Introduction to Learning

cs540 section 2
Louis Oliphant
oliphant@cs.wisc.edu

Announcements

- Homework 3
  - Typo on 2b. The horizontal line should be just above the last sentence

$$\frac{\neg(N \Rightarrow I)}{\neg I \Rightarrow C}$$
$$C$$

Deduction and Induction
(dictionary.com)

- Deduction
  - The process of reasoning in which a conclusion follows necessarily from the stated premises; inference by reasoning from the general to the specific.
- Induction
  - The process of deriving general principles from particular facts or instances.

Deduction and Induction
(dictionary.com)

- Deduction – everything we have done so far
  - Uninformed Search
  - Informed Search
  - Game Playing
  - Logical Reasoning and inference
- Induction – The remainder of the course
  - Decision Trees
  - K-means clustering
  - Neural Networks
  - Statistical Models (a little deduction here too)

Inductive Learning

"deriving general principles from particular facts"

Particular Facts:
1+1=2
3+7=10
13+15=28
...

General Principles:
??? how to add
odd plus odd is even

Types of Inductive Learning

- Three types of learning that depend upon the type of Feedback given to the learner:
  - unsupervised learning
    - no feedback
  - Reinforcement learning
    - feedback is in the form of "good job"
    - Learners decisions are Reinforced or not
  - supervised learning
    - feedback specifies the correct answer for given input

Unsupervised Learning

- No supervision – no teacher saying what is right or wrong
- What learning can be done?
  - Find patterns in the particular examples
- Clustering
- Self-Organizing Maps

Reinforcement Learning

- Lazy Teacher – Doesn't tell you the right answer, just gives reinforcement
- Most Learning in real life
  - burning finger on hot stove
  - 

% written on assignment
  - Lack of a tip for poor service at restaurant

Supervised Learning

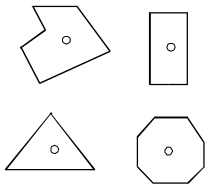
- Teacher tells you the correct answer for a given example
  - Lots of this in "real" life too -- "real" meaning "in the classroom"
- Supervised Learning is what will be covered in this course (except for a day on clustering)

Supervised Learning (the setup)

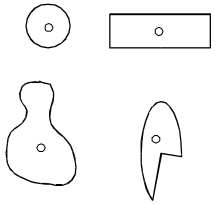
- Given: a bunch of examples with their correct answer
- Do: Learn a model that can give the correct answer for each of the examples. And can generalize well to unseen examples.

The Setup cont.

Positive Examples

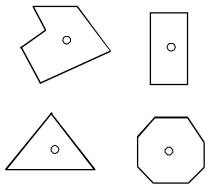


Negative Examples

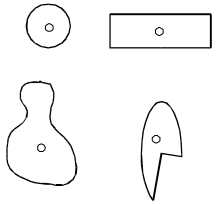


The Setup cont.

Positive Examples



Negative Examples



Test Example?

Negative Example

What is the general principle?
Polygon with yellow circle in it.

Fixed Length Feature Vectors

- Convert each example into a fixed set of features
- Types of Features
  - Discrete
    - Boolean, Nominal, Ordered
  - Continuous or Real valued
- Convert the output (correct answer) to a similar feature
  - Discrete
    - Classification Problem
  - Continuous
    - Regression Problem

Sample Data sets
(in fixed length feature vector form)

Main Actor/Actress	Opening Day	Category	Director	Writer	Oscar		Opening Gross
Daniel Radcliffe	11/14/01	Adventure	Columbus	Steven Kloves	No		> 2 million
Marlon Brando	03/15/72	Drama	Coppola	Puzo	Yes		<2 million
Mark Hamill	05/25/77	Action	Lucas	Lucas	Yes		> 2 million
Keanu Reeves	03/31/99	Action	Wachowski	Wachowski	Yes		> 2 million
Peter Sellers	01/29/64	Comedy	Kubrick	George	No		<2 million
Jodie Foster	02/14/91	Crime	Demme	Harris	Yes		>2 million

Day	Outlook	Temperature	Humidity	Wind	>1,000?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Learning a Model

- How to capture general principles from data?
- answer: use a model that can represent the data
- One of the simplest models is Nearest Neighbor
- To make a prediction on any new example, just find the nearest neighbor in the dataset and predict the same outcome for the new example

- We need a measure of distance between any two examples

What would the prediction be using Nearest Neighbor

- Which data point is closest to the new example?
  Outlook=Overcast, Temperature=Hot, Humidity=Normal, Wind=Strong

Day	Outlook	Temperature	Humidity	Wind	>1,000?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

What would the prediction be using Nearest Neighbor

- Which data point is closest to the new example?
  Outlook=Overcast, Temperature=Hot, Humidity=Normal, Wind=Strong – So Predict >1,000.

Day	Outlook	Temperature	Humidity	Wind	>1,000?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Distance Metric in Nearest Neighbor

- Use Euclidean Distance
  - For Discrete features try to convert to integers
  - Otherwise just use "0" if they match and "1" if they don't

Day	Outlook	Temperature	Humidity	Wind	>1,000?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Distance Metric in Nearest Neighbor

- Use Euclidean Distance
  - For Discrete features try to convert to integers
  - Otherwise just use "0" if they match and "1" if they don't

Day	Outlook	Temperature	Humidity	Wind	>1,000?
1	2	2	1	0	0
2	2	2	1	1	0
3	1	2	1	0	1
4	0	1	1	0	1
5	0	0	0	0	1
6	0	0	0	1	0
7	1	0	0	1	1
8	2	1	1	0	0
9	2	0	0	0	1
10	0	1	0	0	1
11	2	1	0	1	1
12	1	1	1	1	1
13	1	2	0	0	1
14	0	1	1	1	0

Prediction using Nearest Neighbor

- Anytime you need to predict the result for a new example, just calculate the distance between the new example and each example in data set
  Outlook=1, Temperature=1, Humidity=0, Wind=1
- Take the outcome from the example with the smallest distance

Day	Outlook	Temperature	Humidity	Wind	>1,000?	Distance
1	2	2	1	0	0	4
2	2	2	1	1	0	3
3	1	2	1	0	1	3
4	0	1	1	0	1	3
5	0	0	0	0	1	3
6	0	0	0	1	0	2
7	2	0	0	1	1	2
8	2	1	1	0	0	3
9	2	0	0	0	1	3
10	0	1	0	0	1	2
11	2	0	0	1	1	2
12	1	1	1	1	1	1
13	1	2	0	0	1	2
14	0	1	1	1	0	2

K – Nearest Neighbor

- Instead of taking just the Nearest Neighbor, take the K closest neighbors
- Take a majority vote among those K neighbors

K-NN and K=5

Day	Outlook	Temperature	Humidity	Wind	>1,000?	Distance
1	2	2	1	0	0	1
2	2	2	1	1	0	2
3	1	2	1	0	1	2
4	0	1	1	0	1	4
5	0	0	0	0	1	4
6	0	0	0	1	0	3
7	2	0	0	1	1	4
8	2	1	1	0	0	2
9	2	0	0	0	1	3
10	0	1	0	0	1	3
11	2	0	0	1	1	4
12	1	1	1	1	1	4
13	1	2	0	0	1	1
14	0	1	1	1	0	5

Other Variations on Nearest Neighbors

- Weighted Nearest Neighbor
  - Take a weighted vote where examples that have the smallest distance get the largest vote
    - K can be the size of the dataset, no need to limit it
- Scaled Features in Nearest Neighbor
  - Scale more relevant features to have more weight when deciding distance

Evaluating The Model

- How well does this method work?
- Evaluate the model on some unseen, "test" examples and see how many predications are correct
  - Why not evaluate on the training set?
    Think if 1-NN results

Conclusion

- Types of Inductive Learning
- Converting examples to Fixed Length Feature Vectors
- Nearest Neighbor Models
  - K-NN, weighted nearest neighbor, scaling features
- Evaluating the model