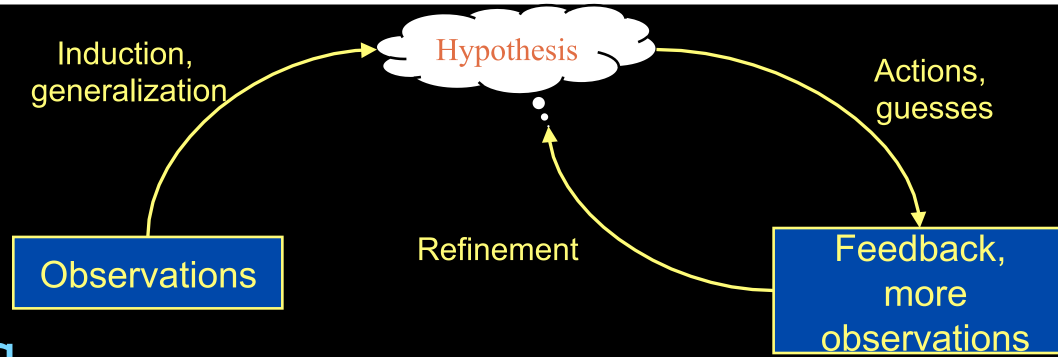
CS 461: Machine Learning
Lecture 2

Dr. Kiri Wagstaff
wkiri@wkiri.com

Today’s Topics

- Review and Reading Questions
- Homework 1
- Data Representation (Features)
- Decision Trees
- Evaluation
- Weka

Review



- Machine Learning
  - Computers learn from their past experience
- Inductive Learning
  - Generalize to new data
- Supervised Learning
  - Training data: <*x*, *g(x)*> pairs
  - Known label or output value for training data
  - Classification and regression
- Instance-Based Learning
  - 1-Nearest Neighbor
  - k-Nearest Neighbors

Reading Questions

- Introduction / Machine Learning (Ch. 1)
  - Classification: What is a discriminant?
  - Regression: to train an autonomous car to predict what angle to turn the steering wheel, where could the training data come from?
- Supervised Learning (Ch. 2.1, 2.4-2.9)
  - Is the most specific hypothesis S a member of the version space? Why or why not?
  - What happens if the true concept C is not in the version space?
  - What is Occam's Razor?

Homework 1

- Solution/Discussion

Data Representation: Which Features?

Decision Trees

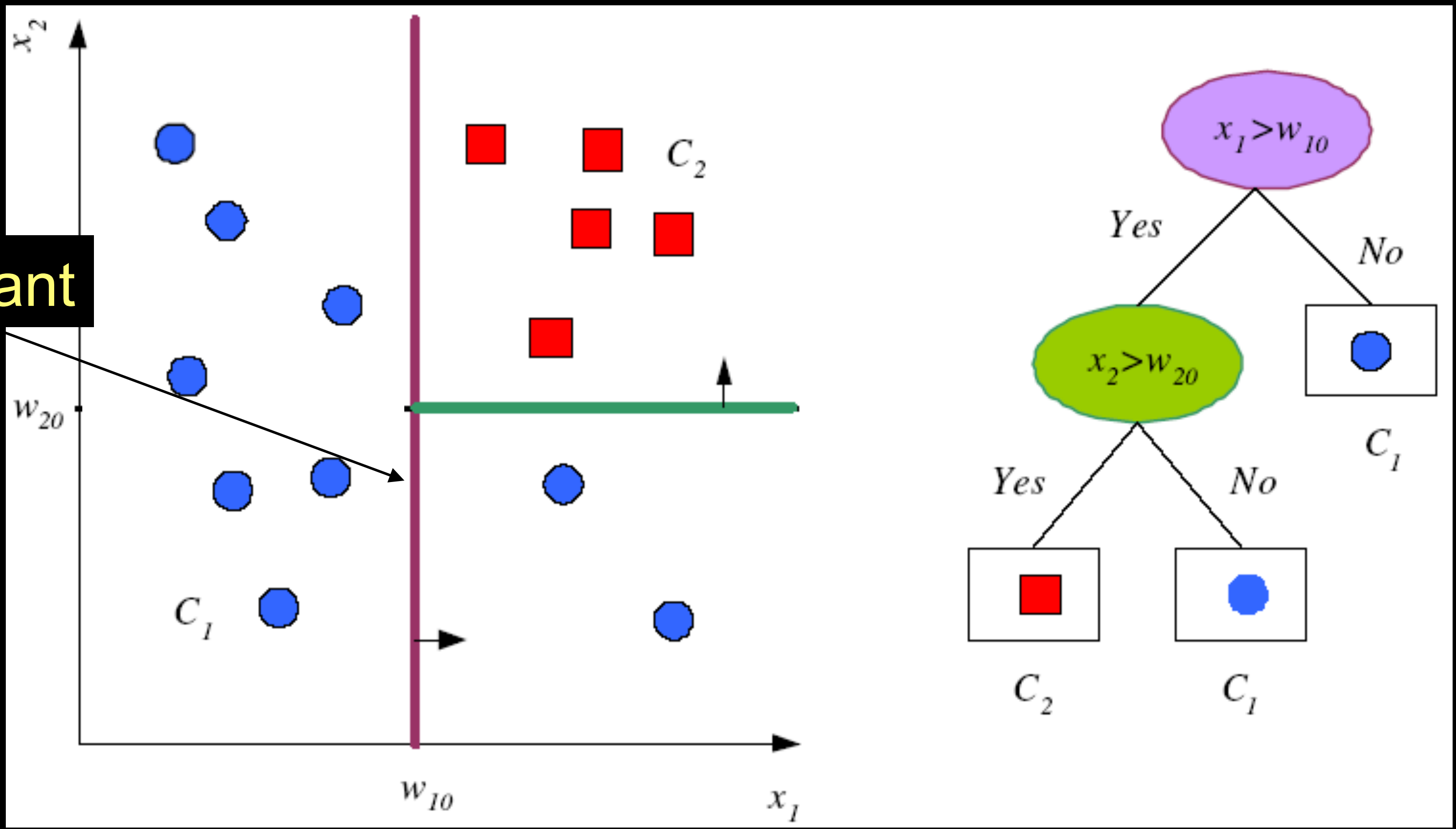
Chapter 9

Decision Trees

- Parametric method
- PredictionWorks
  - "Increasing customer loyalty through targeted marketing"
- Decision Tree Interactive Demo

(Hyper-)Rectangles == Decision Tree

discriminant

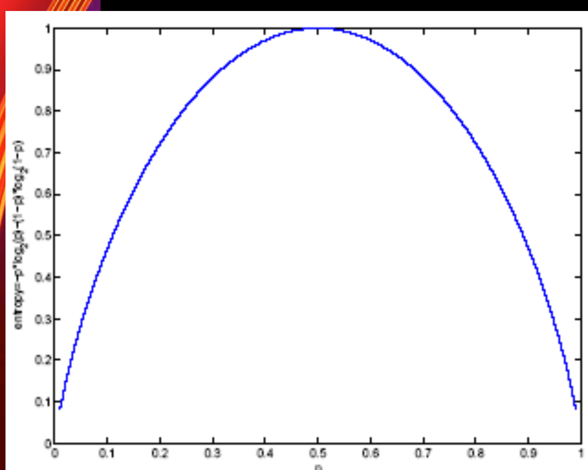


 [Alpaydin 2004 © The MIT Press]

Measuring Impurity

$$\hat{P}(C_i \mid \mathbf{x}, m) \equiv p_m^i = \frac{N_m^i}{N_m}$$

1. Impurity = error using majority label $\quad I_m = 1 - \max_i (p_m^i)$
   - After a split
     $$I'_m = \sum_{j=1}^{n} \frac{N_{mj}}{N_m} 1 - \max_i (p_{mj}^i)$$
2. More sensitive: use entropy
   - For node m, N_m instances reach m, N^i_m belong to C_i



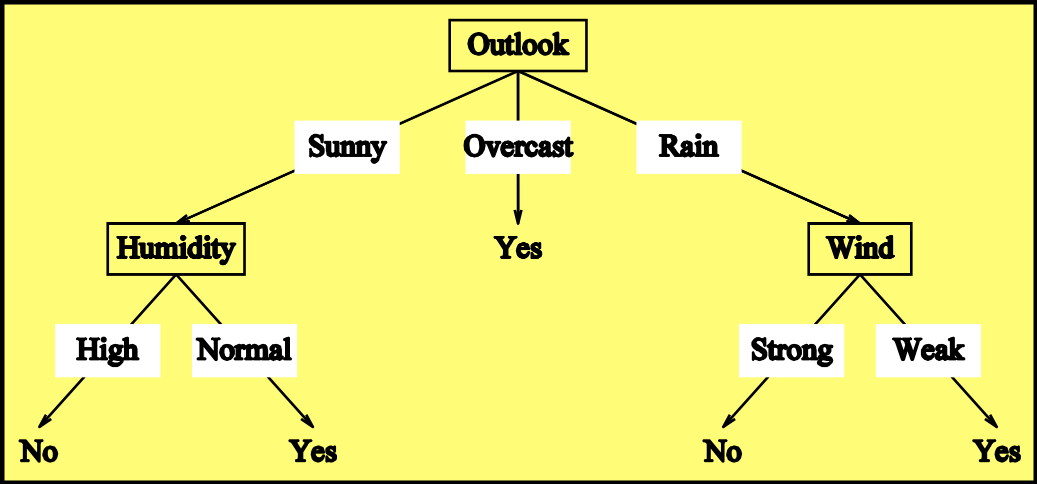
- Node m is pure if p^i_m is 0 or 1
- Entropy: $$I_m = -\sum_{i=1}^{K} p_m^i \log_2 p_m^i$$
- After a split: $$I'_m = -\sum_{j=1}^{n} \frac{N_{mj}}{N_m} \sum_{i=1}^{K} p_{mj}^i \log_2 p_{mj}^i$$

Should we play tennis?

Outlook	Temperature	Humidity	Wind	PlayTennis
Training Sets				
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

[Tom Mitchell]

How well does it generalize?



Outlook	Temperature	Humidity	Wind	PlayTennis
Testing Examples				
Sunny	Mild	Normal	Weak	Yes
Overcast	Hot	High	Strong	Yes
Sunny	Mild	High	Strong	No
Rain	Hot	High	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	High	Strong	Yes

 [Tom Dietterich, Tom Mitchell]

Decision Tree Construction Algorithm

```
GenerateTree(𝒳)
    If NodeEntropy(𝒳)< θ_I /* eq. 9.3
      Create leaf labelled by majority class in 𝒳
      Return
    i ← SplitAttribute(𝒳)
    For each branch of x_i
      Find 𝒳_i falling in branch
      GenerateTree(𝒳_i)
SplitAttribute(𝒳)
    MinEnt← MAX
    For all attributes i = 1,...,d
        If x_i is discrete with n values
          Split 𝒳 into 𝒳_1,...,𝒳_n by x_i
          e ← SplitEntropy(𝒳_1,...,𝒳_n) /* eq. 9.8 */
          If e<MinEnt MinEnt ← e; bestf ← i
        Else /* x_i is numeric */
          For all possible splits
              Split 𝒳 into 𝒳_1,𝒳_2 on x_i
              e←SplitEntropy(𝒳_1,𝒳_2)
              If e<MinEnt MinEnt ← e; bestf ← i
    Return bestf
```


Evaluating a Single Algorithm

Chapter 14

Measuring Error

	Predicted class	
True Class	Yes	No
Yes	TP: True Positive	FN: False Negative
No	FP: False Positive	TN: True Negative

Breast Cancer

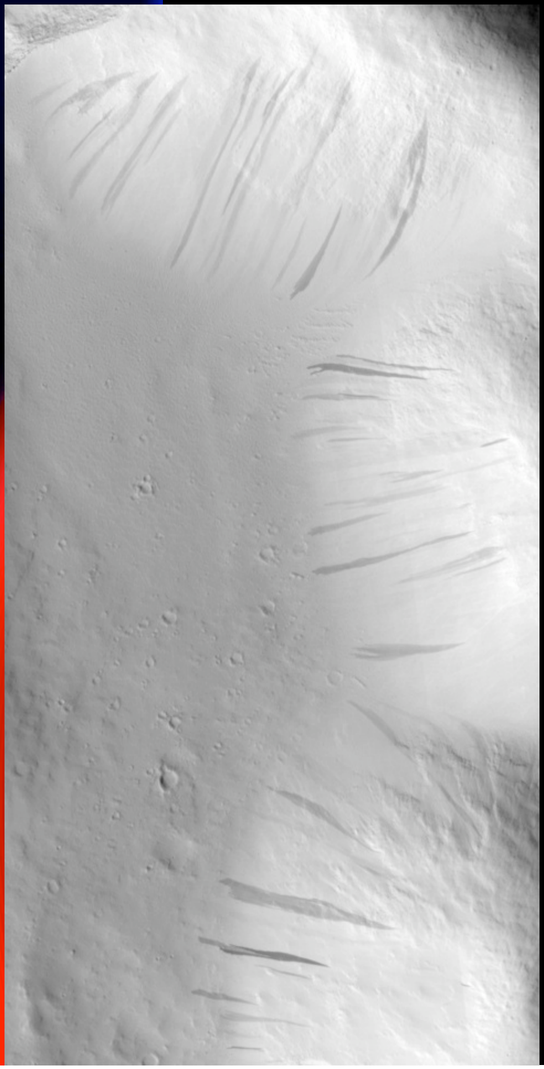
	Survived	Died
Survived	9	3
Died	4	4

Iris

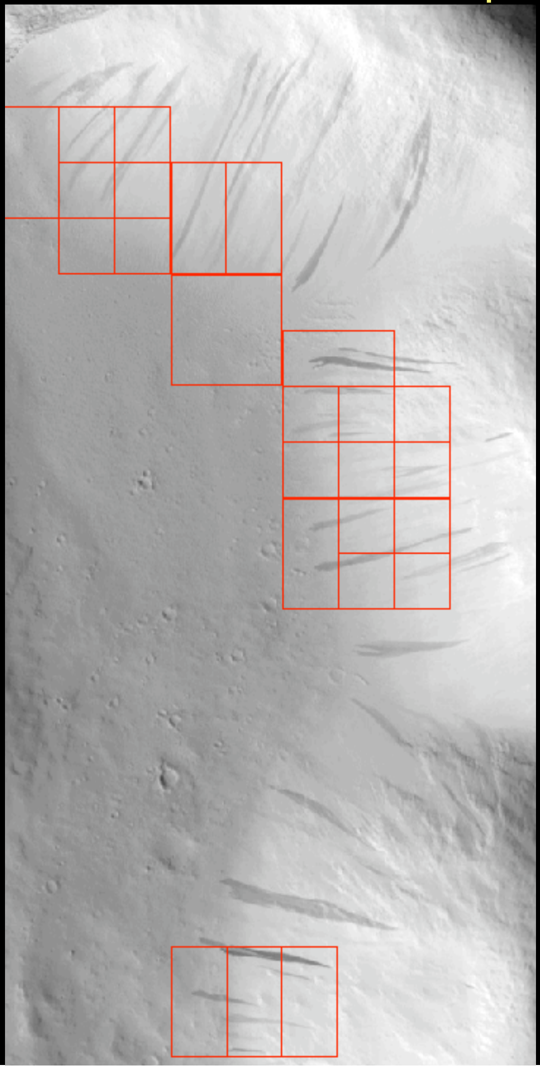
	Setosa	Versicolor	Virginica
Setosa	10	0	0
Versicolor	0	10	0
Virginica	0	1	9

Example: Finding Dark Slope Streaks on Mars

Marte Vallis,
HiRISE on MRO



Output of statistical
landmark detector: top 10%



Results

TP: 13

FP: 1

FN: 16

Recall = 13/29 = 45%

Precision = 13/14 = 93%

Evaluation Methodology

- Metrics: What will you measure?
  - Accuracy / error rate
  - TP/FP, recall, precision…
- What train and test sets?
  - Cross-validation
  - LOOCV
- What baselines (or competing methods)?
- Are the results significant?

Baselines

- Simple rule
- “Straw man”
- If you can’t beat this… don’t bother!

- Imagine:

vs.

Weka Machine Learning Library

Weka Explorer's Guide

Homework 2

Summary: What You Should Know

- Supervised Learning
  - Representation: features available
- Decision Trees
  - Hierarchical, non-parametric, greedy
    - Nodes: test a feature value
    - Leaves: classify items (or predict values)
  - Minimize impurity (%error or entropy)
- Evaluation
  - (10-fold) Cross-Validation
  - Confusion Matrix

Next Time

- Reading:
  - Decision Trees
    (read Ch. 9.1-9.4)
  - Evaluation
    (read Ch. 14.1-14.4)
  - Weka Manual
    (read p. 25-27, 33-35, 39-42, 48-49)

- Questions to answer from the reading:
  - Posted on the website (calendar)
  - Three volunteers: Lewis, Natalia, and T.K.