



COMPUTER SCIENCE

INDIANA UNIVERSITY

School of Informatics and Computing
Bloomington

Machine Learning

CSCI-B 555

Fall 2015

Martha White

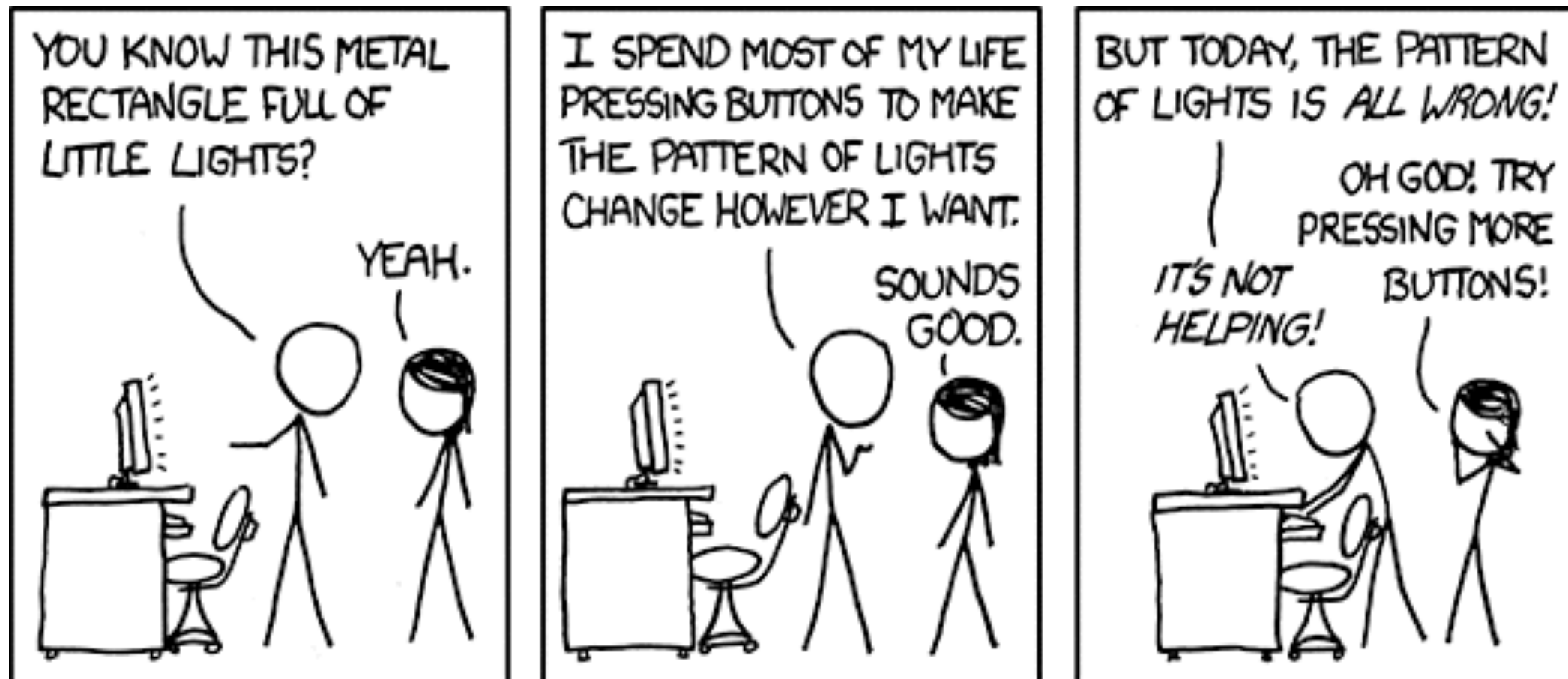
What is this course about?

- The world is full of information and data
- Much of that data is noisy and has an element of uncertainty
 - We have incomplete knowledge of the environment
 - Actions of other actors not provided
- **Goal:** understand machine learning algorithms to analyze this data by deriving them from the beginning
 - our focus will be on prediction (on new data)



What is this course about?

- **One line goal:** demystify the seemingly huge collection of machine learning algorithms by exploring their mathematical foundation



What is this course about?

- The focus is on fundamental data analysis/statistics concepts, and not necessarily on application of these algorithms
 - though you will get to do that too
- Application of algorithms is simpler when you understand their development and underlying assumptions



Basic information

Class meets:

Time: MW 11:15pm – 12:30pm

Place: Informatics East, Room 130

Instructor:

Martha White

Office: Lindley Hall 401E

Email: martha@indiana.edu

Web: <http://homes.soic.indiana.edu/martha/>

Office Hours:

Time: M 2:00pm-4:00pm (is this a good time?)
or by appointment

Place: Lindley Hall 401E

Class Web Site:

<https://iu.instructure.com/courses/1480234>



Associate Instructors (Teaching Assistants)

Shantanu Jain

Email: shajain@indiana.edu

Office: LH 401A

Times: M 1:30pm-3:00pm

Zeeshan Ali Sayyed

Email: zasayyed@indiana.edu

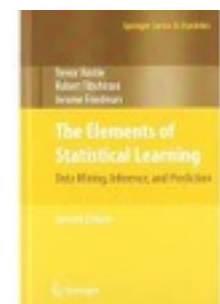
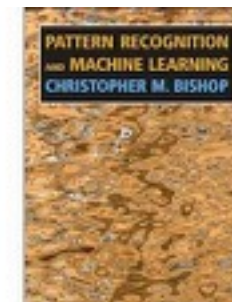
Office: LH 401A

Times: W 1:00pm-2:30pm



Textbook information

- **Main notes** provided in Canvas
 - written by Predrag Radivojac and myself (about 130 pages)
- **Additional reference material/Recommended readings:**
 - Pattern Recognition and Machine Learning by C. M. Bishop, Springer 2006.
 - The Elements of Statistical Learning by T. Hastie, R. Tibshirani, and J. Friedman, 2009



Grading policy

- 40%: Assignments (4), mostly mathematical with some small programming exercises
- 25%: Class project, (likely) solving a problem on Kaggle
- 25%: Final exam, December 16, 12:30 p.m. - 2:30 p.m.
- 10%: Thought questions (4), show that you're reading and thinking about the material



More on grading

- Top performers will get A+
 - This course is not curved; I decide the letter thresholds (Note that I don't enjoy having to assign you a grade, but it helps you learn)
- I'll try to maintain a distribution of scores; Canvas should show you at least your grades and the average
- **All assignments are individual**; you can chat about the material, but follow the “in your head” rule: leave with everything in your head, and not recorded/written down
- Acknowledge sources (classmates, websites, books) in your typed up documents
 - assignments are in LaTeX; I have posted examples/tutorials on Canvas



Submission policy

- Homework assignments are due on the date specified in Canvas
 - always Wednesday, at 11:59 p.m.
- Submit assignments, thought questions and project in Canvas
- Project and thought questions cannot be submitted late
- Late assignments will be accepted according to the following rules

– points	(on time)	}	recommended!
– points x 0.9	(1 day late)		
– points x 0.7	(2 days late)	}	not recommended!
– points x 0.5	(3 days late)		
– points x 0.3	(4 days late)		
– points x 0.1	(5 days late)		
– 0	(after 5 days)		
- There can be legitimate issues; if something is wrong, talk to me early



Academic honesty

- I have to report every cheating incident to the university
- Code of student rights, responsibilities and conduct: <http://www.indiana.edu/~code/>
 - e.g. Students are responsible to “facilitate the learning environment and the process of learning, including attending class regularly, completing class assignments, and coming to class prepared”.
- You are in graduate school, so I hope one of your priorities is to learn and become a more independent thinker

My expectations

- Basic mathematical skills
 - calculus
 - probabilities
 - linear algebra
- You are hardworking and motivated to learn (machine learning)
- You are motivated to succeed in class
- You are motivated to think beyond the material and ask open-ended questions (one of the goals for graduate school)



What to expect from this course

- I know you have expectation of me too
 - I will try to be transparent in marking and course choices
 - I am here to help you learn; I will treat you with respect and listen thoughtfully to your questions (I love to answer questions and give advice!)
 - Feel free to give me feedback (e.g., Miss Martha, you are talking too fast)
 - I will provide an mechanism for anonymous feedback
- This course will be quite mathematical, with derivations of details
 - sometimes I'll say “we need to eat our broccoli” or “pain is good for you”
- By the end of the course, you should have a good grasp of fundamental concepts in ML and algorithm derivation for ML



Thought questions

- Graduate school is about thinking and asking questions
 - maybe surprisingly, research is actually about asking good questions
 - a question does not have to have an answer, but it can be thought-provoking or make you think about a topic differently
- “Thought questions” correspond to (short) readings in the notes, and should demonstrate you’ve read and thought about the topics
- There are no stupid questions (so ask any questions in class/office hours/email), but “thought questions” are to demonstrate insight and so I have some requirements



Examples of “good” thought questions

- After reading about independence, I wonder how one could check in practice if two variables are independent, given a database of samples? Is this even possible?
- Where do the axioms of probability come from? Why does p have to satisfy these properties and what would it mean to add other ones? What would it mean to drop a required property on the probability function?
- “In the given example about expected utility, the parameters (+10 for correct and -20 for incorrect) seems a bit arbitrary. What are common decision making tools practitioners use once outcomes are known?” — student from B554
- “Why not always use gradient descent, as we did to learn the hyperparameters for Bayesian linear models? Why do we ever use EM?” — student from B554
- Why is PCA so widely used? It seems simple and I would not expect it to be able to encode complex properties.



Examples of “bad” thought questions

- I don't understand linear regression. Could you explain it again? (i.e. a request for me to explain something, without any insight)
- Derive the maximum likelihood approach for a Gaussian. (i.e., an exercise question from a textbook)
- What is the difference between a probability mass function and a probability density function? (i.e., a question that could easily be answered from reading the definitions in the notes)
- How are Boltzmann machines and feedforward neural networks different? (i.e., again a definition)
 - But the following modification would be good: “I can see that Boltzmann machines and feedforward neural networks are different, in that the first is undirected and the second directed. How does this difference impact modeling properties and accuracy of estimation in practice?”



What is machine learning?

- We could label it in many different ways including
 - Data-driven approach to artificial intelligence
 - Applied statistics
- Some keywords associated with machine learning
 - Prediction, probability, samples, data, function, optimization, stochastic, ...



ICML 2015 word cloud



Where did machine learning come from?

- Let's first step back to the goals of artificial intelligence
- Many original AI approaches were expert-based
 - logic approaches for theorem proving
 - expert systems
- Machine learning arose as a data-based approach to solve artificial intelligence problems
 - why the shift? increased computation, availability of data and efficacy of data driven approaches (largely driven by availability of data)

What are the ultimate goals?

- The focus for many ML researchers has shifted from AI towards generally solving important (practical) problems
 - computer vision, speech recognition, clustering, modeling temporal data, ...
- This includes a focus on understanding intrinsic properties of a learning problem
 - is it difficult to learn (e.g., NP-hard)?
 - how can it be formulated in a precise way? (e.g., convex objective for sparse coding, preference for “simpler” hypotheses)
 - how many samples are needed to learn the model (epsilon) accurately?
 - how well does the learned model generalize to new samples?

How do learning problems differ?

- They can be categorized across several dimensions
 - **Control versus prediction:** though a control algorithm will likely use predictions to improve decision-making (e.g., reinforcement learning)
 - **Supervised and unsupervised:** supervised learning is for prediction, unsupervised learning is usually for visualization or representation learning; some algorithms combine these two important components
 - **How they are used:** empower decision-making of end user OR autonomously control system
 - ... there are more, but these are some main ones
- **Algorithms** also differ in many ways, even for similar problems
 - e.g., process data incrementally (as stream) or in batch; low computation (or memory) versus heavy computation (or memory); approximate or exact



Let's look at some fun examples!

- Commute times (mostly independent data, iid data)
- Weather prediction (temporally connected data)
 - machine learning is often used for time series, but in the specific case of weather, mostly expert models appear to be used (for now...)
- Robotics and octopus arm simulator (machine learning for control)
 - as a caveat, we will not look at control algorithms; however, their development uses the fundamental concepts in this course
 - check out my Stochastic optimization for ML seminar in Spring for control

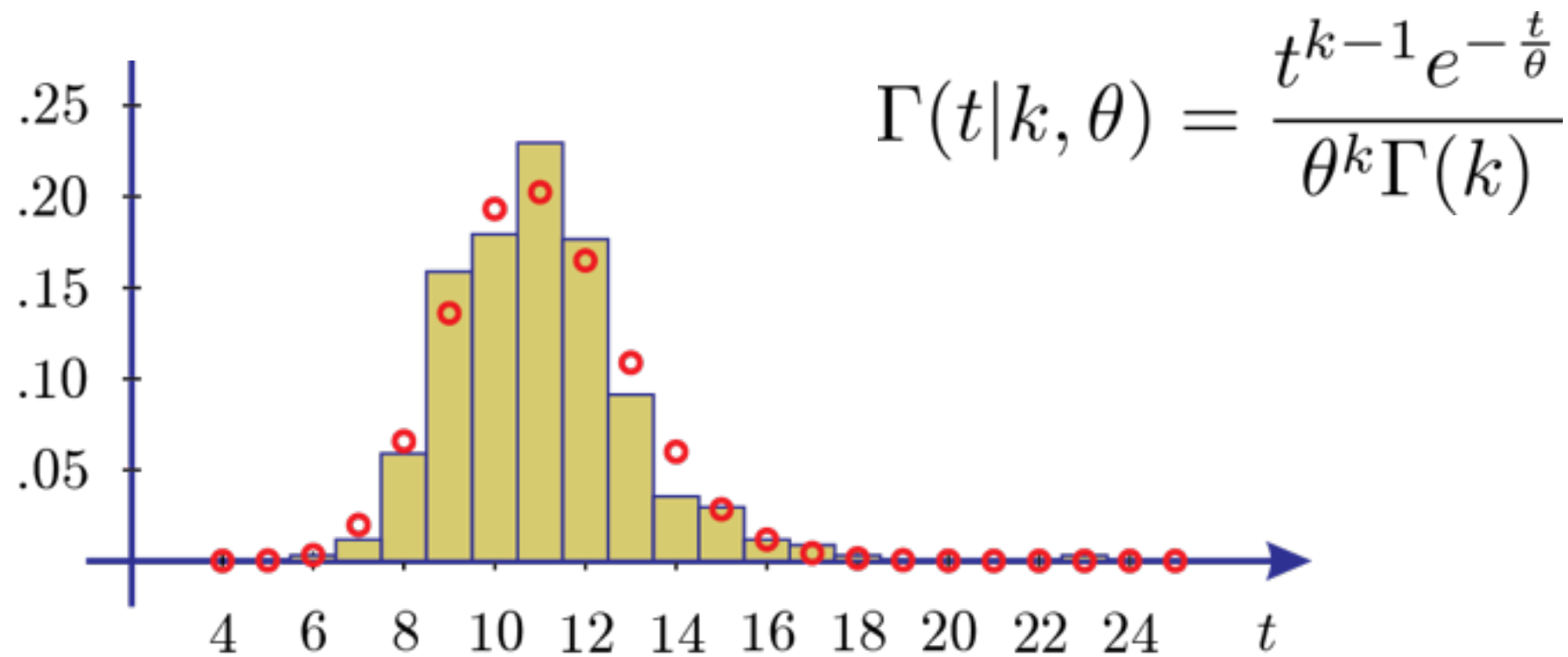


Commute times



LILAC PETAL	9	—	—	10	8
10	10	11	11	9	12
11	12	13	14	11	10
12	13	14	15	12	11
13	14	15	16	13	12
14	15	16	17	14	13
15	16	17	18	15	14
16	17	18	19	16	15
17	18	19	20	17	16
18	19	20	21	18	17
19	20	21	22	19	18
20	21	22	23	20	19
21	22	23	24	21	20
22	23	24	25	22	21
23	24	25	26	23	22
24	25	26	27	24	23
25	26	27	28	25	24
26	27	28	29	26	25
27	28	29	30	27	26
28	29	30	31	28	27
29	30	31	32	29	28
30	31	32	33	30	29
31	32	33	34	31	30
32	33	34	35	32	31
33	34	35	36	33	32
34	35	36	37	34	33
35	36	37	38	35	34
36	37	38	39	36	35
37	38	39	40	37	36
38	39	40	41	38	37
39	40	41	42	39	38
40	41	42	43	40	39
41	42	43	44	41	40
42	43	44	45	42	41
43	44	45	46	43	42
44	45	46	47	44	43
45	46	47	48	45	44
46	47	48	49	46	45
47	48	49	50	47	46
48	49	50	51	48	47
49	50	51	52	49	48
50	51	52	53	50	49
51	52	53	54	51	50
52	53	54	55	52	51
53	54	55	56	53	52
54	55	56	57	54	53
55	56	57	58	55	54
56	57	58	59	56	55
57	58	59	60	57	56
58	59	60	61	58	57
59	60	61	62	59	58
60	61	62	63	60	59
61	62	63	64	61	60
62	63	64	65	62	61
63	64	65	66	63	62
64	65	66	67	64	63
65	66	67	68	65	64
66	67	68	69	66	65
67	68	69	70	67	66
68	69	70	71	68	67
69	70	71	72	69	68
70	71	72	73	70	69
71	72	73	74	71	70
72	73	74	75	72	71
73	74	75	76	73	72
74	75	76	77	74	73
75	76	77	78	75	74
76	77	78	79	76	75
77	78	79	80	77	76
78	79	80	81	78	77
79	80	81	82	79	78
80	81	82	83	80	79
81	82	83	84	81	80
82	83	84	85	82	81
83	84	85	86	83	82
84	85	86	87	84	83
85	86	87	88	85	84
86	87	88	89	86	85
87	88	89	90	87	86
88	89	90	91	88	87
89	90	91	92	89	88
90	91	92	93	90	89
91	92	93	94	91	90
92	93	94	95	92	91
93	94	95	96	93	92
94	95	96	97	94	93
95	96	97	98	95	94
96	97	98	99	96	95
97	98	99	100	97	96
98	99	100	101	98	97
99	100	101	102	99	98
100	101	102	103	100	99
101	102	103	104	101	100
102	103	104	105	102	101
103	104	105	106	103	102
104	105	106	107	104	103
105	106	107	108	105	104
106	107	108	109	106	105
107	108	109	110	107	106
108	109	110	111	108	107
109	110	111	112	109	108
110	111	112	113	110	109
111	112	113	114	111	110
112	113	114	115	112	111
113	114	115	116	113	112
114	115	116	117	114	113
115	116	117	118	115	114
116	117	118	119	116	115
117	118	119	120	117	116
118	119	120	121	118	117
119	120	121	122	119	118
120	121	122	123	120	119
121	122	123	124	121	120
122	123	124	125	122	121
123	124	125	126	123	122
124	125	126	127	124	123
125	126	127	128	125	124
126	127	128	129	126	125
127	128	129	130	127	126
128	129	130	131	128	127
129	130	131	132	129	128
130	131	132	133	130	129
131	132	133	134	131	130
132	133	134	135	132	131
133	134	135	136	133	132
134	135	136	137	134	133
135	136	137	138	135	134
136	137	138	139	136	135
137	138	139	140	137	136
138	139	140	141	138	137
139	140	141	142	139	138
140	141	142	143	140	139
141	142	143	144	141	140
142	143	144	145	142	141
143	144	145	146	143	142
144	145	146	147	144	143
145	146	147	148	145	144
146	147	148	149	146	145
147	148	149	150	147	146
148	149	150	151	148	147
149	150	151	152	149	148
150	151	152	153	150	149
151	152	153	154	151	150
152	153	154	155	152	151
153	154	155	156	153	152
154	155	156	157	154	153
155	156	157	158	155	154
156	157	158	159	156	155
157	158	159	160	157	156
158	159	160	161	158	157
159	160	161	162	159	158
160	161	162	163	160	159
161	162	163	164	161	160
162	163	164	165	162	161
163	164	165	166	163	162
164	165	166	167	164	163
165	166	167	168	165	164
166	167	168	169	166	165
167	168	169	170	167	166
168	169	170	171	168	167
169	170	171	172	169	168
170	171	172	173	170	169
171	172	173	174	171	170
172	173	174	175	172	171
173	174	175	176	173	172
174	175	176	177	174	173
175	176	177	178	175	174
176	177	178	179	176	175
177	178	179	180	177	176
178	179	180	181	178	177
179	180	181	182	179	178
180	181	182	183	180	179
181	182	183	184	181	180
182	183	184	185	182	181
183	184	185	186	183	182
184	185	186	187	184	183
185	186	187	188	185	184
186	187	188	189	186	185
187	188	189	190	187	186
188	189	190	191	188	187
189	190	191	192	189	188
190	191	192	193	190	189
191	192	193	194	191	190
192	193	194	195	192	191
193	194	195	196	193	192
194	195	196	197	194	193
195	196	197	198	195	194
196	197	198	199	196	195
197	198	199	200	197	196
198	199	200	201	198	197
199	200	201	202	199	198
200	201	202	203	200	199
201	202	203	204	201	200
202	203	204	205	202	201
203	204	205	206	203	202
204	205	206	207	204	203
205	206	207	208	205	204
206	207	208	209	206	205
207	208	209	210	207	206
208	209	210	211	208	207
209	210	211	212	209	208
210	211	212	213	210	209
211	212	213	214	211	210
212	213	214	215	212	211
213	214	215	216	213	212
214	215	216	217	214	213
215	216	217	218	215	214
216	217	218	219	216	215
217	218	219	220	217	216
218	219	220	221	218	217
219	220	221	222	219	218
220	221	222	223	220	219
221	222	223	224	221	220
222	223	224	225	222	221
223	224	225	226	223	222
224	225	226	227	224	223
225	226	227	228	225	224
226	227	228	229	226	225
227	228	229	230	227	226
228	229	230	231	228	227
229	230	231	232	229	228
230	231	232	233	230	229
231	232	233	234	231	230
232	233	234	235	232	231
233	234	235	236	233	232
234	235	236	237	234	233
235	236	237	238	235	234
236	237	238	239	236	235
237	238	239	240	237	236
238	239	240	241	238	237
239	240	241	242	239	238
240	241	242	243	240	239
241	242	243	244	241	240
242	243	244	245	242	241
243	244	245	246	243	242
244	245	246	247	244	243
245	246	247	248	245	244
246	247	248	249	246	245
247	248	249	250	247	246
248	249	250	251	248	247
249	250	251	252	249	248
250	251	252	253	250	249
251	252	253	254	251	250
252	253	254	255	252	251
253	254	255	256	253	252
254	255	256	257	254	

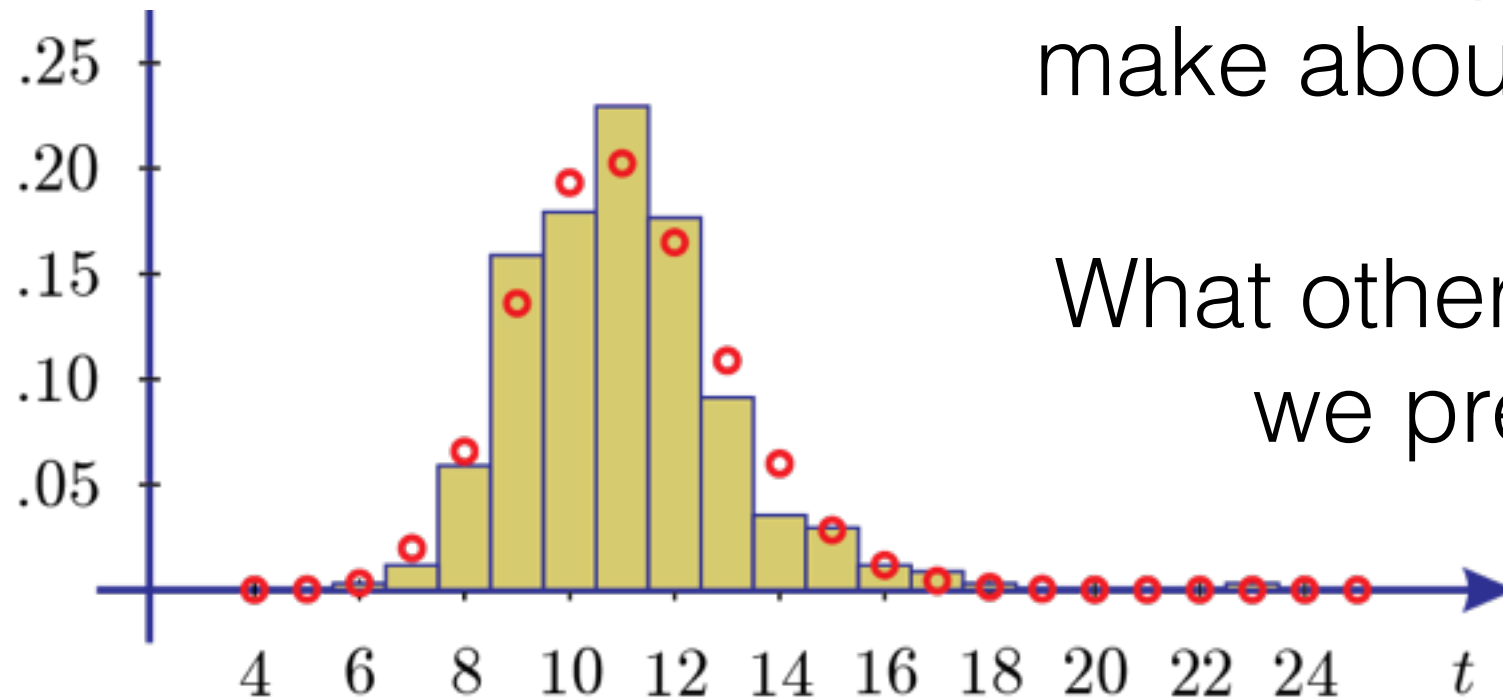
Commute times (2)



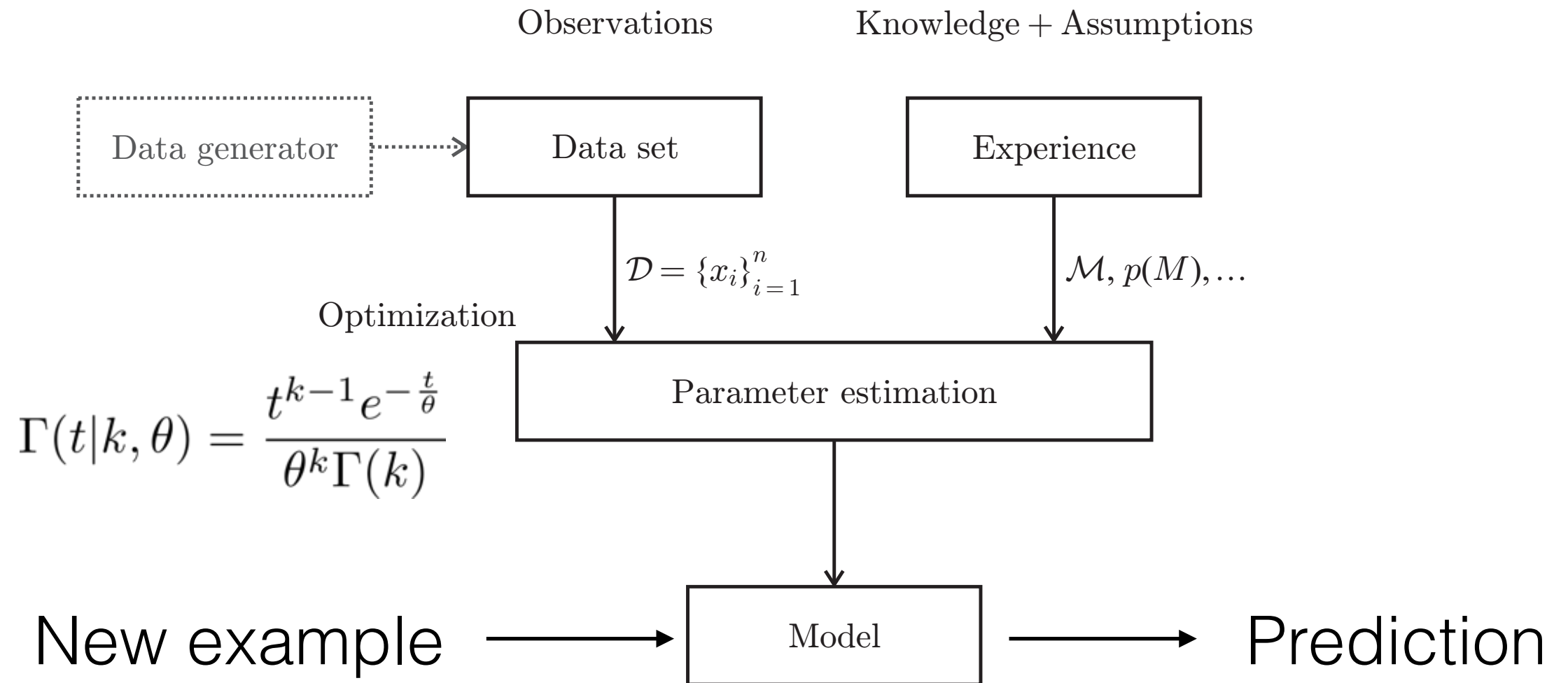
Commute times (3)

What assumptions can we make about the data?

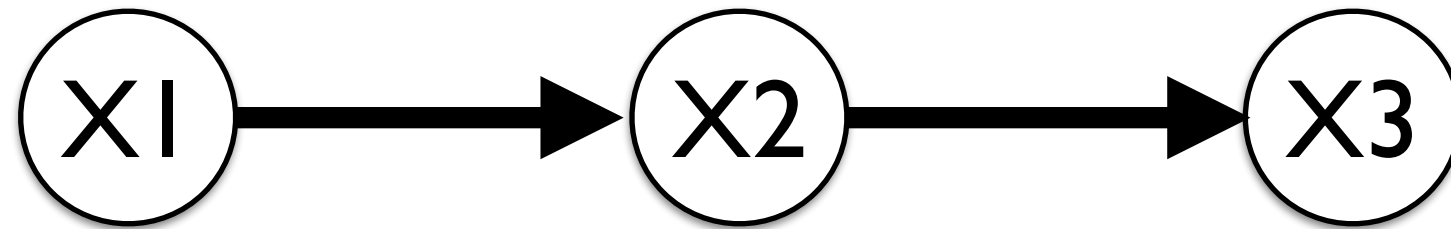
What other ways can we predict?



Summarized flow



Weather prediction (time series)

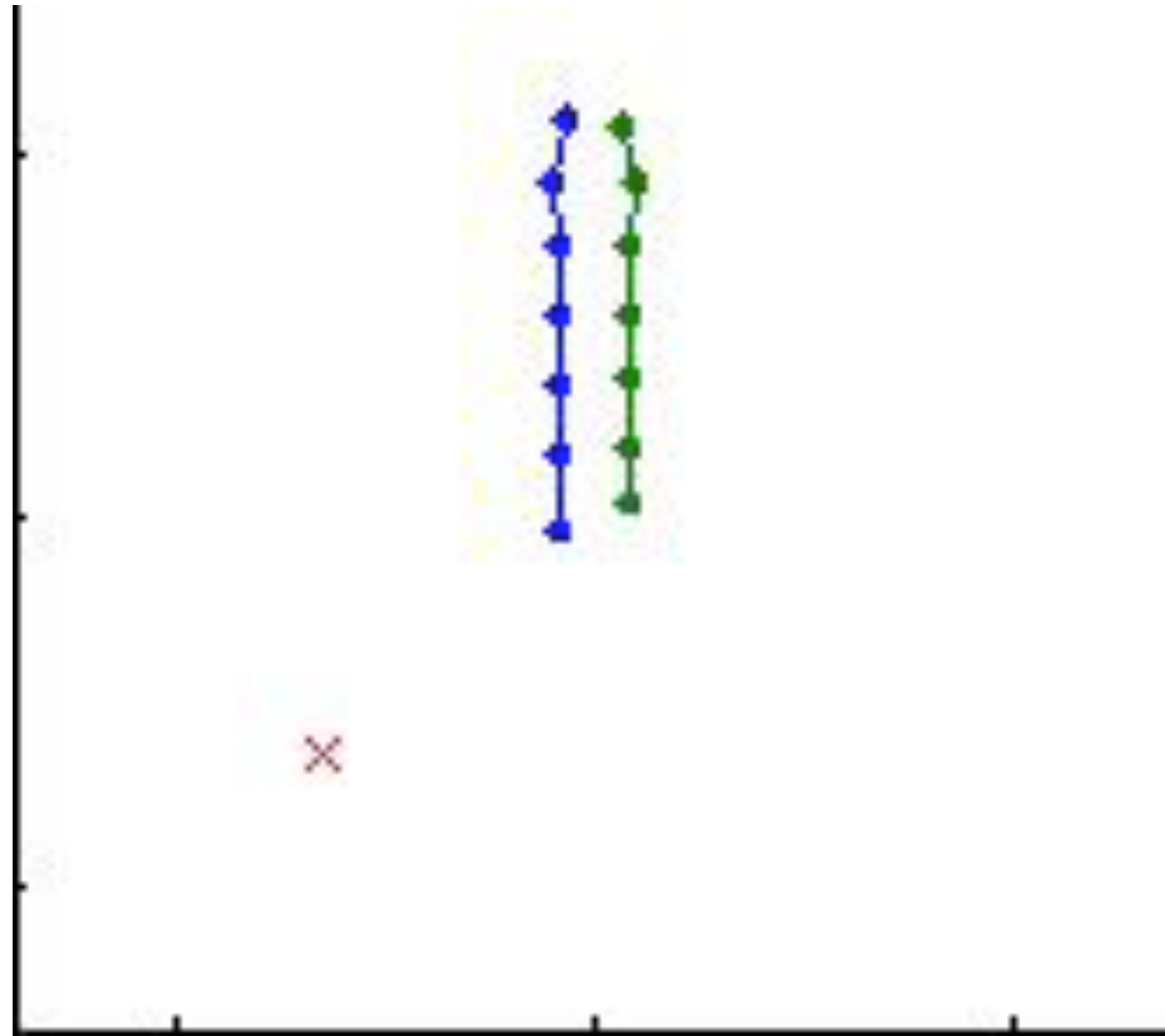


- Imagine we want to predict the probability of rain tomorrow, 2 days from now, 3 days, ...
- One common strategy for time series is to use the last p points as features to predict the next point, 2 points into future, etc.
- What other strategies can you imagine?
- How do you predict a probability value, rather than say a binary value (0 or 1) or a real value?
 - Hint: these are things we will learn

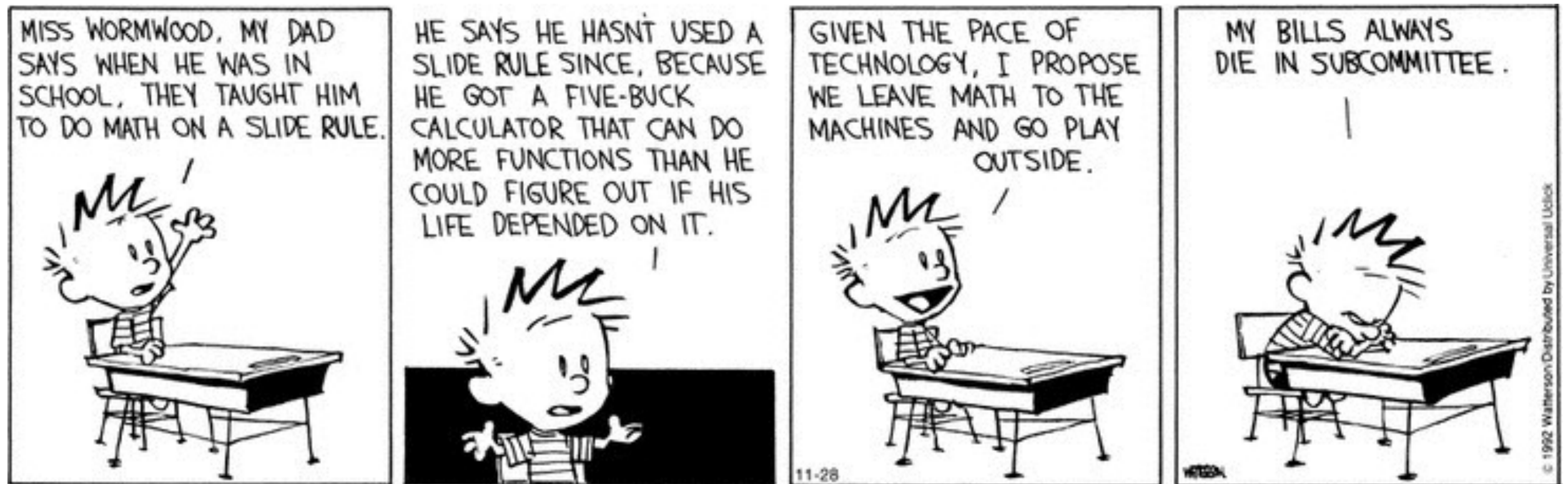
Simple robot (control)

- <https://www.youtube.com/watch?v=CkTftoNFeGY>
 - time 2:50 to 4:10
- An example of a simple implementation of RL algorithms (not heavily engineered, heavily data-based)
 - something you could likely implement after an intro RL course
 - goal to work towards more autonomous systems
- Though simple, does input a relatively large number of sensors

Octopus arm (control)



Next: crash course in probability



- Probabilities underly much of machine learning
 - enable precise modeling of uncertainty

Topic overview

- Parameter estimation and prediction problems (Chapters 2 and 3)
 - the core background for modeling in machine learning
- Linear regression (Chapter 3)
- Generalized linear models (Chapter 4)
- Linear classifiers (Chapter 5)
- Representations for ML (Chapter 6)
 - help make linear predictors more powerful (e.g., neural networks)
- Empirical evaluation of learning algorithms (Chapter 7)
- ... and any other topics listed on the syllabus if we have time (e.g., boosting)

