

CS 445/545
Machine Learning
Winter, 2014

See syllabus at

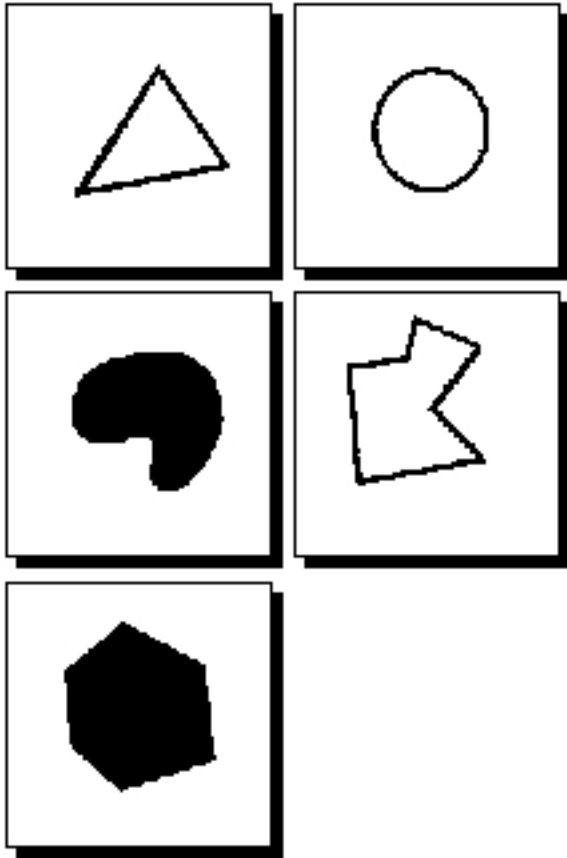
<http://web.cecs.pdx.edu/~mm/MachineLearningWinter2016/>

Lecture slides will be posted on this website before each class.

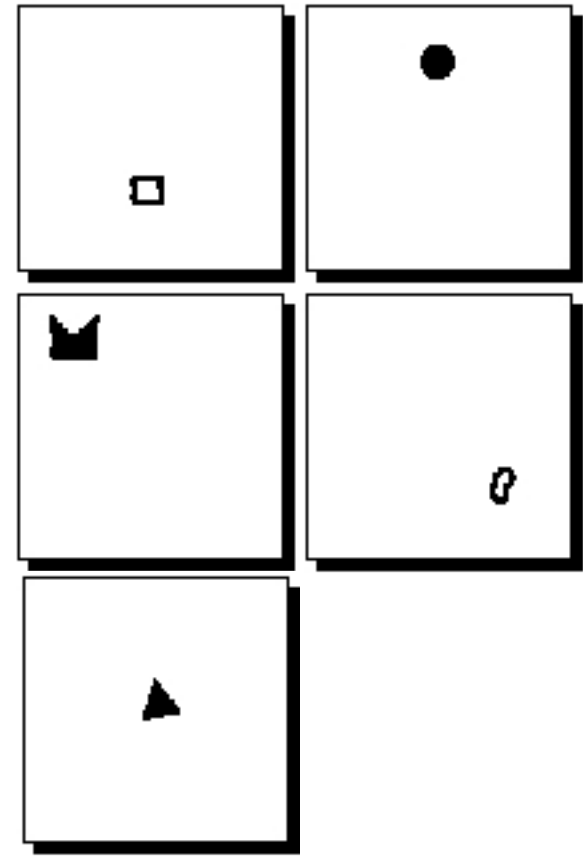
What is machine learning?

- Textbook definitions of “machine learning”:
 - Detecting patterns and regularities with a good and generalizable approximation (“model” or “hypothesis”)
 - Execution of a computer program to optimize the parameters of the model using training data or past experience.

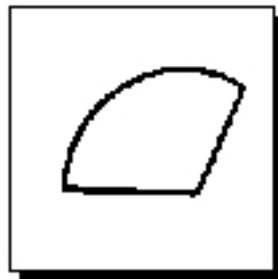
Training Examples: Class 1



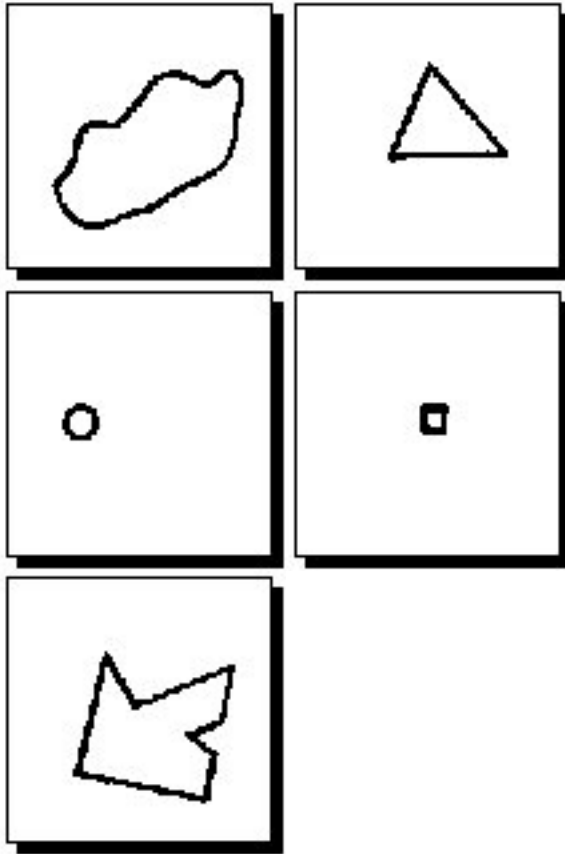
Training Examples: Class 2



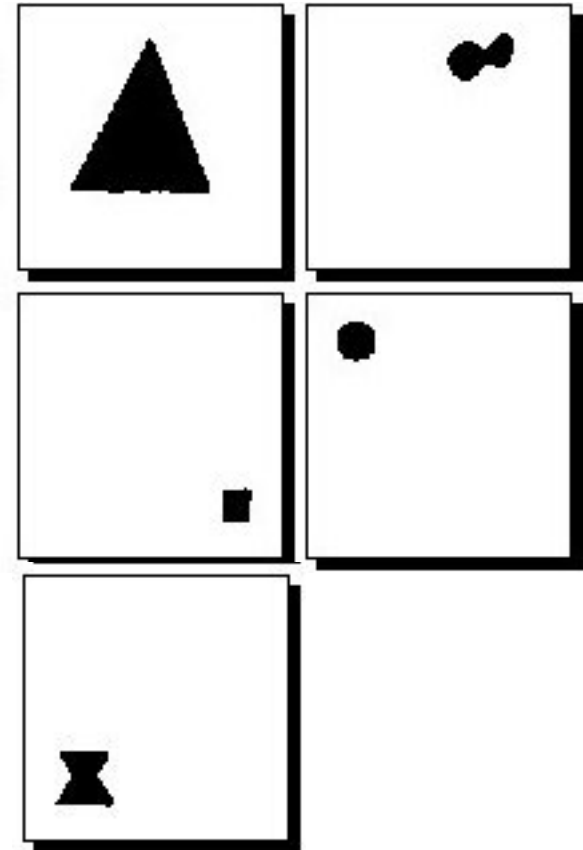
Test example: Class = ?



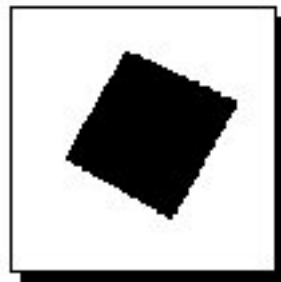
Training Examples: Class 1



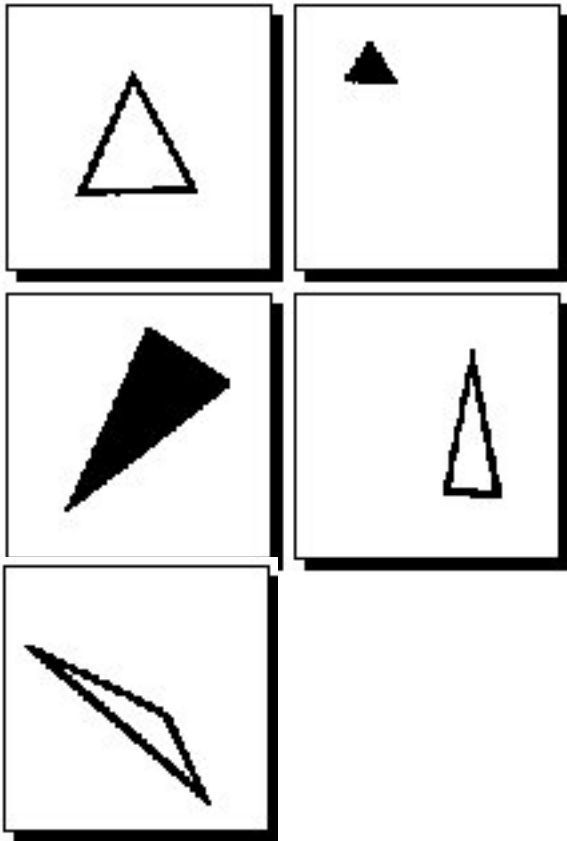
Training Examples: Class 2



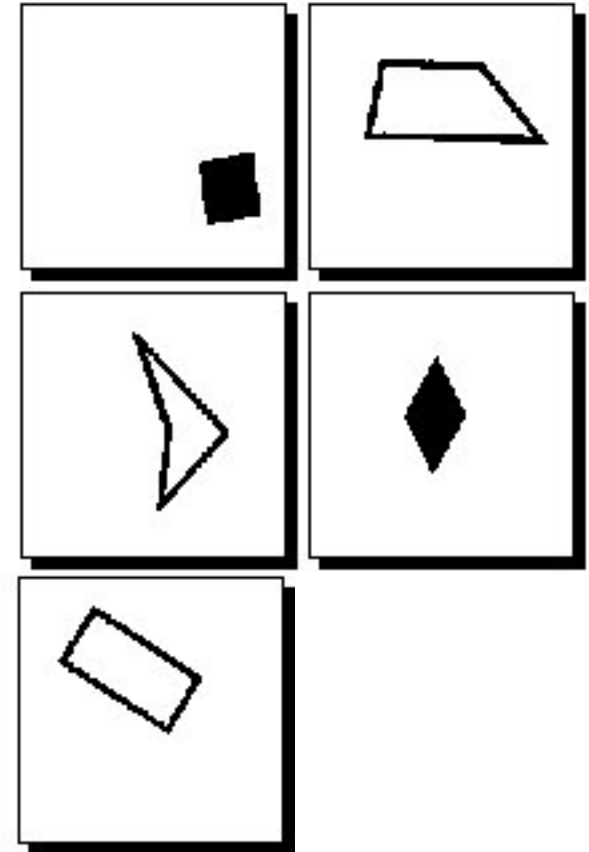
Test example: Class = ?



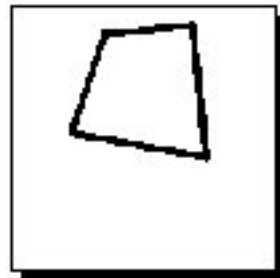
Training Examples: Class 1



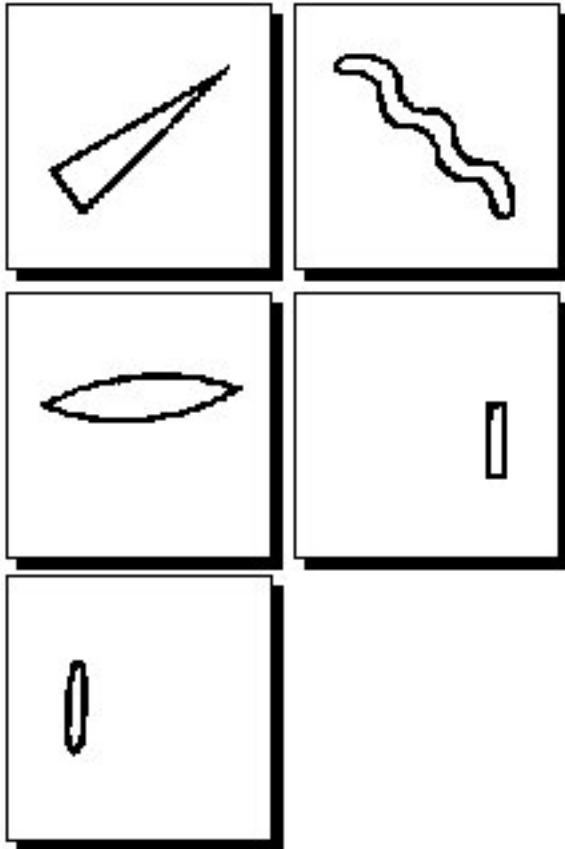
Training Examples: Class 2



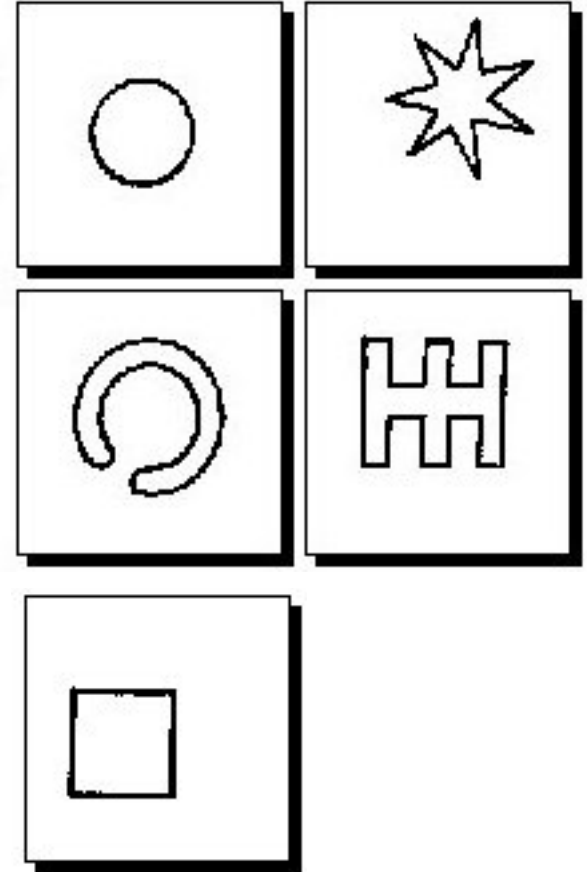
Test example: Class = ?



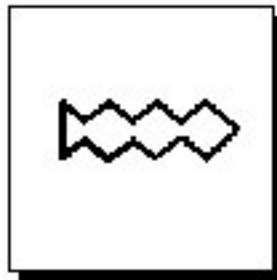
Training Examples: Class 1



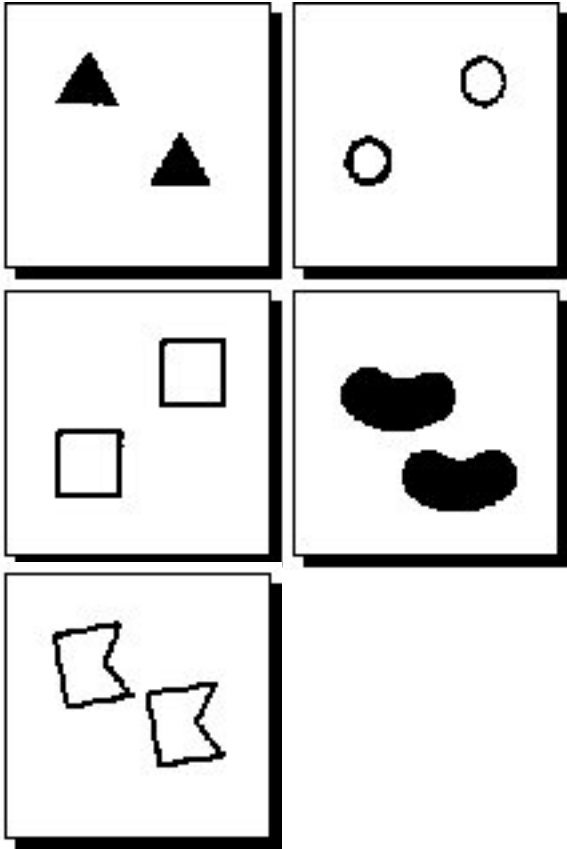
Training Examples: Class 2



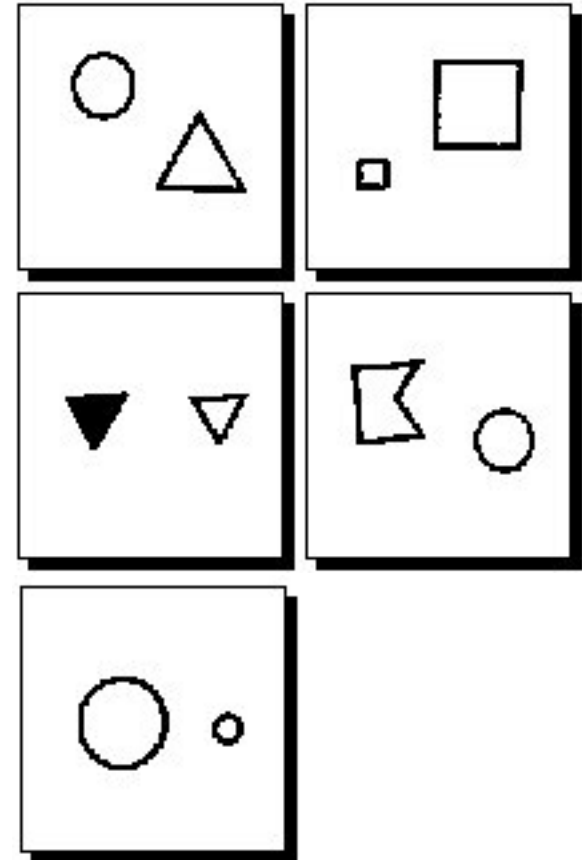
Test example: Class = ?



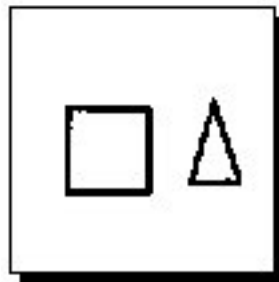
Training Examples: Class 1



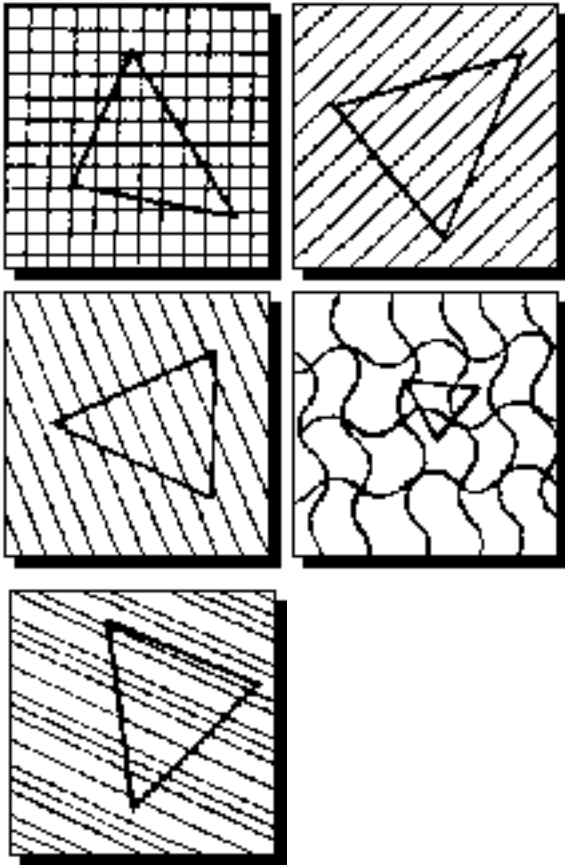
Training Examples: Class 2



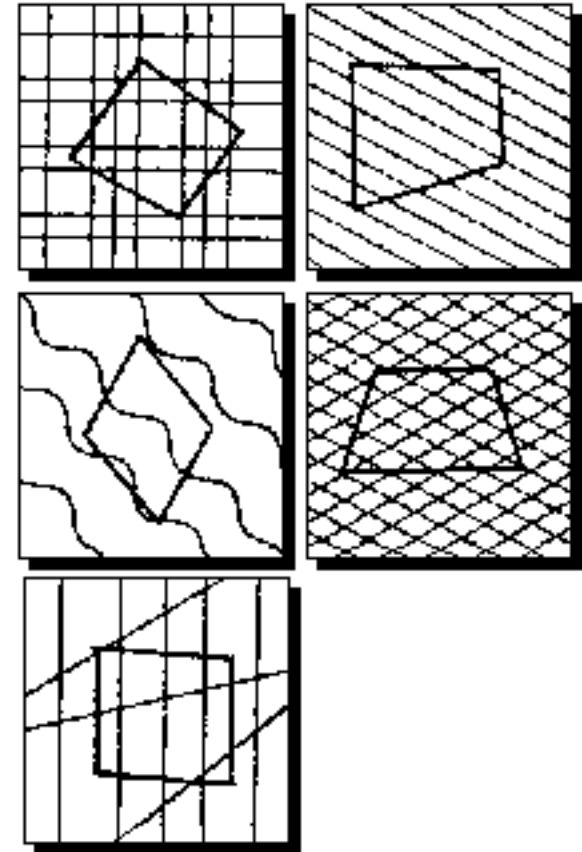
Test example: Class = ?



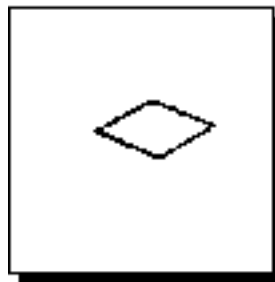
Training Examples: Class 1



Training Examples: Class 2



Test example: Class = ?



Types of machine learning tasks

- Classification
 - Output is one of a number of classes (e.g., ‘A’)
- Regression
 - Output is a real value (e.g., ‘\$35/share’)

Types of Machine Learning Methods

- **Supervised**
 - provide explicit training examples with correct answers
 - e.g. neural networks with back-propagation
- **Unsupervised**
 - no feedback information is provided
 - e.g., unsupervised clustering based on similarity

- **“Semi-supervised”**

- some feedback information is provided but it is not detailed

- e.g., only a fraction of examples are labeled

- e.g., reinforcement learning: reinforcement signal is single-valued assessment of current state

Relation between “artificial intelligence” and
“machine learning”?

Key Ingredients for Any Machine Learning Method

- **Features (or “attributes”)**

0	4	1	9	2	1	3	1	4	3
5	3	6	1	7	2	8	6	9	4
0	9	1	1	2	4	3	2	7	3
8	6	9	0	5	6	0	7	6	1
8	7	9	3	9	8	5	9	3	3
0	7	4	9	8	0	9	4	1	4
4	6	0	4	5	6	1	0	0	1
7	1	6	3	0	2	1	1	7	9
0	2	6	7	8	3	9	0	4	6
7	4	6	8	0	7	8	3	1	5

From: "TopShop" <mm@pdx.edu>
To: <mm@pdx.edu>
Subject: Dear mm get 77% off
List-Unsubscribe:
<mailto:unsubscribe-5E587F53488BBC388181@harkinsroofing.com>
MIME-Version: 1.0
Content-Type: text/plain; charset=ISO-8859-1; format=flowed
Content-Transfer-Encoding: 7bit

Wassup mm, how are you?!
Hurry up, last day huge discount...

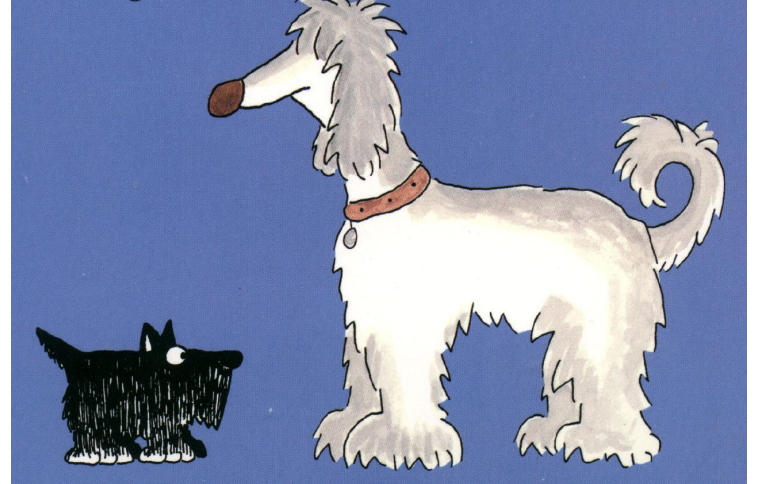
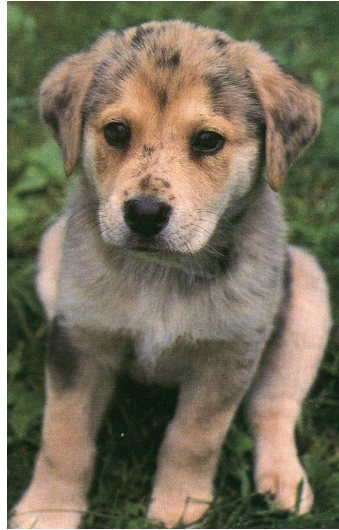
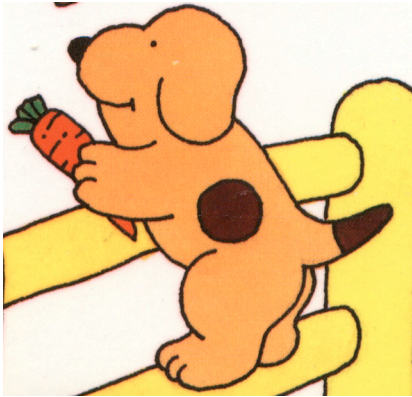
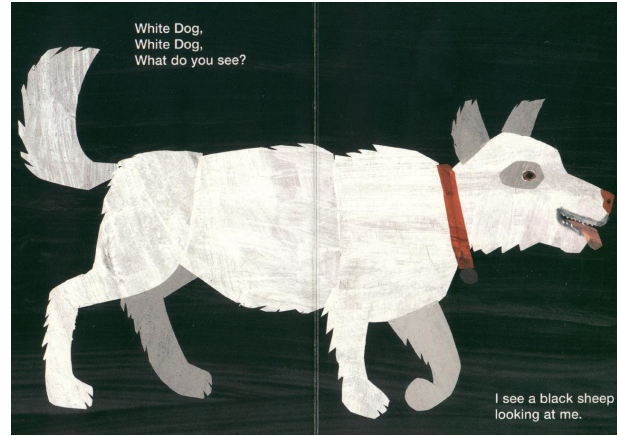
Our shop have more than 3085 pisl online.
Buy more, pay less!

Today TopSellers:

- ** Proepcia - 0.12\$
- ** Levitra - 1.64\$
- ** Ciails - 1.78\$
- ** Vaigra - 0.79\$

Purchase here <http://TvBqeUP.ngadoctor.ru/>



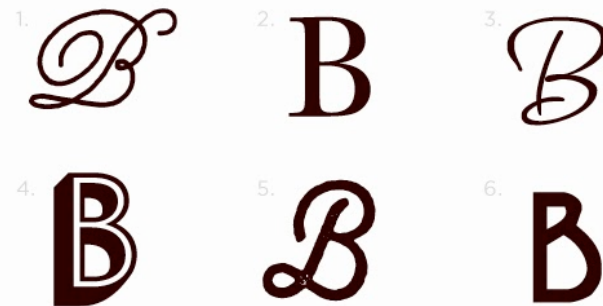


Data: Instances and Features

Instance: Individual example of a particular category or *class*



56 instances of the class “A”

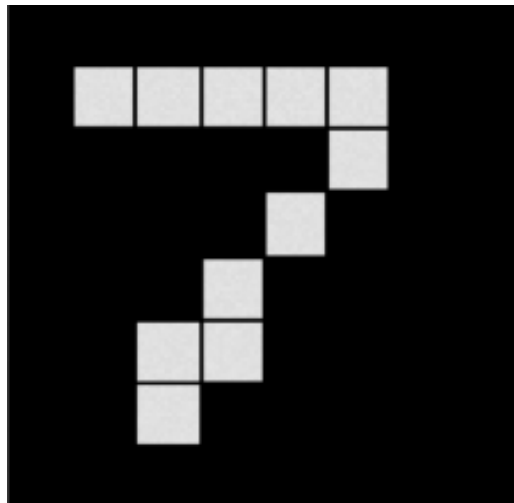


6 instances of the class “B”

Data: Instances and Features

Features: Collection of attributes of a single instance

Feature Vector: N -dimensional vector describing a single instance



→

0	0	0	0	0	0	0	0
0	1	1	1	1	1	0	0
0	0	0	0	0	1	0	0
0	0	0	0	1	0	0	0
0	0	0	1	0	0	0	0
0	0	1	1	0	0	0	0
0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0

→ 0000000000111...

Another example of computing features

[illegible]

$$\mathbf{x} = (f_1, f_2, \dots, f_{64})$$

$$= (0, 2, 13, 16, 16, 16, 2, 0, 0, \dots)$$

Notation for Instances and Features

Instance: \mathbf{x} (boldface \Rightarrow vector)

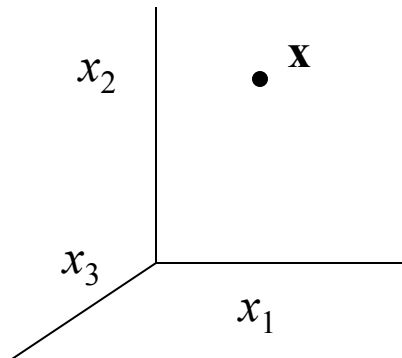
Set of M instances:

$$\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M\}$$

Instance as feature vector, with N features:

$$\mathbf{x} = (x_1, x_2, \dots, x_N)$$

Instance as a point in an N-dimensional space:



Key Ingredients for Any Machine Learning Method

- **Features** (or “attributes”)
- Underlying **Representation** for “hypothesis”, “model”, or “target function”
- **Hypothesis space**

- **Learning method**
- **Data:**
 - **Training data**
 - Used to train the model
 - **Validation (or Development) data**
 - Used to select model complexity, to determine when to stop training, or to alter training method
 - **Test data**
 - Used to evaluate trained model
- **Evaluation method**

Assumption of all ML methods:

Inductive learning hypothesis:

Any hypothesis that approximates target concept well over sufficiently large set of training examples will also approximate the concept well over other examples outside of the training set.

Difference between “induction” and “deduction”?

Goals of this course

- Broad survey of modern ML methods
- Learn by hands-on experience
- Good preparation to go further in the field, with more advanced courses or self-learning

Class Syllabus

<http://web.cecs.pdx.edu/~mm/MachineLearningWinter2016/>

Class Logistics

My office hours: M, W 2-3pm or by appt.

Class Website: <http://www.cs.pdx.edu/~mm/MachineLearningWinter2016>

Mailing list: ml2016@cs.pdx.edu

Readings / Slides: On class website

Homework: Programming / lab/ problem-solving assignments

- You may use any programming language
- You will turn in code and lab report
- Lab report will be graded

Late HW policy: Students must request and be granted an extension on any homework assignment *before* the assignment is due. Otherwise, 5% of the assignment grade will be subtracted for each day the homework is late.

Weekly Quizzes:

- 30 minutes, on Thursdays
- 1 double-sided page of notes allowed

Final Exam:

- Monday, March 14, 10:15am - 12:05pm
- 4 double-sided pages of notes allowed

Grading:

- Homework 50%
- Quizzes 20%
- Final exam 30%