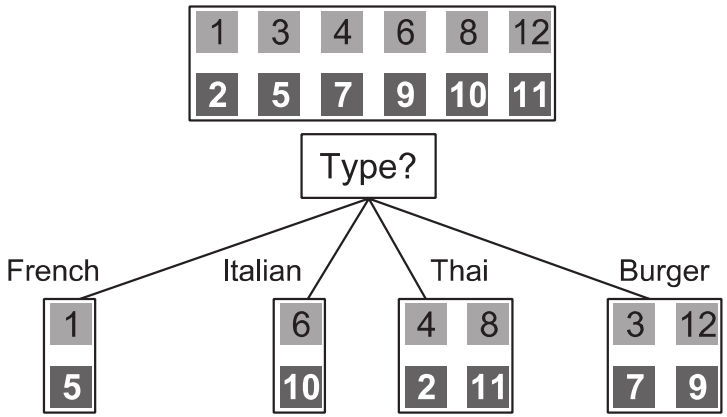
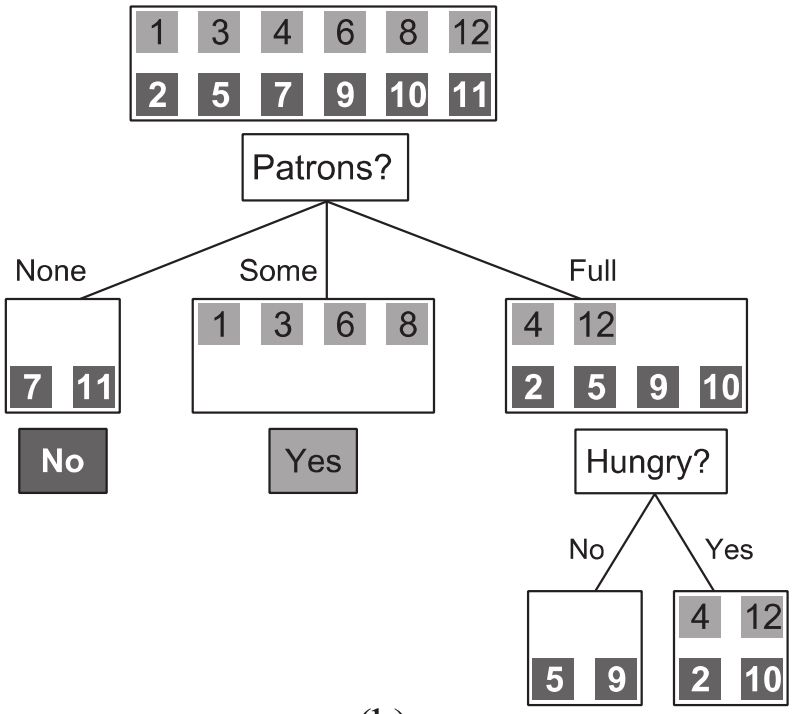
Lecture 3: Overfitting

Course Teacher: Md. Shariful Islam Bhuyan

Recap


1 3 4 6 8 12
2 5 7 9 10 11
Type?
French
Italian
Thai
Burger
1
5
6
10
4 8
2 11
3 12
7 9

(a)

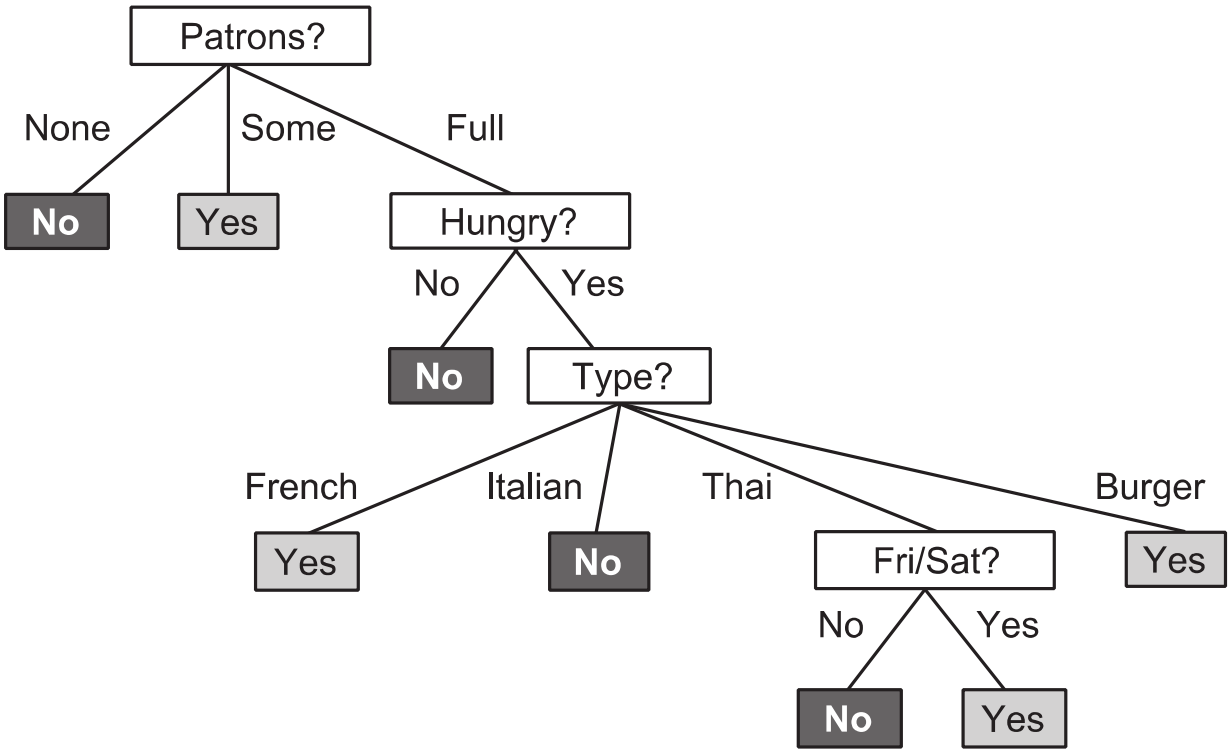

1 3 4 6 8 12
2 5 7 9 10 11
Patrons?
None
Some
Full
7 11
1 3 6 8
4 12
2 5 9 10
No
Yes
Hungry?
No
Yes
5 9
4 12
2 10

(b)

Errata on Last Class

- Divides example into sets with
  - less uniform distribution or less entropy
- Entropy is a measure of the uncertainty of a random variable
  - High entropy, less information, more encoding bits for data
  - Low entropy, more information, less encoding bits for data
  - For binary variable $B(q) = -q \log_2 q - (1-q) \log_2 (1-q)$
- Information gain $IG(A) = B\left(\frac{p}{p+n}\right) - \sum_{k=1}^{d} \frac{p_k+n_k}{p+n} B\left(\frac{p_k}{p_k+n_k}\right)$

Decision tree



Expressiveness of Binary Decision Trees

- Goal $\Leftrightarrow$ ($Path_1 \vee Path_2 \vee \ldots$)
- Path = ($Attribute_1 = a_1 \wedge Attribute_2 = a_2$)
- Boolean function in disjunctive normal form
- Hypothesis space: number of possible function for n attributes 2^{2^n}
- Table example
- Function approximation

Overfitting

- "If you torture the data long enough, it will confess" - Coase
- Example: Will a dice roll give 6?
  - For fair dice, learn a tree with a single node that says "NO"
  - Overfitting: Only 7-gram blue die with fingers crossed rolls 6
  - Irrelevant attributes: color, weight, time, is fingers crossed
- Increases with hypothesis space and number of attributes
- Decreases as we increase the number of training examples

Detecting Overfitting

- Train-test split/holdout cross validation
- Poor performance on test data
- Did not learn to generalize
  - Extreme case: table lookup
- Peeking
- Combat pruning

Continuous Valued Input

- Find the split point that gives the highest information gain
- Sort examples according to attribute values
- Consider only split points that are between two examples in sorted order that have different classifications
- Keep track of the running totals of positive and negative examples on each side of the split point

Example

Cheat	No		No		No		Yes		Yes		Yes		No		No		No		No	
	Taxable Income																			
Sorted Values →	60		70		75		85		90		95		100		120		125		220	
Split Positions →	55		65		72		80		87		92		97		110		122		172		230	
	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>
Yes	0	3	0	3	0	3	0	3	1	2	2	1	3	0	3	0	3	0	3	0	3	0
No	0	7	1	6	2	5	3	4	3	4	3	4	3	4	4	3	5	2	6	1	7	0
Gini	0.420		0.400		0.375		0.343		0.417		0.400		*0.300*		0.343		0.375		0.400		0.420	

Continuous Valued Output (Regression)

- Approximate function with continuous Range
- Apply continuous function of some subset of attributes at leaves
  - Linear regression
  - Mean value
- Decide when to stop splitting and begin applying leaf function