CS 171: Intro to ML and DM

Christian Shelton

UC Riverside

Slide Set 1: Introduction

Slides from CS 171

- From UC Riverside
  - CS 171: Introduction to Machine Learning and Data Mining
  - Professor Christian Shelton
- DO NOT REDISTRIBUTE
  - These slides contain copyrighted material (used with permission) from
    - Elements of Statistical Learning (Hastie, et al.)
    - Pattern Recognition and Machine Learning (Bishop)
    - An Introduction to Machine Learning (Kubat)
    - Machine Learning: A Probabilistic Perspective (Murphy)
  - For use only by enrolled students in the course

Syllabus

Lecture:
Mondays, Wednesdays, and Fridays
6:10pm – 7:00pm
Bourns A125

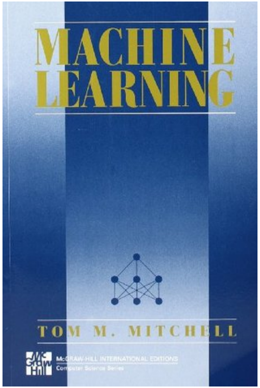
Instructor:
Christian Shelton
cshelton@cs.ucr.edu
off hrs: Wednesdays, 10am–12pm, Chung, 327

TA:
Jacob Fauber
jfaub001@ucr.edu

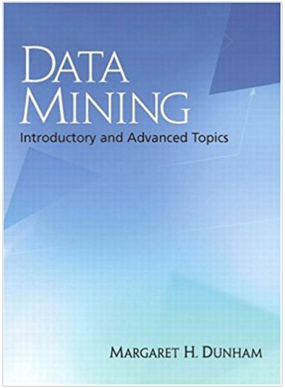
Syllabus

Possible (all optional) Texts:

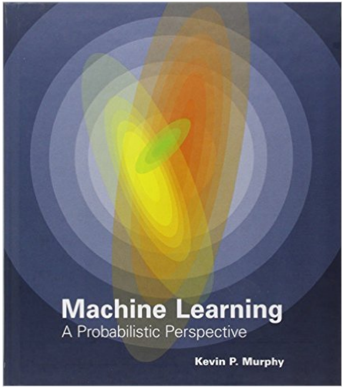
Machine Learning
by Tom M. Mitchell



Data Mining: Introductory and Advanced Topics
by Margaret H. Dunham



Machine Learning: A Probabilistic Perspective
by Kevin P. Murphy



Syllabus

due		Monday		Wednesday		Friday
	4/3	Introduction	4/5	Overfitting	4/7	Bayes classifier
	4/10	Linear regression	4/12	Linear regression	4/14	Linear regression
PS1	4/17	Linear classification	4/19	Linear classification	4/21	Linear classification
	4/24	**Midterm**	4/26	Nearest neighbor	4/28	Nearest neighbor
PS2	5/1	Nonlinear class.	5/3	Nonlinear class.	5/5	Neural networks
	5/8	Neural networks	5/10	Neural networks	5/12	Decision trees
PS3	5/15	Decision trees	5/17	Association Analysis	5/19	Association Analysis
	5/22	**Midterm**	5/24	Data normalization	5/26	Dim. reduction
PS4	5/29	Feature Generation	5/31	Clustering	6/2	Clustering
	6/5	Data Encodings	6/7	Data Visualization	6/9	Review
Final: Wednesday June 14, 2017, 7–10pm						

Syllabus

- Course Work:
  - 4 problem sets (25 pts each)
  - 2 midterms (50 pts each)
  - 1 final (100 pts)
- Problem Sets
  - Due Sunday night (11:59pm)
  - 10% deduction for every hour late!
  - programming in Matlab
  - non-programming questions
  - no external sources (friends, web, etc), except on clarification on Matlab
- Discussion sections Wednesdays
- On-line discussions at Piazza
- Matlab tutorials on iLearn

What's this Course About

Machine Learning:

- Grew out of AI
- How can a computer improve based on experience?

Data Mining:

- Grew out of Databases
- How can we make useful information out of a large database?

What's this Course About

Machine Learning:

- Grew out of AI
- How can a computer improve based on experience?

Data Mining:

- Grew out of Databases
- How can we make useful information out of a large database?

Statistics:

- What can we say mathematically about data?

Optimization:

- How can we improve a function, utility, or score

Face Detection

Face Detection

Face Detection

↓

Is this a face?

↓

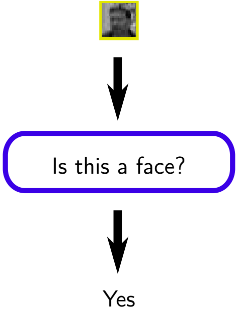
No

Face Detection

Is this a face?

No

Face Detection



Face Detection

Is this a face?

No

Face Detection

Is this a face?

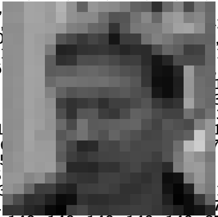
Yes

Face or Not Face?

[142, 142, 142, 143, 143, 143, 164, 164, 164, 162, 162, 162, 156, 156, 156, 179, 179, 179, 146, 146, 146, 97, 97, 97, 150, 150, 150, 141, 141,
141, 138, 138, 138, 124, 124, 124, 98, 98, 98, 94, 94, 94, 47, 47, 47, 104, 104, 104, 132, 132, 132, 74, 74, 74, 121, 121, 121, 103, 103, 103, 145,
145, 145, 146, 146, 146, 164, 164, 164, 163, 163, 163, 154, 154, 154, 174, 174, 174, 145, 145, 145, 138, 138, 138, 137, 137, 137, 110, 110, 110,
83, 83, 83, 85, 85, 85, 114, 114, 114, 136, 136, 136, 103, 103, 103, 149, 149, 149, 157, 157, 157, 128, 128, 128, 136, 136, 136, 106, 106, 106,
141, 141, 141, 141, 141, 141, 161, 161, 161, 161, 161, 161, 158, 158, 158, 152, 152, 152, 128, 128, 128, 123, 123, 123, 103, 103, 103, 98, 98, 98,
51, 51, 51, 88, 88, 88, 120, 120, 120, 124, 124, 124, 125, 125, 125, 143, 143, 143, 161, 161, 161, 164, 164, 164, 143, 143, 143, 102, 102, 102,
144, 144, 144, 142, 142, 142, 162, 162, 162, 156, 156, 156, 148, 148, 148, 85, 85, 85, 72, 72, 72, 56, 56, 56, 69, 69, 69, 60, 60, 60, 21, 21, 21, 60,
60, 60, 61, 61, 61, 113, 113, 113, 129, 129, 129, 120, 120, 120, 105, 105, 105, 98, 98, 98, 101, 101, 101, 99, 99, 99, 143, 143, 143, 144, 144, 144,
162, 162, 162, 157, 157, 157, 122, 122, 122, 34, 34, 34, 37, 37, 37, 49, 49, 49, 51, 51, 51, 54, 54, 54, 47, 47, 47, 37, 37, 37, 44, 44, 44, 43, 43,
43, 75, 75, 75, 102, 102, 102, 94, 94, 94, 62, 62, 62, 65, 65, 65, 103, 103, 103, 142, 142, 142, 144, 144, 144, 162, 162, 162, 159, 159, 159, 93, 93,
93, 47, 47, 47, 51, 51, 51, 77, 77, 77, 76, 76, 76, 80, 80, 80, 72, 72, 72, 61, 61, 61, 62, 62, 62, 43, 43, 43, 21, 21, 21, 37, 37, 37, 75, 75, 75, 73,
73, 73, 73, 73, 73, 102, 102, 102, 141, 141, 141, 144, 144, 144, 163, 163, 163, 161, 161, 161, 92, 92, 92, 95, 95, 95, 93, 93, 93, 102, 102, 102,
104, 104, 104, 98, 98, 98, 90, 90, 90, 81, 81, 81, 76, 76, 76, 51, 51, 51, 19, 19, 19, 15, 15, 15, 38, 38, 38, 154, 154, 154, 103, 103, 103, 106, 106,
106, 142, 142, 142, 146, 146, 146, 162, 162, 162, 160, 160, 160, 106, 106, 106, 116, 116, 116, 115, 115, 115, 112, 112, 112, 110, 110, 110, 112,
112, 112, 105, 105, 105, 97, 97, 97, 86, 86, 86, 53, 53, 53, 14, 14, 14, 12, 12, 12, 30, 30, 30, 151, 151, 151, 110, 110, 110, 104, 104, 104, 141,
141, 141, 145, 145, 145, 162, 162, 162, 159, 159, 159, 130, 130, 130, 98, 98, 98, 83, 83, 83, 89, 89, 89, 87, 87, 87, 80, 80, 80, 84, 84, 84, 89, 89,
89, 80, 80, 80, 72, 72, 72, 23, 23, 23, 12, 12, 12, 31, 31, 31, 150, 150, 150, 111, 111, 111, 107, 107, 107, 140, 140, 140, 146, 146, 146, 162, 162,
162, 156, 156, 156, 144, 144, 144, 64, 64, 64, 49, 49, 49, 56, 56, 56, 63, 63, 63, 47, 47, 47, 63, 63, 63, 64, 64, 64, 74, 74, 74, 84, 84, 84, 57, 57,
57, 44, 44, 44, 52, 52, 52, 163, 163, 163, 114, 114, 114, 105, 105, 105, 142, 142, 142, 144, 144, 144, 161, 161, 161, 156, 156, 156, 148, 148, 148,
79, 79, 79, 49, 49, 49, 86, 86, 86, 88, 88, 88, 61, 61, 61, 55, 55, 55, 75, 75, 75, 97, 97, 97, 92, 92, 92, 75, 75, 75, 66, 66, 66, 66, 66, 66, 114, 114,
114, 92, 92, 92, 115, 115, 115, 141, 141, 141, 143, 143, 143, 163, 163, 163, 158, 158, 158, 146, 146, 146, 106, 106, 106, 74, 74, 74, 116, 116,
116, 90, 90, 90, 92, 92, 92, 96, 96, 96, 99, 99, 99, 101, 101, 101, 80, 80, 80, 67, 67, 67, 86, 86, 86, 80, 80, 80, 74, 74, 74, 128, 128, 128, 152,
152, 152, 141, 141, 141, 145, 145, 145, 164, 164, 164, 149, 149, 149, 149, 149, 149, 87, 87, 87, 99, 99, 99, 114, 114, 114, 93, 93, 93, 81, 81, 81,
100, 100, 100, 88, 88, 88, 80, 80, 80, 68, 68, 68, 70, 70, 70, 80, 80, 80, 86, 86, 86, 96, 96, 96, 200, 200, 200, 152, 152, 152, 138, 138, 138, 145,
145, 145, 162, 162, 162, 155, 155, 155, 155, 155, 155, 97, 97, 97, 69, 69, 69, 38, 38, 38, 42, 42, 42, 71, 71, 71, 68, 68, 68, 81, 81, 81, 71, 71, 71,
73, 73, 73, 73, 73, 73, 80, 80, 80, 110, 110, 110, 186, 186, 186, 205, 205, 205, 145, 145, 145, 142, 142, 142, 149, 149, 149, 158, 158, 158, 157,
157, 157, 156, 156, 156, 91, 91, 91, 62, 62, 62, 52, 52, 52, 61, 61, 61, 67, 67, 67, 68, 68, 68, 70, 70, 70, 72, 72, 72, 71, 71, 71, 62, 62, 62, 82, 82,
82, 120, 120, 120, 157, 157, 157, 143, 143, 143, 119, 119, 119, 142, 142, 142, 147, 147, 147, 159, 159, 159, 155, 155, 155, 158, 158, 158, 110,
110, 110, 40, 40, 40, 40, 40, 40, 38, 38, 38, 51, 51, 51, 50, 50, 50, 63, 63, 63, 70, 70, 70, 57, 57, 57, 55, 55, 55, 42, 42, 42, 125, 125, 125, 153,
153, 153, 156, 156, 156, 128, 128, 128, 150, 150, 150, 151, 151, 151, 134, 134, 134, 142, 142, 142, 111, 111, 111, 73, 73, 73, 57, 57, 57, 51, 51,
51, 50, 50, 50, 64, 64, 64, 65, 65, 65, 60, 60, 60, 53, 53, 53, 48, 48, 48, 57, 57, 57, 13, 13, 13, 18, 18, 18, 124, 124, 124, 152, 152, 152, 126, 126,
126, 147, 147, 147, 138, 138, 138, 104, 104, 104, 62, 62, 62, 63, 63, 63, 40, 40, 40, 33, 33, 33, 61, 61, 61, 63, 63, 63, 66, 66, 66, 61, 61, 61, 49,
49, 49, 49, 49, 49, 55, 55, 55, 59, 59, 59, 19, 19, 19, 5, 5, 5, 34, 34, 34, 138, 138, 138, 122, 122, 122, 103, 103, 103, 75, 75, 75, 46, 46, 46, 49,
49, 49, 32, 32, 32, 10, 10, 10, 14, 14, 14, 50, 50, 50, 44, 44, 44, 45, 45, 45, 50, 50, 50, 50, 50, 50, 56, 56, 56, 61, 61, 61, 57, 57, 57, 31, 31, 31, 2,
2, 2, 12, 12, 12, 56, 56, 56, 101, 101, 101, 51, 51, 51, 52, 52, 52, 39, 39, 39, 32, 32, 32, 9, 9, 9, 2, 2, 2, 61, 61, 61, 70, 70, 70, 54, 54, 54, 53, 53,
53, 59, 59, 59, 61, 61, 61, 63, 63, 63, 62, 62, 62, 55, 55, 55, 45, 45, 45, 32, 32, 32, 12, 12, 12, 19, 19, 19, 42, 42, 42]

Face Detection

Face or Not Face?



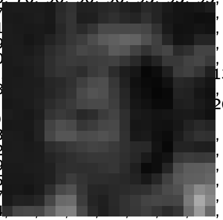
[142, 142, 142, 143, 143, 143, 164, 164, 164, 162, 162, 162, 156, 156, 156, 179, 179, 179, 146, 146, 146, 97, 97, 97, 150, 150, 150, 141, 141,
141, 138, 138, 138, 124, 124, 124, 98, 98, 98, 94, 94, 94, 47, 47, 47, 104, 104, 104, 132, 132, 132, 74, 74, 74, 121, 121, 121, 103, 103, 103, 145,
145, 145, 146, 146, 146, 164, 164, 164, 163, 163, 163, 154, 154, 154, 174, 174, 174, 145, 145, 145, 138, 138, 138, 137, 137, 137, 110, 110, 110,
83, 83, 83, 85, 85, 85, 114, 114, 114, 136, 136, 136, 103, 103, 103, 149, 149, 149, 157, 157, 157, 128, 128, 128, 136, 136, 136, 106, 106, 106,
141, 141, 141, 141, 141, 141, 161, 161, 161, 161, 161, 161, 158, 158, 158, 152, 152, 152, 128, 128, 128, 123, 123, 123, 103, 103, 103, 98, 98, 98,
51, 51, 51, 88, 88, 88, 120, 120, 120, 124, 124, 124, 125, 125, 125, 143, 143, 143, 161, 161, 161, 164, 164, 164, 143, 143, 143, 102, 102, 102,
144, 144, 144, 142, 142, 142, 162, 162, 162, 156, 156, 156, 148, 148, 148, 85, 85, 85, 72, 72, 72, 56, 56, 56, 69, 69, 69, 60, 60, 60, 21, 21, 21, 60,
60, 60, 61, 61, 61, 113, 113, 113, 129, 129, 129, 120, 120, 120, 105, 105, 105, 98, 98, 98, 101, 101, 101, 99, 99, 99, 143, 143, 143, 144, 144, 144,
162, 162, 162, 157, 157, 157, 122, 122, 122, 34, 34, 34, 37, 51, 54, 54, 54, 47, 47, 47, 37, 37, 37, 44, 44, 44, 43, 43,
43, 75, 75, 75, 102, 102, 102, 94, 94, 94, 62, 62, 62, 65, 65, 2, 142, 144, 144, 144, 162, 162, 162, 159, 159, 159, 93, 93,
93, 47, 47, 47, 51, 51, 51, 77, 77, 77, 76, 76, 76, 80, 80, 80 , 62, 62, 43, 43, 43, 21, 21, 21, 37, 37, 37, 75, 75, 75, 73,
73, 73, 73, 73, 73, 102, 102, 102, 141, 141, 141, 144, 144, 1, 161, 92, 92, 92, 95, 95, 95, 93, 93, 93, 102, 102, 102,
104, 104, 104, 98, 98, 98, 90, 90, 90, 81, 81, 81, 76, 76, 76 15, 15, 38, 38, 38, 154, 154, 154, 103, 103, 103, 106, 106,
106, 142, 142, 142, 146, 146, 146, 162, 162, 162, 160, 160, 16, 116, 115, 115, 115, 112, 112, 112, 110, 110, 110, 112,
112, 112, 105, 105, 105, 97, 97, 97, 86, 86, 86, 53, 53, 53, 30, 30, 151, 151, 151, 110, 110, 110, 104, 104, 104, 141,
141, 141, 145, 145, 145, 162, 162, 162, 159, 159, 159, 130, 3, 83, 89, 89, 89, 87, 87, 87, 80, 80, 80, 84, 84, 84, 89, 89,
89, 80, 80, 80, 72, 72, 72, 23, 23, 23, 12, 12, 12, 31, 31, 31 111, 107, 107, 107, 140, 140, 140, 146, 146, 146, 162, 162,
162, 156, 156, 156, 144, 144, 144, 64, 64, 64, 49, 49, 49, 5 , 47, 63, 63, 63, 64, 64, 64, 74, 74, 74, 84, 84, 84, 57, 57,
57, 44, 44, 44, 52, 52, 52, 163, 163, 163, 114, 114, 114, 10 144, 144, 144, 161, 161, 161, 156, 156, 156, 148, 148, 148,
79, 79, 79, 49, 49, 49, 86, 86, 86, 88, 88, 88, 61, 61, 61, 55 97, 92, 92, 92, 75, 75, 75, 66, 66, 66, 66, 66, 66, 114, 114,
114, 92, 92, 92, 115, 115, 115, 141, 141, 141, 143, 143, 143 158, 146, 146, 146, 106, 106, 106, 74, 74, 74, 116, 116,
116, 90, 90, 90, 92, 92, 92, 96, 96, 96, 99, 99, 99, 101, 101 7, 86, 86, 86, 80, 80, 80, 74, 74, 74, 128, 128, 128, 152,
152, 152, 141, 141, 141, 145, 145, 145, 164, 164, 164, 149, 149, 149, 149, 149, 149, 87, 87, 87, 99, 99, 99, 114, 114, 114, 93, 93, 93, 81, 81, 81,
100, 100, 100, 88, 88, 88, 80, 80, 80, 68, 68, 68, 70, 70, 70, 80, 80, 80, 86, 86, 86, 96, 96, 96, 200, 200, 200, 152, 152, 152, 138, 138, 138, 145,
145, 145, 162, 162, 162, 155, 155, 155, 155, 155, 155, 97, 97, 97, 69, 69, 69, 38, 38, 38, 42, 42, 42, 71, 71, 71, 68, 68, 68, 81, 81, 81, 71, 71, 71,
73, 73, 73, 73, 73, 73, 80, 80, 80, 110, 110, 110, 186, 186, 186, 205, 205, 205, 145, 145, 145, 142, 142, 142, 149, 149, 149, 158, 158, 158, 157,
157, 157, 156, 156, 156, 91, 91, 91, 62, 62, 62, 52, 52, 52, 61, 61, 61, 67, 67, 67, 68, 68, 68, 70, 70, 70, 72, 72, 72, 71, 71, 71, 62, 62, 62, 82, 82,
82, 120, 120, 120, 157, 157, 157, 143, 143, 143, 119, 119, 119, 142, 142, 142, 147, 147, 147, 159, 159, 159, 155, 155, 155, 158, 158, 158, 110,
110, 110, 40, 40, 40, 40, 40, 40, 38, 38, 38, 51, 51, 51, 50, 50, 50, 63, 63, 63, 70, 70, 70, 57, 57, 57, 55, 55, 55, 42, 42, 42, 125, 125, 125, 153,
153, 153, 156, 156, 156, 128, 128, 128, 150, 150, 150, 151, 151, 151, 134, 134, 134, 142, 142, 142, 111, 111, 111, 73, 73, 73, 57, 57, 57, 51, 51,
51, 50, 50, 50, 64, 64, 64, 65, 65, 65, 60, 60, 60, 53, 53, 53, 48, 48, 48, 57, 57, 57, 13, 13, 13, 18, 18, 18, 124, 124, 124, 152, 152, 152, 126, 126,
126, 147, 147, 147, 138, 138, 138, 104, 104, 104, 62, 62, 62, 63, 63, 63, 40, 40, 40, 33, 33, 33, 61, 61, 61, 63, 63, 63, 66, 66, 66, 61, 61, 61, 49,
49, 49, 49, 49, 49, 55, 55, 55, 59, 59, 59, 19, 19, 19, 5, 5, 5, 34, 34, 34, 138, 138, 138, 122, 122, 122, 103, 103, 103, 75, 75, 75, 46, 46, 46, 49,
49, 49, 32, 32, 32, 10, 10, 10, 14, 14, 14, 50, 50, 50, 44, 44, 44, 45, 45, 45, 50, 50, 50, 50, 50, 50, 56, 56, 56, 61, 61, 61, 57, 57, 57, 31, 31, 31, 2,
2, 2, 12, 12, 12, 56, 56, 56, 101, 101, 101, 51, 51, 51, 52, 52, 52, 39, 39, 39, 32, 32, 32, 9, 9, 9, 2, 2, 2, 61, 61, 61, 70, 70, 70, 54, 54, 54, 53, 53,
53, 59, 59, 59, 61, 61, 61, 63, 63, 63, 62, 62, 62, 55, 55, 55, 45, 45, 45, 32, 32, 32, 12, 12, 12, 19, 19, 19, 42, 42, 42]

Face Detection

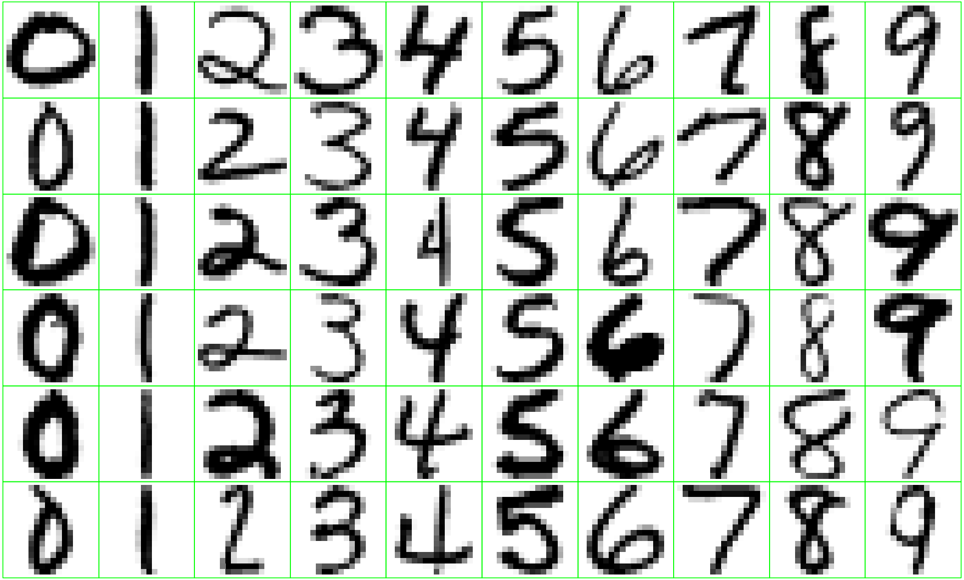
Face or Not Face?

[27, 27, 27, 28, 28, 28, 28, 28, 28, 28, 28, 28, 25, 25, 25, 41, 41, 41, 50, 50, 50, 33, 33, 33, 34, 34, 34, 39, 39, 39, 42, 42, 42, 34, 34, 34, 28, 28, 28, 41, 41, 41, 36, 36, 36, 26, 26, 26, 20, 20, 20, 20, 20, 20, 23, 23, 23, 29, 29, 29, 26, 26, 26, 25, 25, 25, 27, 27, 27, 29, 29, 29, 25, 25, 25, 43, 43, 43, 52, 52, 52, 30, 30, 30, 31, 31, 31, 42, 42, 42, 48, 48, 48, 40, 40, 40, 36, 36, 36, 48, 48, 48, 36, 36, 36, 23, 23, 23, 20, 20, 20, 18, 18, 18, 15, 15, 15, 16, 16, 16, 28, 28, 28, 26, 26, 26, 25, 25, 25, 26, 26, 26, 24, 24, 24, 39, 39, 39, 52, 52, 52, 42, 42, 42, 53, 53, 53, 73, 73, 73, 73, 73, 73, 72, 72, 72, 73, 73, 73, 69, 69, 69, 50, 50, 50, 26, 26, 26, 17, 17, 17, 18, 18, 18, 21, 21, 21, 18, 18, 18, 28, 28, 28, 27, 27, 27, 22, 22, 22, 26, 26, 26, 34, 34, 34, 43, 43, 43, 54, 54, 54, 61, 61, 61, 76, 76, 76, 90, 90, 90, 95, 95, 95, 101, 101, 101, 98, 98, 98, 80, 80, 80, 64, 64, 64, 37, 37, 37, 19, 19, 19, 21, 21, 21, 36, 36, 36, 36, 36, 36, 27, 27, 27, 28, 28, 28, 26, 26, 26, 31, 31, 31, 43, 43, 43, 51, 51, 51, 60, 60, 60, 71, 71, 71, 67, 67, 67, 71, 71, 71, 90, 90, 90, 85, 85, 85, 66, 66, 66, 51, 51, 51, 62, 62, 62, 69, 69, 69, 42, 42, 42, 48, 48, 48, 67, 67, 67, 67, 67, 67, 26, 26, 26, 28, 28, 28, 30, 30, 30, 34, 34, 34, 46, 46, 46, 55, 55, 55, 59, 59, 59, 57, 57, 57, 60, 60, 60, 62, 62, 62, 65, 65, 65, 54, 54, 54, 41, 41, 41, 24, 24, 24, 35, 35, 35, 56, 56, 56, 56, 56, 56, 66, 66, 66, 70, 70, 70, 58, 58, 58, 26, 26, 26, 25, 25, 25, 28, 28, 28, 33, 33, 33, 46, 46, 46, 53, 53, 53, 51, 51, 51, 45, 45, 45, 58, 58, 58, 64, 64, 64, 44, 44, 44, 29, 29, 29, 22, 22, 22, 7 7, 7 13, 13, 13, 31, 31, 31, 59, 59, 59, 64, 64, 64, 53, 53, 53, 32, 32, 32, 26, 26, 26, 27, 27, 27, 28, 28, 28, 29, 29, 29, 38, 38, 38, 41, 41, 41, 49, 49, 49, 63, 63, 63, 75, 75, 75, 78, 78, 78, 55, 55, 55, 33, 33, 33, 16, 16, 16, 10, 10, 10, 8 8, 8 17, 17, 17, 40, 40, 40, 37, 37, 37, 29, 29, 29, 21, 21, 21, 26, 26, 26, 28, 28, 28, 28, 28, 28, 31, 31, 31, 37, 37, 37, 44, 44, 44, 52, 52, 52, 73, 73, 73, 83, 83, 83, 79, 79, 79, 59, 59, 59, 38, 38, 38, 19, 19, 19, 13, 13, 13, 6 6, 6 9, 9 9, 35, 35, 35, 33, 33, 33, 22, 22, 22, 17, 17, 17, 24, 24, 24, 26, 26, 26, 28, 28, 28, 33, 33, 33, 43, 43, 43, 51, 51, 51, 58, 58, 58, 69, 69, 69, 73, 73, 73, 73, 73, 73, 62, 62, 62, 40, 40, 40, 22, 22, 22, 17, 17, 17, 10, 10, 10, 10, 10, 10, 30, 30, 30, 32, 32, 32, 26, 26, 26, 20, 20, 20, 29, 29, 29, 26, 26, 26, 24, 24, 24, 30, 30, 30, 51, 51, 51, 58, 58, 58, 53, 53, 53, 40, 40, 40, 45, 45, 45, 59, 59, 59, 67, 67, 67, 47, 47, 47, 26, 26, 26, 23, 23, 23, 19, 19, 19, 21, 21, 21, 25, 25, 25, 26, 26, 26, 23, 23, 23, 17, 17, 17, 24, 24, 24, 25, 25, 25, 25, 25, 25, 34, 34, 34, 66, 66, 66, 78, 78, 78, 73, 73, 73, 54, 54, 54, 60, 60, 60, 70, 70, 70, 72, 72, 72, 47, 47, 47, 22, 22, 22, 23, 23, 23, 18, 18, 18, 17, 17, 17, 23, 23, 23, 34, 34, 34, 41, 41, 41, 36, 36, 36, 28, 28, 28, 27, 27, 27, 24, 24, 24, 34, 34, 34, 70, 70, 70, 85, 85, 85, 85, 85, 85, 74, 74, 74, 76, 76, 76, 81, 81, 81, 75, 75, 75, 49, 49, 49, 24, 24, 24, 26, 26, 26, 16, 16, 16, 12, 12, 12, 28, 28, 28, 41, 41, 41, 50, 50, 50, 48, 48, 48, 27, 27, 27, 28, 28, 28, 26, 26, 26, 28, 28, 28, 44, 44, 44, 50, 50, 50, 57, 57, 57, 62, 62, 62, 67, 67, 67, 65, 65, 65, 58, 58, 58, 49, 49, 49, 41, 41, 41, 40, 40, 40, 19, 19, 19, 8 8, 8, 35, 35, 35, 35, 35, 35, 31, 31, 31, 30, 30, 30, 28, 28, 28, 28, 28, 28, 27, 27, 27, 28, 28, 28, 36, 36, 36, 42, 42, 42, 47, 47, 47, 52, 52, 52, 55, 55, 55, 55, 55, 55, 54, 54, 54, 47, 47, 47, 39, 39, 39, 26, 26, 26, 14, 14, 14, 19, 19, 19, 55, 55, 55, 60, 60, 60, 57, 57, 57, 59, 59, 59, 26, 26, 26, 26, 26, 26, 28, 28, 28, 29, 29, 29, 32, 32, 32, 45, 45, 45, 47, 47, 47, 46, 46, 46, 47, 47, 47, 51, 51, 51, 54, 54, 54, 44, 44, 44, 31, 31, 31, 10, 10, 10, 14, 14, 14, 37, 37, 37, 79, 79, 79, 92, 92, 92, 93, 93, 93, 97, 97, 97, 25, 25, 25, 25, 25, 25, 25, 25, 25, 28, 28, 28, 32, 32, 32, 47, 47, 47, 60, 60, 60, 51, 51, 51, 45, 45, 45, 45, 45, 45, 43, 43, 43, 40, 40, 40, 35, 35, 35, 24, 24, 24, 38, 38, 38, 64, 64, 64, 88, 88, 88, 95, 95, 95, 96, 96, 96, 102, 102, 102, 23, 23, 23, 24, 24, 24, 21, 21, 21, 27, 27, 27, 46, 46, 46, 73, 73, 73, 81, 81, 81, 54, 54, 54, 38, 38, 38, 37, 37, 37, 41, 41, 41, 43, 43, 43, 46, 46, 46, 48, 48, 48, 60, 60, 60, 74, 74, 74, 81, 81, 81, 112, 112, 112, 138, 138, 138, 127, 127, 127, 25, 25, 25, 33, 33, 33, 32, 32, 32, 38, 38, 38, 70, 70, 70, 105, 105, 105, 108, 108, 108, 60, 60, 60, 36, 36, 36, 31, 31, 31, 41, 41, 41, 50, 50, 50, 57, 57, 57, 67, 67, 67, 70, 70, 70, 68, 68, 68, 70, 70, 70, 122, 122, 122, 169, 169, 169, 136, 136, 136, 47, 47, 47, 70, 70, 70, 73, 73, 73, 71, 71, 71, 96, 96, 96, 125, 125, 125, 128, 128, 128, 72, 72, 72, 41, 41, 41, 30, 30, 30, 40, 40, 40, 50, 50, 50, 56, 56, 56, 62, 62, 62, 53, 53, 53, 42, 42, 42, 52, 52, 52, 108, 108, 108, 154, 154, 154, 104, 104, 104]

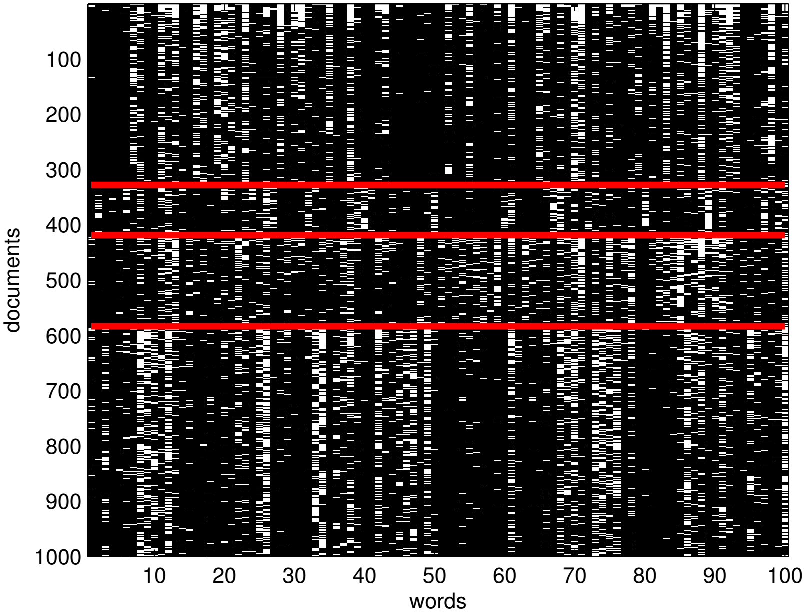
Face or Not Face?

[27, 27, 27, 28, 28, 28, 28, 28, 28, 28, 28, 28, 25, 25, 25, 41, 41, 41, 50, 50, 50, 33, 33, 33, 34, 34, 34, 39, 39, 39, 42, 42, 42, 34, 34, 34, 28, 28,
28, 41, 41, 41, 36, 36, 36, 26, 26, 26, 20, 20, 20, 20, 20, 20, 23, 23, 23, 29, 29, 29, 26, 26, 26, 25, 25, 25, 27, 27, 27, 29, 29, 29, 25, 25, 25, 43,
43, 43, 52, 52, 52, 30, 30, 30, 31, 31, 31, 42, 42, 42, 48, 48, 48, 40, 40, 40, 36, 36, 36, 48, 48, 48, 36, 36, 36, 23, 23, 23, 20, 20, 20, 18, 18, 18,
15, 15, 15, 16, 16, 16, 28, 28, 28, 26, 26, 26, 25, 25, 25, 26, 26, 26, 24, 24, 24, 39, 39, 39, 52, 52, 52, 42, 42, 42, 53, 53, 53, 73, 73, 73, 73, 73,
73, 72, 72, 72, 73, 73, 73, 69, 69, 69, 50, 50, 50, 26, 26, 26, 17, 17, 17, 18, 18, 18, 21, 21, 21, 18, 18, 18, 28, 28, 28, 27, 27, 27, 22, 22, 22, 26,
26, 26, 34, 34, 34, 43, 43, 43, 54, 54, 54, 61, 61, 61, 76, 76, 76, 90, 90, 90, 95, 95, 95, 101, 101, 101, 98, 98, 98, 80, 80, 80, 64, 64, 64, 37, 37,
37, 19, 19, 19, 21, 21, 21, 36, 36, 36, 36, 36, 36, 27, 27, 27  , 31, 31, 43, 43, 43, 51, 51, 51, 60, 60, 60, 71, 71, 71, 67,
67, 67, 71, 71, 71, 90, 90, 90, 85, 85, 85, 66, 66, 66, 51, 51 , 42, 42, 42, 48, 48, 48, 67, 67, 67, 67, 67, 67, 26, 26, 26,
28, 28, 28, 30, 30, 30, 34, 34, 34, 46, 46, 46, 55, 55, 55, 59 , 60, 62, 62, 62, 65, 65, 65, 54, 54, 54, 41, 41, 41, 24, 24,
24, 35, 35, 35, 56, 56, 56, 56, 56, 56, 66, 66, 66, 70, 70, 70 , 25, 25, 28, 28, 28, 33, 33, 33, 46, 46, 46, 53, 53, 53, 51,
51, 51, 45, 45, 45, 58, 58, 58, 64, 64, 64, 44, 44, 44, 29, 29 13, 13, 31, 31, 31, 59, 59, 59, 64, 64, 64, 53, 53, 53, 32, 32,
32, 26, 26, 26, 27, 27, 27, 28, 28, 28, 29, 29, 29, 38, 38, 38 , 63, 63, 75, 75, 75, 78, 78, 78, 55, 55, 55, 33, 33, 33, 16,
16, 16, 10, 10, 10, 8 8, 8 17, 17, 17, 40, 40, 40, 37, 37, 37, 26, 26, 28, 28, 28, 28, 28, 28, 31, 31, 31, 37, 37, 37, 44, 44,
44, 52, 52, 52, 73, 73, 73, 83, 83, 83, 79, 79, 79, 59, 59, 59 13, 13, 6 6, 6 9, 9 9, 35, 35, 35, 33, 33, 33, 22, 22, 22, 17,
17, 17, 24, 24, 24, 26, 26, 26, 28, 28, 28, 33, 33, 33, 43, 43 , 69, 69, 69, 73, 73, 73, 73, 73, 73, 62, 62, 62, 40, 40, 40,
22, 22, 22, 17, 17, 17, 10, 10, 10, 10, 10, 10, 30, 30, 30, 32 , 20, 29, 29, 29, 26, 26, 26, 24, 24, 24, 30, 30, 30, 51, 51,
51, 58, 58, 58, 53, 53, 53, 40, 40, 40, 45, 45, 45, 59, 59, 59 , 26, 26, 23, 23, 23, 19, 19, 19, 21, 21, 21, 25, 25, 25, 26,
26, 26, 23, 23, 23, 17, 17, 17, 24, 24, 24, 25, 25, 25, 25, 25 , 78, 78, 78, 73, 73, 73, 54, 54, 54, 60, 60, 60, 70, 70, 70,
72, 72, 72, 47, 47, 47, 22, 22, 22, 23, 23, 23, 18, 18, 18, 17 , 34, 41, 41, 41, 36, 36, 36, 28, 28, 28, 27, 27, 27, 24, 24,
24, 34, 34, 34, 70, 70, 70, 85, 85, 85, 85, 85, 85, 74, 74, 74 , 75, 75, 49, 49, 49, 24, 24, 24, 26, 26, 26, 16, 16, 16, 12,
12, 12, 28, 28, 28, 41, 41, 41, 50, 50, 50, 48, 48, 48, 27, 27, 27, 28, 28, 28, 26, 26, 26, 28, 28, 28, 44, 44, 44, 50, 50, 50, 57, 57, 57, 62, 62, 62,
67, 67, 67, 65, 65, 65, 58, 58, 58, 49, 49, 49, 41, 41, 41, 40, 40, 40, 19, 19, 19, 8 8, 8, 35, 35, 35, 35, 35, 35, 31, 31, 31, 30, 30, 30, 28, 28, 28,
28, 28, 28, 27, 27, 27, 28, 28, 28, 36, 36, 36, 42, 42, 42, 47, 47, 47, 52, 52, 52, 55, 55, 55, 55, 55, 55, 54, 54, 54, 47, 47, 47, 39, 39, 39, 26, 26,
26, 14, 14, 14, 19, 19, 19, 55, 55, 55, 60, 60, 60, 57, 57, 57, 59, 59, 59, 26, 26, 26, 26, 26, 26, 28, 28, 28, 29, 29, 29, 32, 32, 32, 45, 45, 45, 47,
47, 47, 46, 46, 46, 47, 47, 47, 51, 51, 51, 54, 54, 54, 44, 44, 44, 31, 31, 31, 10, 10, 10, 14, 14, 14, 37, 37, 37, 79, 79, 79, 92, 92, 92, 93, 93, 93,
97, 97, 97, 25, 25, 25, 25, 25, 25, 25, 25, 25, 28, 28, 28, 32, 32, 32, 47, 47, 47, 60, 60, 60, 51, 51, 51, 45, 45, 45, 45, 45, 45, 43, 43, 43, 40, 40,
40, 35, 35, 35, 24, 24, 24, 38, 38, 38, 64, 64, 64, 88, 88, 88, 95, 95, 95, 96, 96, 96, 102, 102, 102, 23, 23, 23, 24, 24, 24, 21, 21, 21, 27, 27, 27,
46, 46, 46, 73, 73, 73, 81, 81, 81, 54, 54, 54, 38, 38, 38, 37, 37, 37, 41, 41, 41, 43, 43, 43, 46, 46, 46, 48, 48, 48, 60, 60, 60, 74, 74, 74, 81, 81,
81, 112, 112, 112, 138, 138, 138, 127, 127, 127, 25, 25, 25, 33, 33, 33, 32, 32, 32, 38, 38, 38, 70, 70, 70, 105, 105, 105, 108, 108, 108, 60, 60, 60,
36, 36, 36, 31, 31, 31, 41, 41, 41, 50, 50, 50, 57, 57, 57, 67, 67, 67, 70, 70, 70, 68, 68, 68, 70, 70, 70, 122, 122, 122, 169, 169, 169, 136, 136,
136, 47, 47, 47, 70, 70, 70, 73, 73, 73, 71, 71, 71, 96, 96, 96, 125, 125, 125, 128, 128, 128, 72, 72, 72, 41, 41, 41, 30, 30, 30, 40, 40, 40, 50, 50,
50, 56, 56, 56, 62, 62, 62, 53, 53, 53, 42, 42, 42, 52, 52, 52, 108, 108, 108, 154, 154, 154, 104, 104, 104]

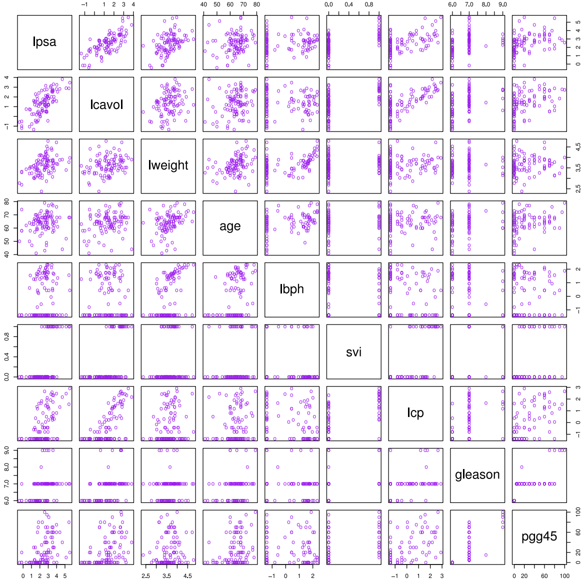
Postal Code Digit Data



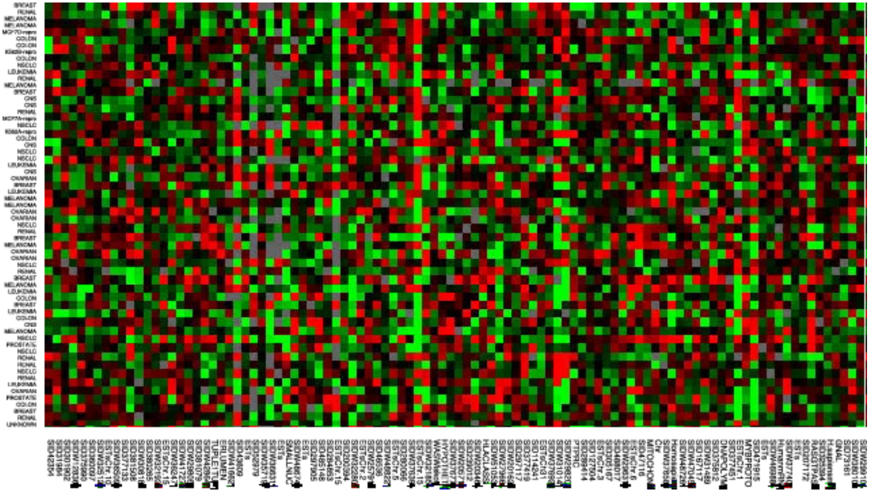
Document Topic Data



Prostate Cancer Data



Genomics Data



Supervised Learning

The Data (**training set**):

- An (unordered) set of examples
- Each example has a set of **features** or **attributes**
  - binary (ex: is the word "health" present in the document or not?)
  - categorical (ex: relationship status)
  - real-valued (ex: weight)
- A **target** or **label** or **output**
  - binary (ex: has cancer)
  - categorical (ex: preferred brand of paper towels)
  - real-valued (ex: traffic delay in minutes)

Supervised Learning

The Data (**training set**):

- An (unordered) set of examples
- Each example has a set of **features** or **attributes**
  - binary (ex: is the word "health" present in the document or not?)
  - categorical (ex: relationship status)
  - real-valued (ex: weight)
- A **target** or **label** or **output**
  - binary (ex: has cancer)
  - categorical (ex: preferred brand of paper towels)
  - real-valued (ex: traffic delay in minutes)

The Goal:

- Given the data above as the input to the learning algorithm,
- Produce a rule (**classifier** or **regressor**)
- that will map the feature values to the target
- The rule should work well on future, as-yet-unseen examples

Supervised Learning, Example

The Data (**training set**):

$$
X = \begin{array}{cccc} \text{sepal length (cm)} & \text{sepal width (cm)} & \text{petal length (cm)} & \text{petal width (cm)} \\ \end{array}
$$

$$
X = \left[\begin{array}{cccc} 5.1 & 3.5 & 1.4 & 0.2 \\ 7.0 & 3.2 & 4.7 & 1.4 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 6.3 & 3.3 & 6.0 & 2.5 \\ \vdots & \vdots & \vdots & \vdots \\ 4.9 & 3.0 & 1.4 & 0.2 \end{array}\right] \quad \begin{array}{c} \text{class} \\ Y = \left[\begin{array}{c} 0 \\ 1 \\ 1 \\ 2 \\ \vdots \\ 0 \end{array}\right] \end{array}
$$

where class 0 = Setosa, 1 = Versicolor, 2 = Virginica

Supervised Learning, Example

The Data (**training set**):

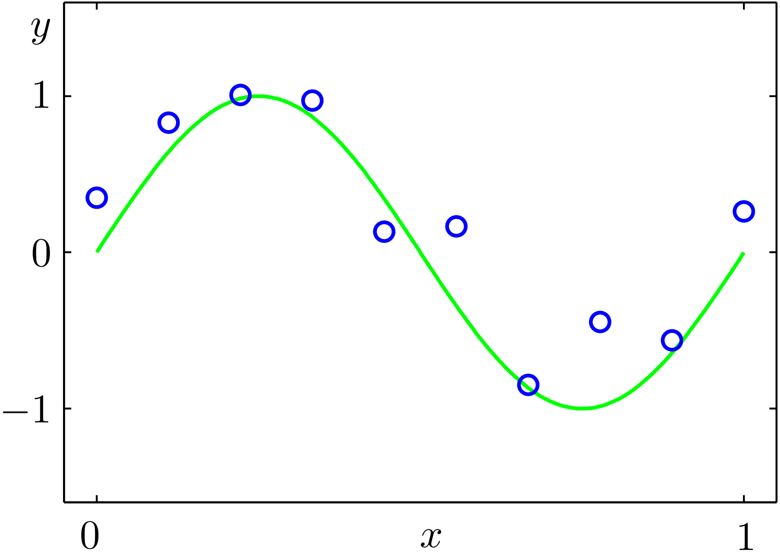
$$
X = \begin{array}{c} \text{sepal length (cm)} \quad \text{sepal width (cm)} \quad \text{petal length (cm)} \quad \text{petal width (cm)} \\ \left[\begin{array}{cccc} 5.1 & 3.5 & 1.4 & 0.2 \\ 7.0 & 3.2 & 4.7 & 1.4 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 6.3 & 3.3 & 6.0 & 2.5 \\ \vdots & \vdots & \vdots & \vdots \\ 4.9 & 3.0 & 1.4 & 0.2 \end{array}\right] \end{array}
\qquad
Y = \begin{array}{c} \text{class} \\ \left[\begin{array}{c} 0 \\ 1 \\ 1 \\ 2 \\ \vdots \\ 0 \end{array}\right] \end{array}
$$

where class 0 = Setosa, 1 = Versicolor, 2 = Virginica

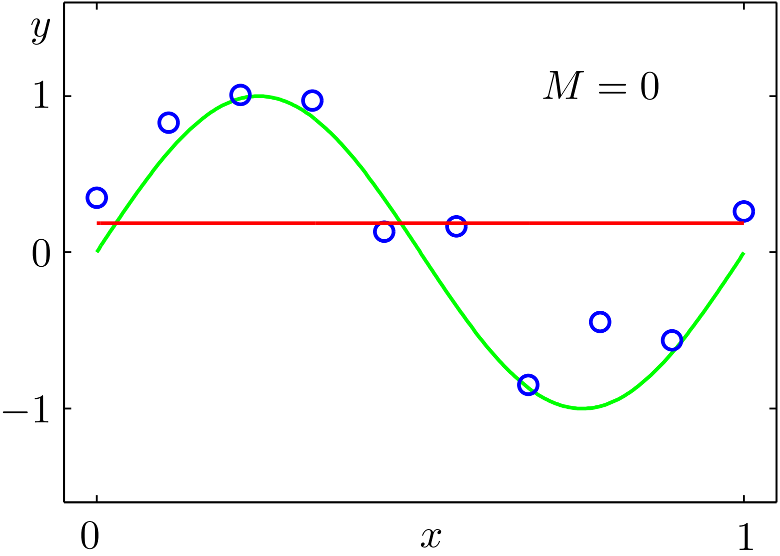
The Output (possible):

f(x):
```
If x(1) < 6 then y = 0
Else If x(3)+x(4) > 8 then y = 2
Else y = 1
```

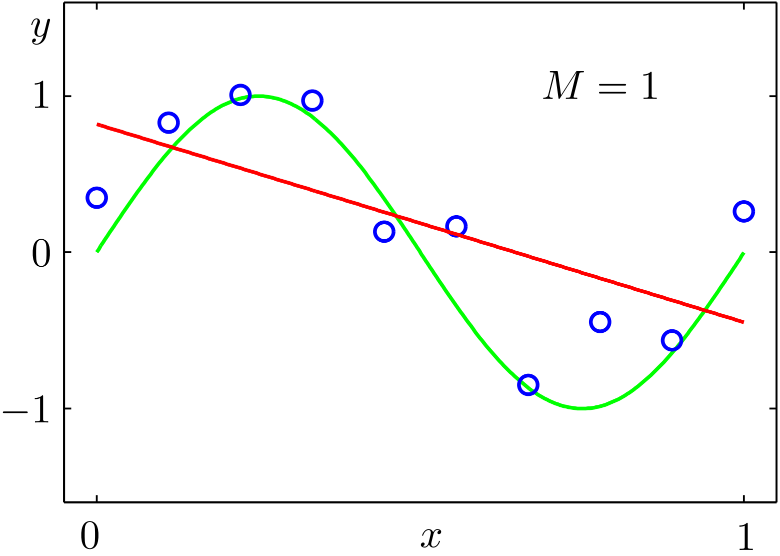
Simple Example



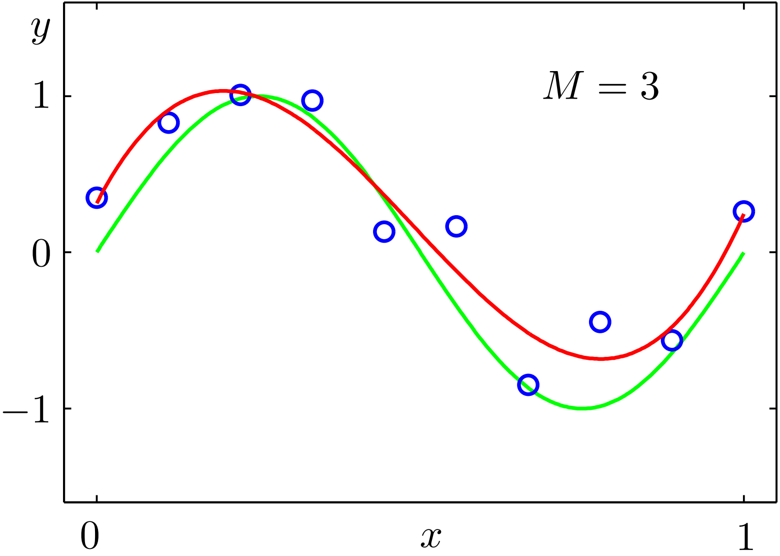
Simple Example



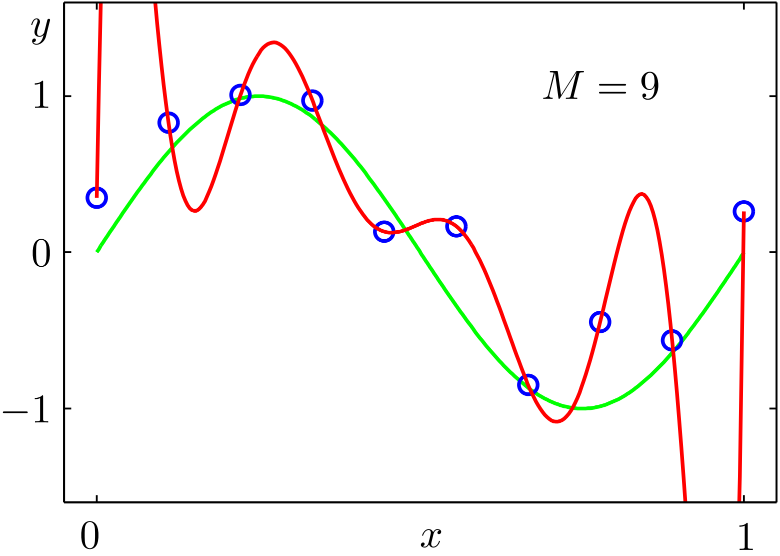
Simple Example



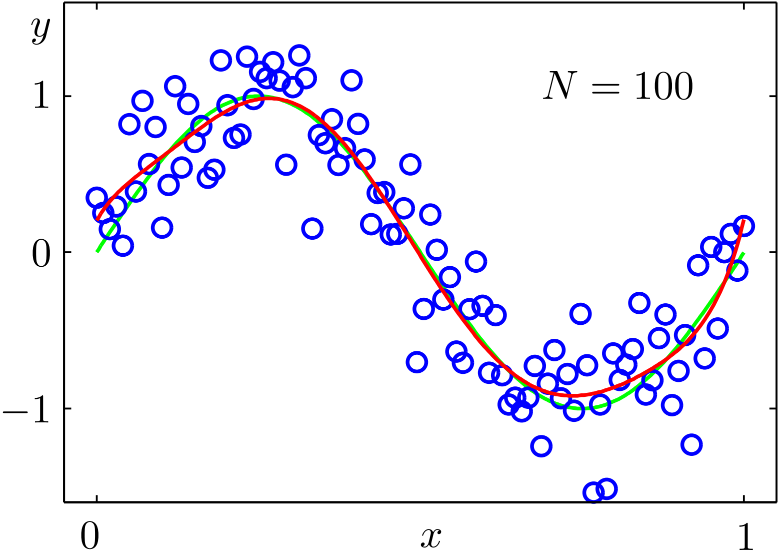
Simple Example



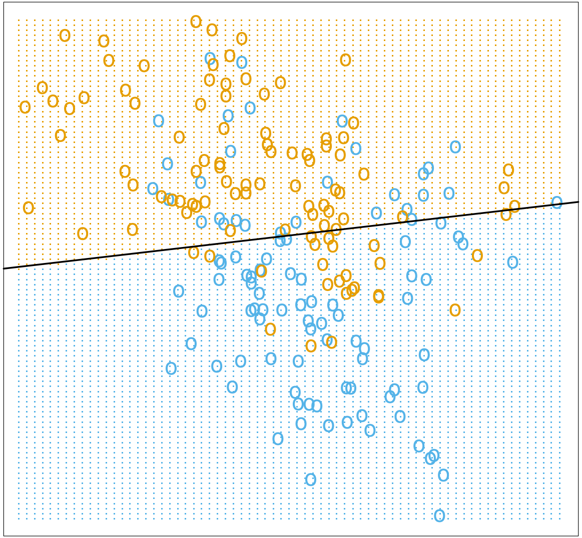
Simple Example



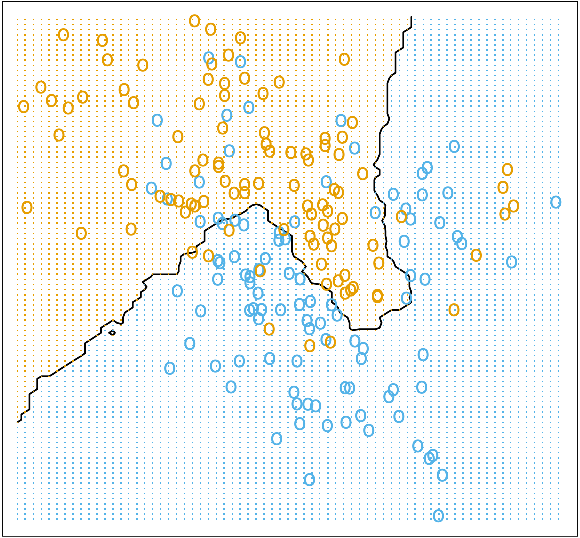
Simple Example



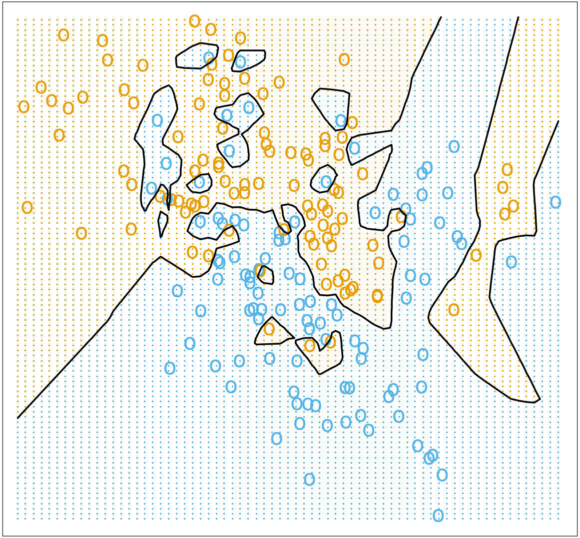
Another Example



Another Example



Another Example



Common Approach

Define the **hypothesis space**, the set of all possible output rules.
Examples:

- Degree-3 polynomials
- Linear dividing surfaces
- Other "crazy" things we'll discuss

Common Approach

Define the **hypothesis space**, the set of all possible output rules.
Examples:

- Degree-3 polynomials
- Linear dividing surfaces
- Other "crazy" things we'll discuss

Pick a definition of error
Examples:

- number incorrect (classification)
- sum of squared errors (regression)

Common Approach

Define the **hypothesis space**, the set of all possible output rules.
Examples:

- Degree-3 polynomials
- Linear dividing surfaces
- Other "crazy" things we'll discuss

Pick a definition of error
Examples:

- number incorrect (classification)
- sum of squared errors (regression)

Develop an algorithm to

- pick the member of the hypothesis space that
- minimizes the error on the training set
- perhaps slightly modified

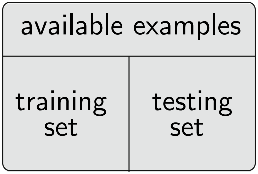
Testing Set

The error on the training set may not be indicative of the performance of the classifier or regressor

Testing Set

The error on the training set may not be indicative of the performance of the classifier or regressor

Thus, we often use a **testing set**



Feed the training set into the learning algorithm.
Find the error of the resulting rule on the testing set.

Why is this hard?

- High dimensional spaces are weird
- Irrelevant attributes
- Redundant attributes
- Missing attributes
- Attribute noise
- Label noise

Terms defined

features or **attributes**: measured values associated with each example
target or **label** or **output**: desired output of the rule for each example
classifier, **regressor**: the rule that maps the features to the label
hypothesis space: the space of all rules that the machine learning algorithm is considering
training set: the data used as input to the machine learning algorithm
testing test: the data used to check the performance of the resulting rule