SC4000/CZ4041/CE4041: Machine Learning

Solutions to L6 Tutorial Questions

Kelly KE

School of Computer Science and Engineering,

NTU, Singapore

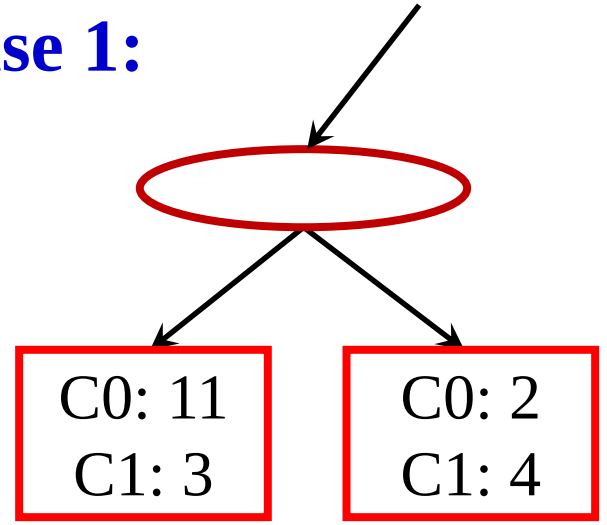
Question 1

- Pessimistic error?

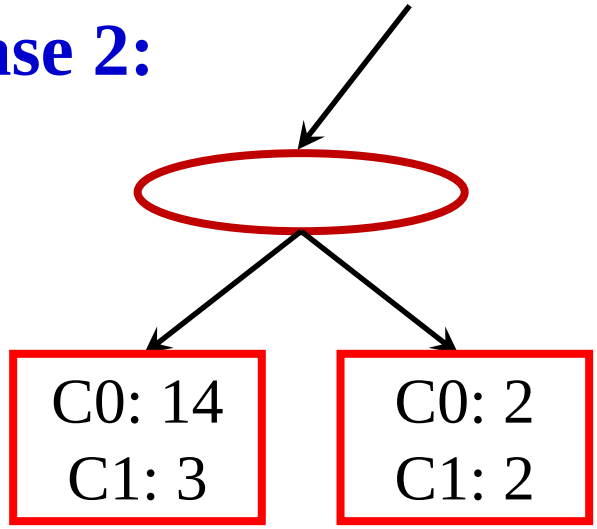
$$e'(T) = e(T) + N \times 0.5$$

PRUNE?

Case 1:



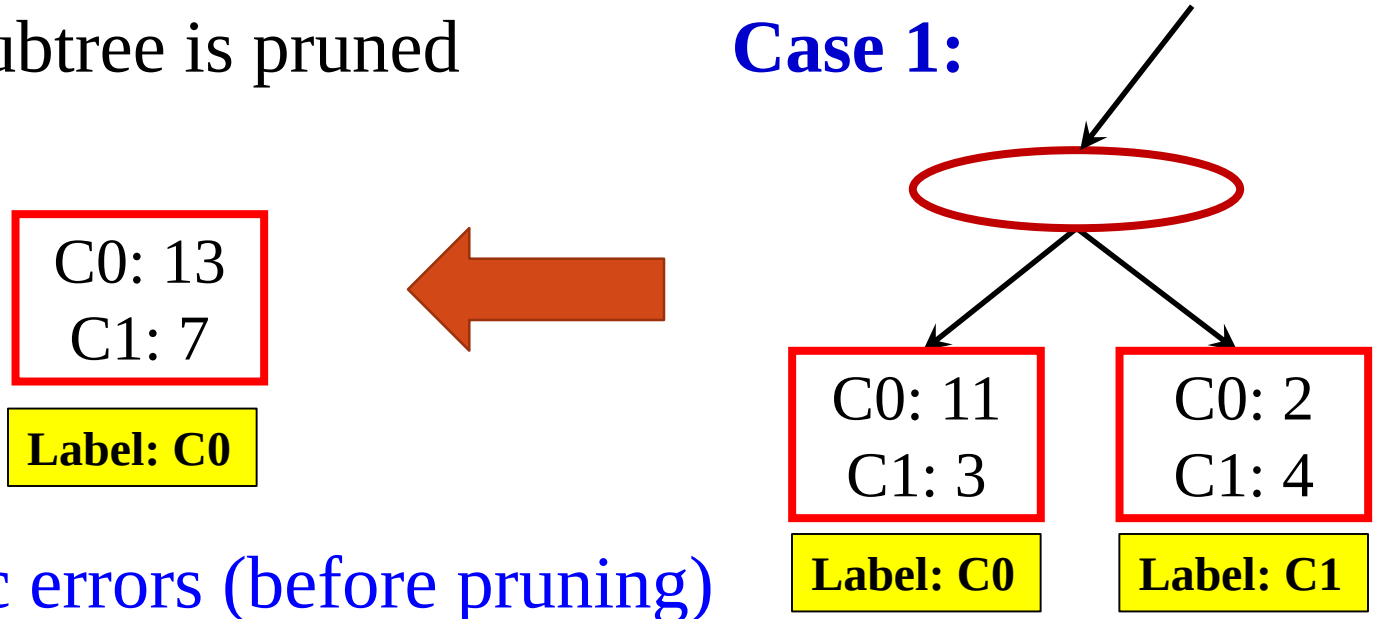
Case 2:



Pessimistic Error: Case 1

If the subtree is pruned

Case 1:



Pessimistic errors (before pruning)

$= 3 + 2 + 2 \times 0.5 = 6$

Training error Model complexity

Pessimistic errors (after pruning)

$= 7 + 0.5 = 7.5$

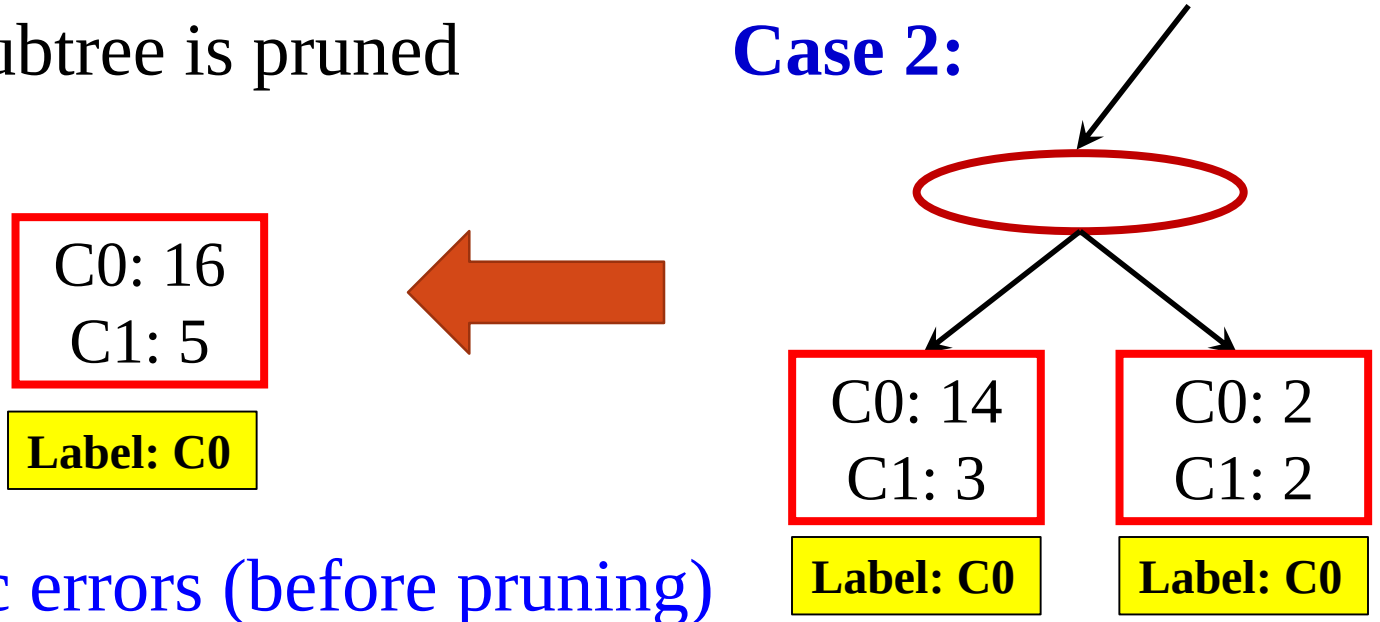
Training error Model complexity

Not prune!

Pessimistic Error: Case 2

If the subtree is pruned

Case 2:



Pessimistic errors (before pruning)

$= 3 + 2 + 2 \times 0.5 = 6$

Training error Model complexity

Pessimistic errors (after pruning)

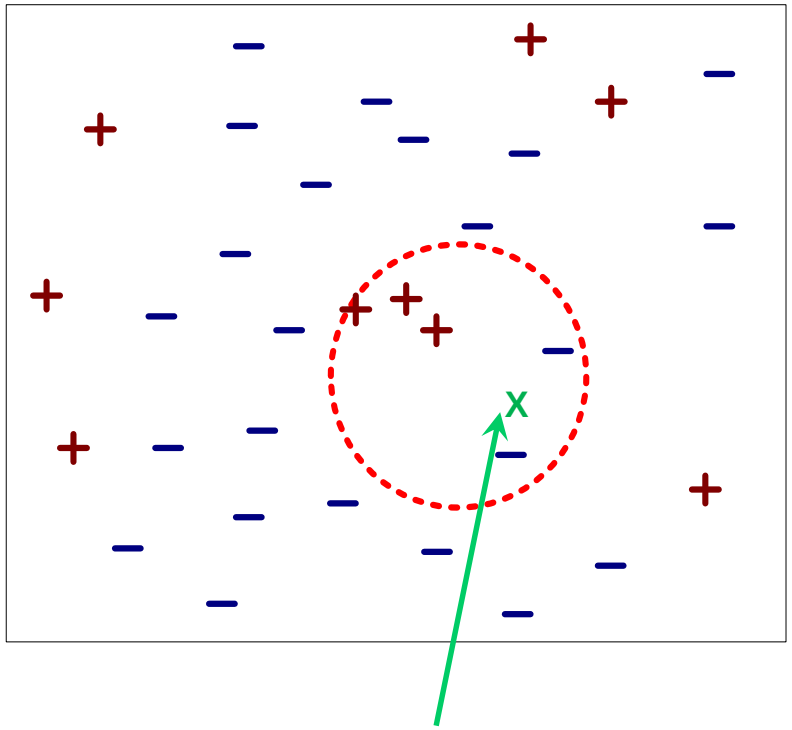
$= 5 + 0.5 = 5.5$

Training error Model complexity

Prune!

Question 2

Consider a binary classification problem, and a 5-NN classifier



Retrieved instances →

Training instances	Class label	Distance to test instance
1	+	3
2	+	3.5
3	+	4
4	-	1.5
5	-	2

- Majority voting:

$+ : 3 > - : 2$

- Distance-weighted voting:

Question 3 (cont.)

Training record	Class label	Distance to test record
1	+	3
2	+	3.5
3	+	4
4	-	1.5
5	-	2

Distance−Weighted voting for +:

$$\left(\frac{1}{3}\right)^2 + \left(\frac{1}{3.5}\right)^2 + \left(\frac{1}{4}\right)^2 = 0.2552$$

Distance−Weighted voting for −:

$$\left(\frac{1}{2}\right)^2 + \left(\frac{1}{1.5}\right)^2 = 0.6944$$

Thank you!