Decision Trees and Ensemble Learning

Shiwali Mohan

Computer Science and Engineering
University of Michigan

May 25, 2012

Outline

Decision Trees Review

- What is a decision tree?

Decision Trees Review

- What is a decision tree?
  - a function that takes as input a vector of attribute values and returns a output value
  - assumption: inputs are discrete and output is boolean
  - given a set of examples: value assignment to attributes in input and corresponding output value

Decision Trees Review

- What is a decision tree?
  - a function that takes as input a vector of attribute values and returns a output value
  - assumption: inputs are discrete and output is boolean
  - given a set of examples: value assignment to attributes in input and corresponding output value
- Entropy of a random variable?

Decision Trees Review

- What is a decision tree?
  - a function that takes as input a vector of attribute values and returns a output value
  - assumption: inputs are discrete and output is boolean
  - given a set of examples: value assignment to attributes in input and corresponding output value
- Entropy of a random variable?
  - measurement of uncertainity of a random variable
  - acquisition of information corresponds to reduction in entropy
  - $$H(V) = -\sum_k P(v_k) \log_2 P(v_k)$$

Decision Trees Review

- What is a decision tree?
  - a function that takes as input a vector of attribute values and returns a output value
  - assumption: inputs are discrete and output is boolean
  - given a set of examples: value assignment to attributes in input and corresponding output value
- Entropy of a random variable?
  - measurement of uncertainity of a random variable
  - acquisition of information corresponds to reduction in entropy
  - $$H(V) = -\sum_k P(v_k) \log_2 P(v_k)$$
  - $H(fair - coin)$?

Decision Trees Review

- What is a decision tree?
  - a function that takes as input a vector of attribute values and returns a output value
  - assumption: inputs are discrete and output is boolean
  - given a set of examples: value assignment to attributes in input and corresponding output value
- Entropy of a random variable?
  - measurement of uncertainity of a random variable
  - acquisition of information corresponds to reduction in entropy
  - $$H(V) = -\sum_k P(v_k) \log_2 P(v_k)$$
  - $H(fair - coin)$?
  - $H(tail - only)$?

Decision Trees Review

- What is a decision tree?
  - a function that takes as input a vector of attribute values and returns a output value
  - assumption: inputs are discrete and output is boolean
  - given a set of examples: value assignment to attributes in input and corresponding output value
- Entropy of a random variable?
  - measurement of uncertainity of a random variable
  - acquisition of information corresponds to reduction in entropy
  - $$H(V) = -\sum_k P(v_k) \log_2 P(v_k)$$
  - $H(fair - coin)$?
  - $H(tail - only)$?
- $B(q)$

Decision Trees Review

- What is a decision tree?
  - a function that takes as input a vector of attribute values and returns a output value
  - assumption: inputs are discrete and output is boolean
  - given a set of examples: value assignment to attributes in input and corresponding output value
- Entropy of a random variable?
  - measurement of uncertainity of a random variable
  - acquisition of information corresponds to reduction in entropy
  - $$H(V) = -\sum_k P(v_k) \log_2 P(v_k)$$
  - $H(fair - coin)$?
  - $H(tail - only)$?
- $B(q)$
  - entropy of a Boolean random variable that is true with a probability q
  - $B(q) = -(q \log_2 q + (1-q) \log_2 (1-q))$

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is the entropy of the result attribute of the whole example set?

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is the entropy of the result attribute of the whole example set?
  - $B(\frac{5}{12})$

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is the entropy of the result attribute of the whole example set?
  - $B(\frac{5}{12})$
  - $-\frac{5}{12}\log_2\frac{5}{12} - \frac{7}{12}\log_2\frac{27}{12} = 0.9803$

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is first attribute to split on?

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is first attribute to split on?
  - That attribute that gives us the most information gain, reduces the entropy by the largest amount.

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is first attribute to split on?
  - That attribute that gives us the most information gain, reduces the entropy by the largest amount.
  - $$Remainder(A) = \sum_{k=1}^{d} \frac{p_k + n_k}{p + n} B(\frac{p_k}{p_k + n_k})$$

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is first attribute to split on?
  - That attribute that gives us the most information gain, reduces the entropy by the largest amount.
  - $$Remainder(A) = \sum_{k=1}^{d} \frac{p_k + n_k}{p + n} B(\frac{p_k}{p_k + n_k})$$
  - $$Gain(A) = B(\frac{p_k}{p_k + n_k}) - Remainder(A)$$

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is first attribute to split on?

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is first attribute to split on?
  - Remainder(Type)?

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is first attribute to split on?
  - Remainder(Type)?
    - $\frac{2}{12}B(\frac{1}{2}) + \frac{3}{12}B(\frac{2}{3}) + \frac{4}{12}B(\frac{1}{2}) + \frac{3}{12}B(0) = 0.6371$

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is first attribute to split on?
  - Remainder(Type)?
    - $\frac{2}{12}B(\frac{1}{2}) + \frac{3}{12}B(\frac{2}{3}) + \frac{4}{12}B(\frac{1}{2}) + \frac{3}{12}B(0) = 0.6371$
  - Remainder(Price)?

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is first attribute to split on?
  - Remainder(Type)?
    - $\frac{2}{12}B(\frac{1}{2}) + \frac{3}{12}B(\frac{2}{3}) + \frac{4}{12}B(\frac{1}{2}) + \frac{3}{12}B(0) = 0.6371$
  - Remainder(Price)?
    - $\frac{5}{12}B(\frac{2}{5}) + \frac{3}{12}B(\frac{2}{3}) + \frac{4}{12}B(\frac{1}{4}) = 0.90448$
  - Remainder(Rush)?

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is first attribute to split on?
  - Remainder(Type)?
    - $\frac{2}{12}B(\frac{1}{2}) + \frac{3}{12}B(\frac{2}{3}) + \frac{4}{12}B(\frac{1}{2}) + \frac{3}{12}B(0) = 0.6371$
  - Remainder(Price)?
    - $\frac{5}{12}B(\frac{2}{5}) + \frac{3}{12}B(\frac{2}{3}) + \frac{4}{12}B(\frac{1}{4}) = 0.90448$
  - Remainder(Rush)?
    - $\frac{6}{12}B(\frac{1}{2}) + \frac{6}{12}B(\frac{2}{3}) = 0.9591$

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is first attribute to split on?
  - Remainder(Type)?
    - $\frac{2}{12}B(\frac{1}{2}) + \frac{3}{12}B(\frac{2}{3}) + \frac{4}{12}B(\frac{1}{2}) + \frac{3}{12}B(0) = 0.6371$
  - Remainder(Price)?
    - $\frac{5}{12}B(\frac{2}{5}) + \frac{3}{12}B(\frac{2}{3}) + \frac{4}{12}B(\frac{1}{4}) = 0.90448$
  - Remainder(Rush)?
    - $\frac{6}{12}B(\frac{1}{2}) + \frac{6}{12}B(\frac{2}{3}) = 0.9591$
  - Gain(Type)? Gain(Price)? Gain(Rush)?

Learning to Lunch

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is first attribute to split on?
  - Remainder(Type)?
    - $\frac{2}{12}B(\frac{1}{2}) + \frac{3}{12}B(\frac{2}{3}) + \frac{4}{12}B(\frac{1}{2}) + \frac{3}{12}B(0) = 0.6371$
  - Remainder(Price)?
    - $\frac{5}{12}B(\frac{2}{5}) + \frac{3}{12}B(\frac{2}{3}) + \frac{4}{12}B(\frac{1}{4}) = 0.90448$
  - Remainder(Rush)?
    - $\frac{6}{12}B(\frac{1}{2}) + \frac{6}{12}B(\frac{2}{3}) = 0.9591$
  - Gain(Type)? Gain(Price)? Gain(Rush)?
    - 0.3432, 0.07585, 0.0212

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is the weight of examples?

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is the weight of examples?
  - $\frac{1}{12}$

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is the weight of examples?
  - $\frac{1}{12}$
- What is the initial decision stump (use information gain)?

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is the weight of examples?
  - $\frac{1}{12}$
- What is the initial decision stump (use information gain)?
  - Food type

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is the weight of examples?
  - $\frac{1}{12}$
- What is the initial decision stump (use information gain)?
  - Food type
- What is the classification error?

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is the weight of examples?
  - $\frac{1}{12}$
- What is the initial decision stump (use information gain)?
  - Food type
- What is the classification error?
  - Sushi (good), Hamburger (good), Pizza (good), Gruel (bad)
  - $\frac{4}{12} = 0.333$

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is the weight of examples?
  - $\frac{1}{12}$
- What is the initial decision stump (use information gain)?
  - Food type
- What is the classification error?
  - Sushi (good), Hamburger (good), Pizza (good), Gruel (bad)
  - $\frac{4}{12} = 0.333$
- What is the weight of the hypothesis?

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- What is the weight of examples?
  - $\frac{1}{12}$
- What is the initial decision stump (use information gain)?
  - Food type
- What is the classification error?
  - Sushi (good), Hamburger (good), Pizza (good), Gruel (bad)
  - $\frac{4}{12} = 0.333$
- What is the weight of the hypothesis?
  - $z[1] = \log(\frac{1-0.333}{0.333}) = 0.695$

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index
High	Sushi	Hurry	Good	x1
Low	Pizza	Hurry	Good	x2
Medium	Hamburger	Hurry	Good	x3
Medium	Pizza	Relax	Good	x4
High	Gruel	Relax	Bad	x5
High	Hamburger	Relax	Good	x6
Low	Gruel	Hurry	Bad	x7
Low	Sushi	Relax	Bad	x8
High	Pizza	Hurry	Bad	x9
High	Hamburger	Hurry	Bad	x10
Medium	Hamburger	Hurry	Bad	x11
Low	Gruel	Relax	Bad	x12

- How will the weight change?

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index	Weight
High	Sushi	Hurry	Good	x1	0.0625
Low	Pizza	Hurry	Good	x2	0.0625
Medium	Hamburger	Hurry	Good	x3	0.0625
Medium	Pizza	Relax	Good	x4	0.0625
High	Gruel	Relax	Bad	x5	0.0625
High	Hamburger	Relax	Good	x6	0.0625
Low	Gruel	Hurry	Bad	x7	0.0625
Low	Sushi	Relax	Bad	x8	0.125
High	Pizza	Hurry	Bad	x9	0.125
High	Hamburger	Hurry	Bad	x10	0.125
Medium	Hamburger	Hurry	Bad	x11	0.125
Low	Gruel	Relax	Bad	x12	0.0625

- How will the weight change?

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index	Weight
High	Sushi	Hurry	Good	x1	0.0625
Low	Pizza	Hurry	Good	x2	0.0625
Medium	Hamburger	Hurry	Good	x3	0.0625
Medium	Pizza	Relax	Good	x4	0.0625
High	Gruel	Relax	Bad	x5	0.0625
High	Hamburger	Relax	Good	x6	0.0625
Low	Gruel	Hurry	Bad	x7	0.0625
Low	Sushi	Relax	Bad	x8	0.125
High	Pizza	Hurry	Bad	x9	0.125
High	Hamburger	Hurry	Bad	x10	0.125
Medium	Hamburger	Hurry	Bad	x11	0.125
Low	Gruel	Relax	Bad	x12	0.0625

- How will the weight change?
  - $w[j] \leftarrow w[j].\frac{error}{(1-error)}$
  - NORMALIZE(w)

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index	Weight
High	Sushi	Hurry	Good	x1	0.0625
Low	Pizza	Hurry	Good	x2	0.0625
Medium	Hamburger	Hurry	Good	x3	0.0625
Medium	Pizza	Relax	Good	x4	0.0625
High	Gruel	Relax	Bad	x5	0.0625
High	Hamburger	Relax	Good	x6	0.0625
Low	Gruel	Hurry	Bad	x7	0.0625
Low	Sushi	Relax	Bad	x8	0.125
High	Pizza	Hurry	Bad	x9	0.125
High	Hamburger	Hurry	Bad	x10	0.125
Medium	Hamburger	Hurry	Bad	x11	0.125
Low	Gruel	Relax	Bad	x12	0.0625

- How will the weight change?
  - $w[j] \leftarrow w[j].\frac{error}{(1-error)}$
  - NORMALIZE(w)
- New entropy of the sample

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index	Weight
High	Sushi	Hurry	Good	x1	0.0625
Low	Pizza	Hurry	Good	x2	0.0625
Medium	Hamburger	Hurry	Good	x3	0.0625
Medium	Pizza	Relax	Good	x4	0.0625
High	Gruel	Relax	Bad	x5	0.0625
High	Hamburger	Relax	Good	x6	0.0625
Low	Gruel	Hurry	Bad	x7	0.0625
Low	Sushi	Relax	Bad	x8	0.125
High	Pizza	Hurry	Bad	x9	0.125
High	Hamburger	Hurry	Bad	x10	0.125
Medium	Hamburger	Hurry	Bad	x11	0.125
Low	Gruel	Relax	Bad	x12	0.0625

- How will the weight change?
  - $w[j] \leftarrow w[j].\frac{error}{(1-error)}$
  - NORMALIZE(w)
- New entropy of the sample
  - $-0.3125 \log_2(0.3125) - 0.6875 \log_2(0.6875)$

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index	Weight
High	Sushi	Hurry	Good	x1	0.0625
Low	Pizza	Hurry	Good	x2	0.0625
Medium	Hamburger	Hurry	Good	x3	0.0625
Medium	Pizza	Relax	Good	x4	0.0625
High	Gruel	Relax	Bad	x5	0.0625
High	Hamburger	Relax	Good	x6	0.0625
Low	Gruel	Hurry	Bad	x7	0.0625
Low	Sushi	Relax	Bad	x8	0.125
High	Pizza	Hurry	Bad	x9	0.125
High	Hamburger	Hurry	Bad	x10	0.125
Medium	Hamburger	Hurry	Bad	x11	0.125
Low	Gruel	Relax	Bad	x12	0.0625

- How will the weight change?
  - $w[j] \leftarrow w[j].\frac{error}{(1-error)}$
  - NORMALIZE(w)
- New entropy of the sample
  - $-0.3125\log_2(0.3125) - 0.6875\log_2(0.6875)$
- Genrate the next stump.

Ensemble of Stumps

Price	Food Type	Rush?	Results	Index	Weight
High	Sushi	Hurry	Good	x1	0.0625
Low	Pizza	Hurry	Good	x2	0.0625
Medium	Hamburger	Hurry	Good	x3	0.0625
Medium	Pizza	Relax	Good	x4	0.0625
High	Gruel	Relax	Bad	x5	0.0625
High	Hamburger	Relax	Good	x6	0.0625
Low	Gruel	Hurry	Bad	x7	0.0625
Low	Sushi	Relax	Bad	x8	0.125
High	Pizza	Hurry	Bad	x9	0.125
High	Hamburger	Hurry	Bad	x10	0.125
Medium	Hamburger	Hurry	Bad	x11	0.125
Low	Gruel	Relax	Bad	x12	0.0625

- How will the weight change?
  - $w[j] \leftarrow w[j].\frac{error}{(1-error)}$
  - NORMALIZE(w)
- New entropy of the sample
  - $-0.3125\log_2(0.3125) - 0.6875\log_2(0.6875)$
- Genrate the next stump.
  - $$Remainder(A) = \sum_{k=1}^{d} \frac{p_k + n_k}{p+n} B(\frac{p_k}{p_k + n_k})$$

PAC Hypothesis

PAC Hypothesis

- Size of the hypothesis space of this problem.

PAC Hypothesis

- Size of the hypothesis space of this problem.
  - 2^{3*4*2}

PAC Hypothesis

- Size of the hypothesis space of this problem.
  - 2^{3*4*2}
- If a hypothesis classifies more than 3 samples wrong, it is a bad hypothesis. What is the epsilon?

PAC Hypothesis

- Size of the hypothesis space of this problem.
  - 2^{3*4*2}
- If a hypothesis classifies more than 3 samples wrong, it is a bad hypothesis. What is the epsilon?
  - $\frac{3}{24}$

PAC Hypothesis

- Size of the hypothesis space of this problem.
  - 2^{3*4*2}
- If a hypothesis classifies more than 3 samples wrong, it is a bad hypothesis. What is the epsilon?
  - $\frac{3}{24}$
- After 12 examples are observed, what is the hypothesis space?

PAC Hypothesis

- Size of the hypothesis space of this problem.
  - 2^{3*4*2}
- If a hypothesis classifies more than 3 samples wrong, it is a bad hypothesis. What is the epsilon?
  - $\frac{3}{24}$
- After 12 examples are observed, what is the hypothesis space?
  - 2^{12}

PAC Hypothesis

- Size of the hypothesis space of this problem.
  - 2^{3*4*2}
- If a hypothesis classifies more than 3 samples wrong, it is a bad hypothesis. What is the epsilon?
  - $\frac{3}{24}$
- After 12 examples are observed, what is the hypothesis space?
  - 2^{12}
- How many of these are approximately correct?

PAC Hypothesis

- Size of the hypothesis space of this problem.
  - 2^{3*4*2}
- If a hypothesis classifies more than 3 samples wrong, it is a bad hypothesis. What is the epsilon?
  - $\frac{3}{24}$
- After 12 examples are observed, what is the hypothesis space?
  - 2^{12}
- How many of these are approximately correct?
  - $1 + 12 + \binom{12}{2}$

PAC Hypothesis

- Size of the hypothesis space of this problem.
  - 2^{3*4*2}
- If a hypothesis classifies more than 3 samples wrong, it is a bad hypothesis. What is the epsilon?
  - $\frac{3}{24}$
- After 12 examples are observed, what is the hypothesis space?
  - 2^{12}
- How many of these are approximately correct?
  - $1 + 12 + \binom{12}{2}$
- What is the probability that a randomly-chosen hypothesis that is consistent with the 8 examples seen so far will be approximately correct?

PAC Hypothesis

- Size of the hypothesis space of this problem.
  - 2^{3*4*2}
- If a hypothesis classifies more than 3 samples wrong, it is a bad hypothesis. What is the epsilon?
  - $\frac{3}{24}$
- After 12 examples are observed, what is the hypothesis space?
  - 2^{12}
- How many of these are approximately correct?
  - $1 + 12 + \binom{12}{2}$
- What is the probability that a randomly-chosen hypothesis that is consistent with the 8 examples seen so far will be approximately correct?
  - $\frac{1+12+\binom{12}{2}}{2^{12}}$