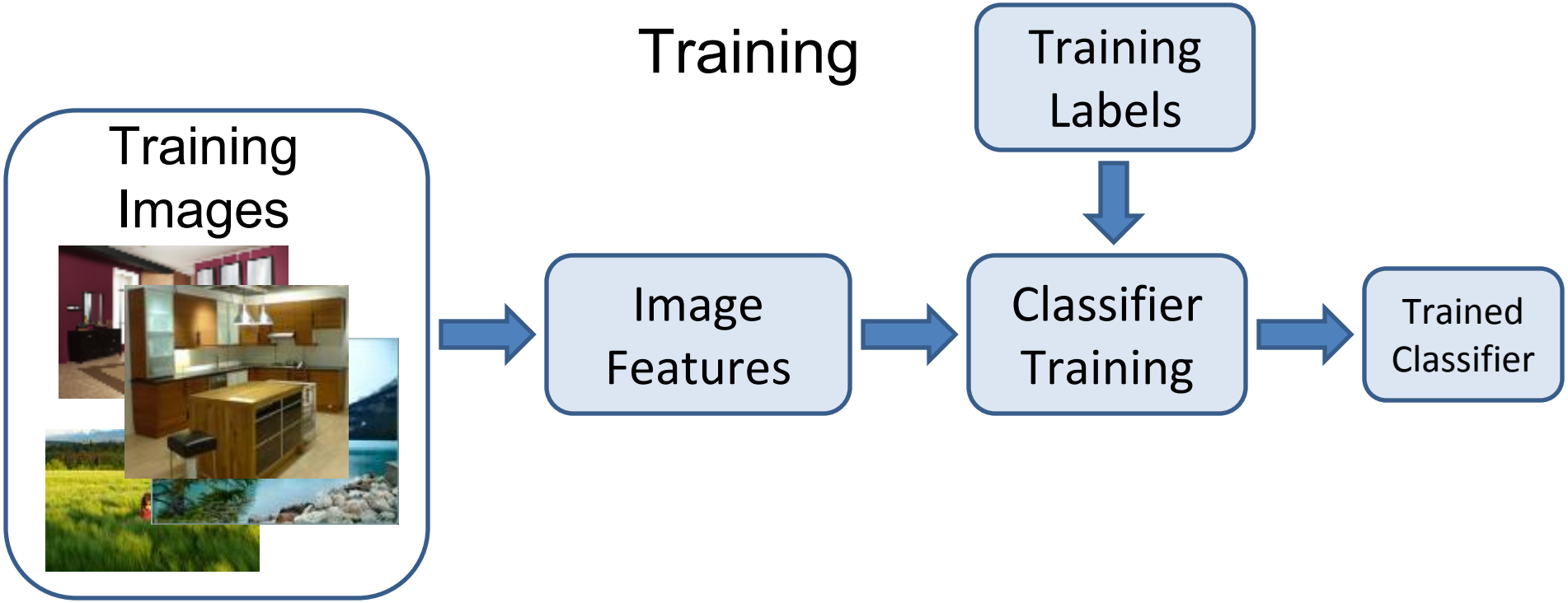
- Machine Learning is arguably the greatest export from computing to other scientific fields.

Machine Learning Applications



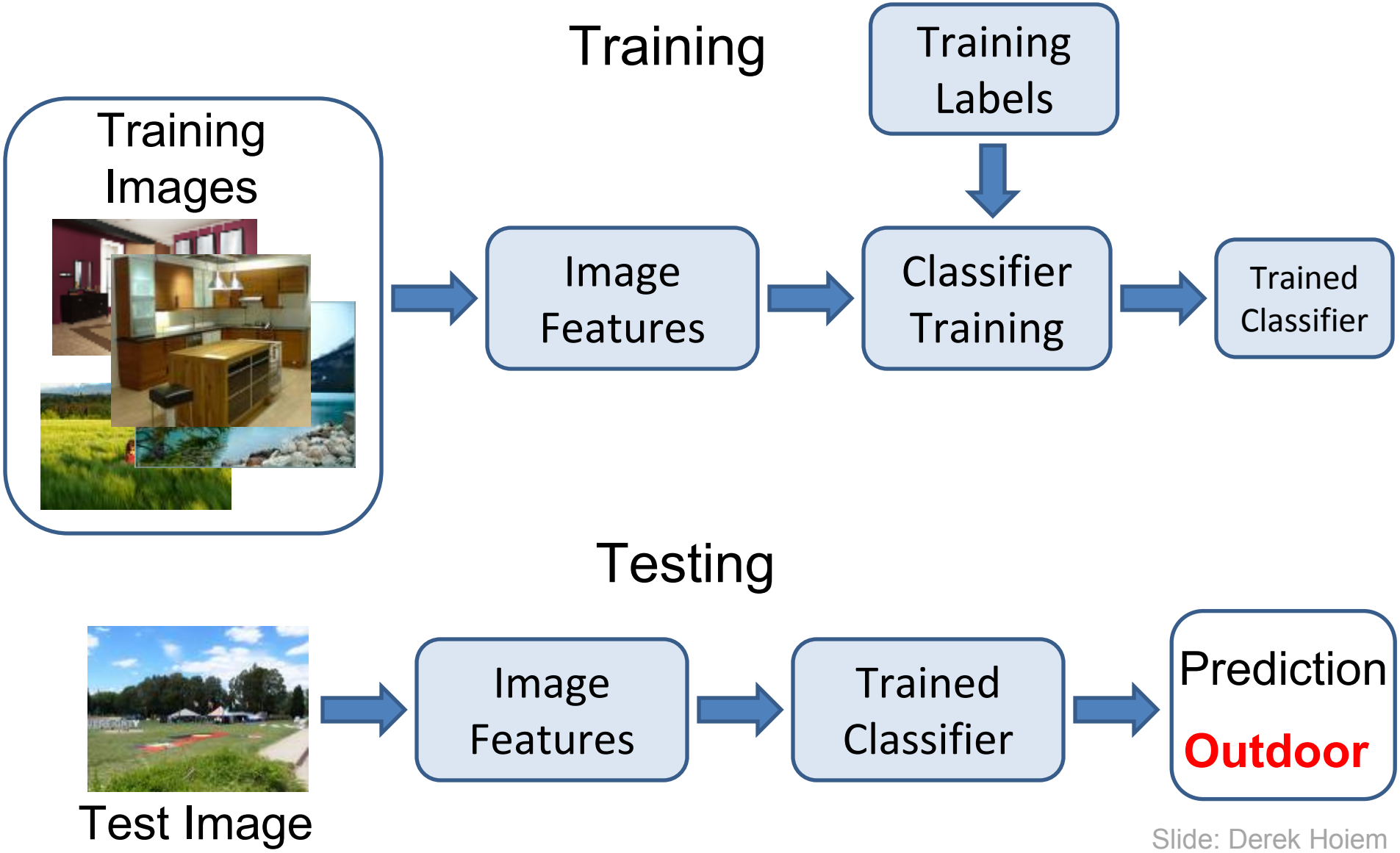
Slide: Isabelle Guyon

Image Categorization



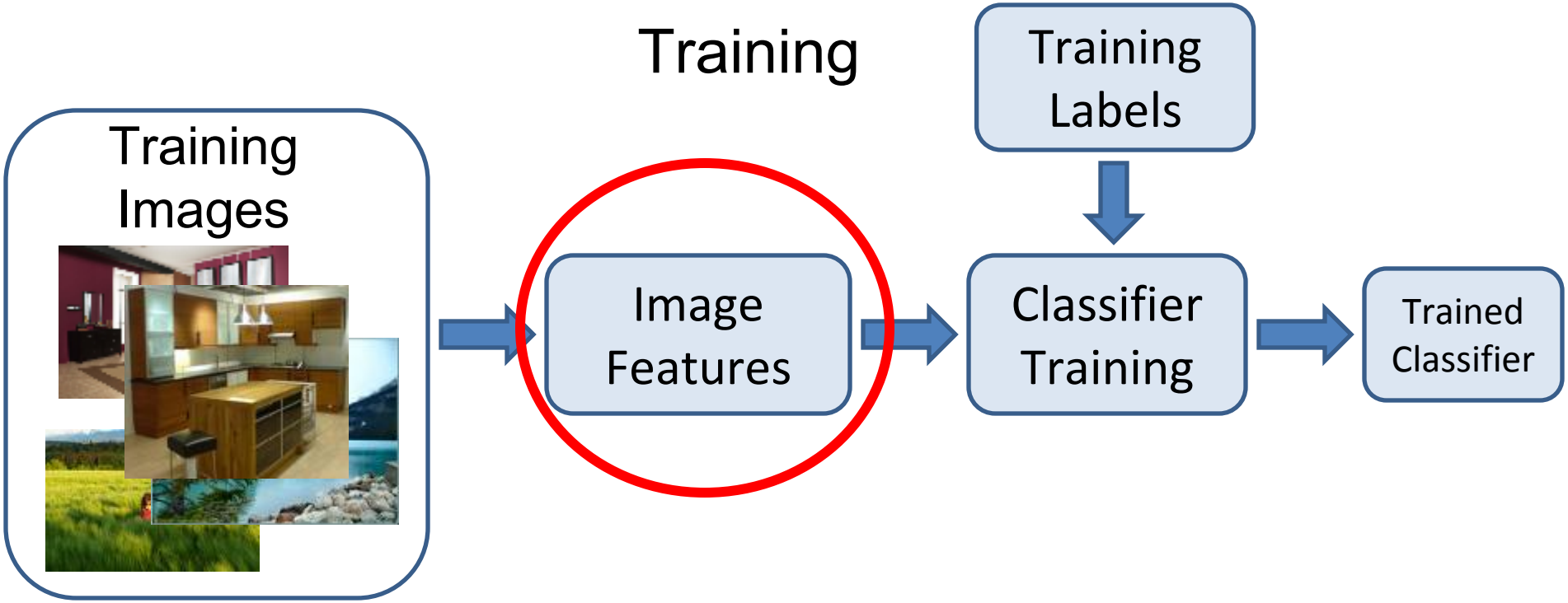
Slide: Derek Hoiem

Image Categorization



Slide: Derek Hoiem

Image features



Slide: Derek Hoiem

General Principles of Representation

- Coverage
  - Ensure that all relevant info is captured
- Concision
  - Minimize number of features without sacrificing coverage
- Directness
  - Ideal features are independently useful for prediction

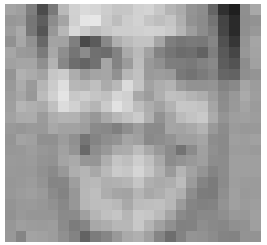
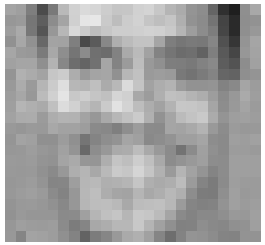


Image Intensity

Slide: Derek Hoiem

Image representations

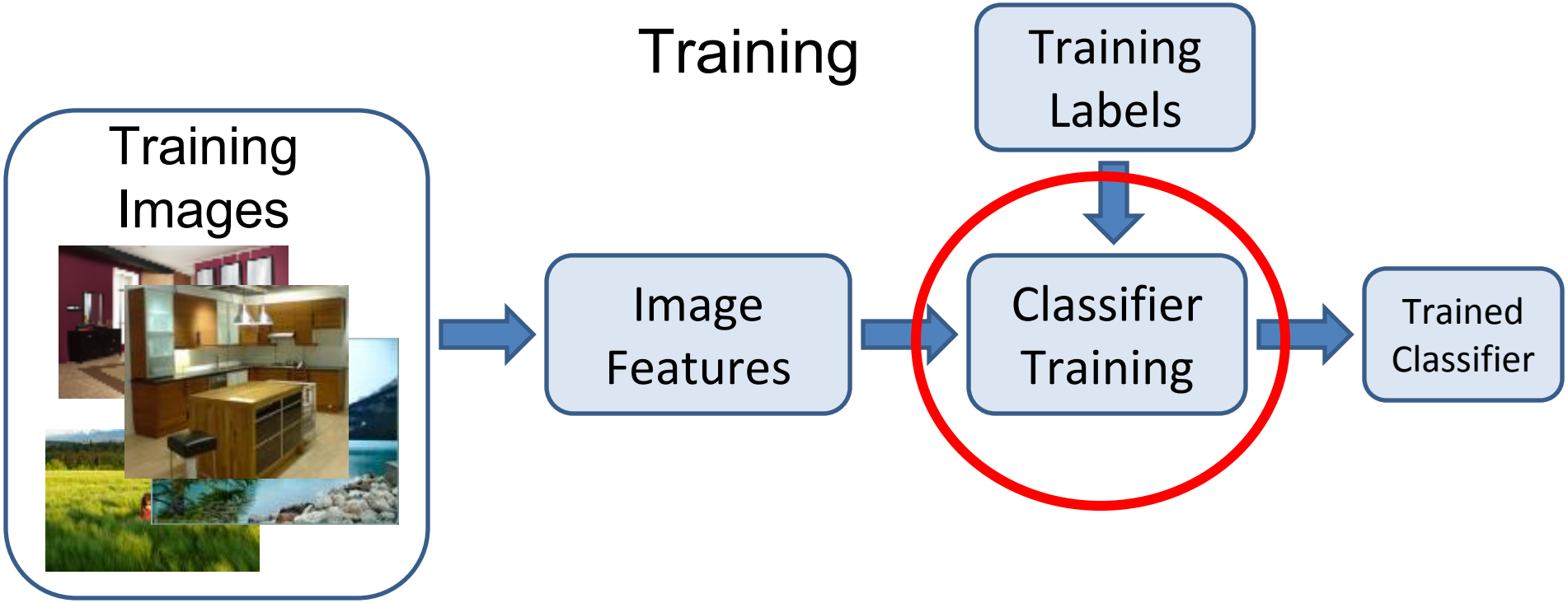
- Templates
  - Intensity, gradients, etc.



- Histograms
  - Color, texture, SIFT descriptors, etc.

Slide: Derek Hoiem

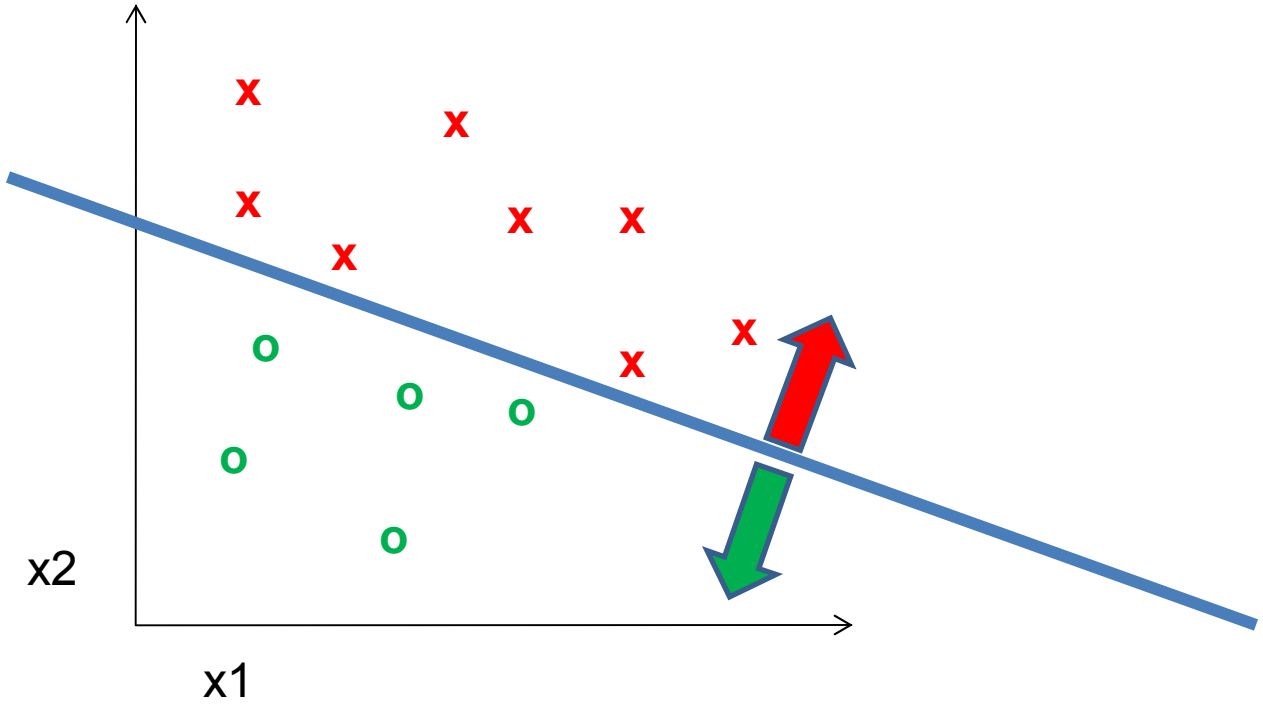
Classifiers



Slide: Derek Hoiem

Learning a classifier

Given some set of features with corresponding labels, learn a function to predict the labels from the features



Slide: Derek Hoiem

One way to think about it…

- Training labels dictate that two examples are the same or different, in some sense
- Features and distance measures define visual similarity
- Classifiers try to learn weights or parameters for features and distance measures so that visual similarity predicts label similarity

Slide: Derek Hoiem

Machine Learning Problems

	Supervised Learning	*Unsupervised Learning*
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

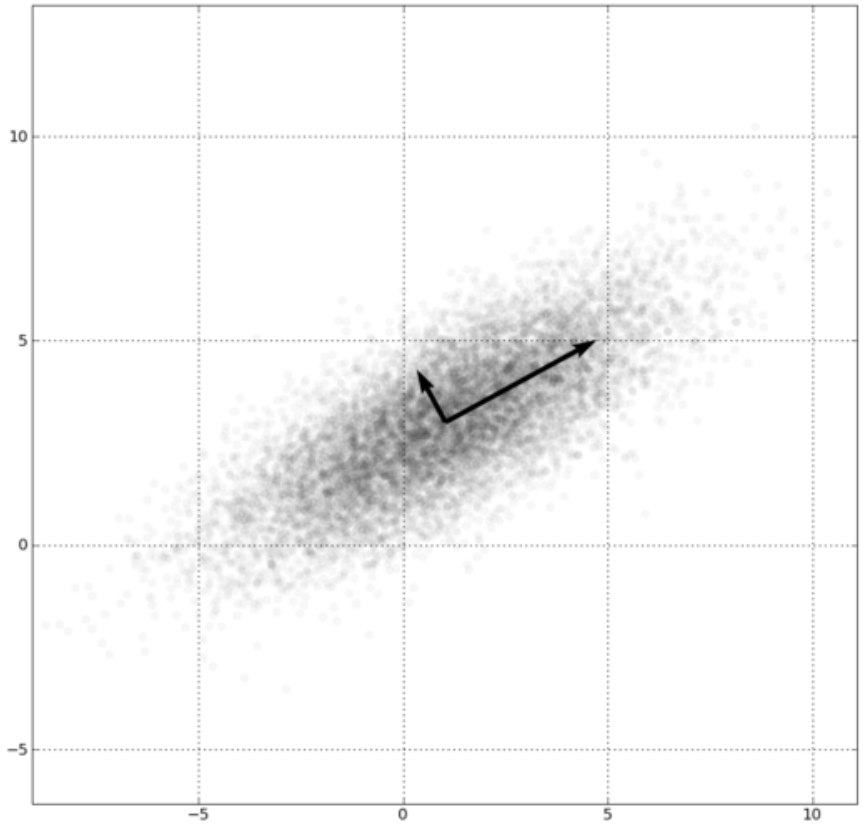
Slide: Erik Sudderth

Dimensionality Reduction

- **PCA**, ICA, LDA, Isomap

- PCA is the most important technique to know. It takes advantage of correlations in data dimensions to produce the best possible lower dimensional representation, according to reconstruction error.

- PCA should be used for dimensionality reduction, not for discovering patterns or making predictions. Don't try to assign semantic meaning to the bases.



Many classifiers to choose from

- SVM
- Neural networks
- Naïve Bayes
- Bayesian network
- Logistic regression
- Randomized Forests
- Boosted Decision Trees
- K-nearest neighbor
- RBMs
- Etc.

Which is the best one?

Slide: Derek Hoiem