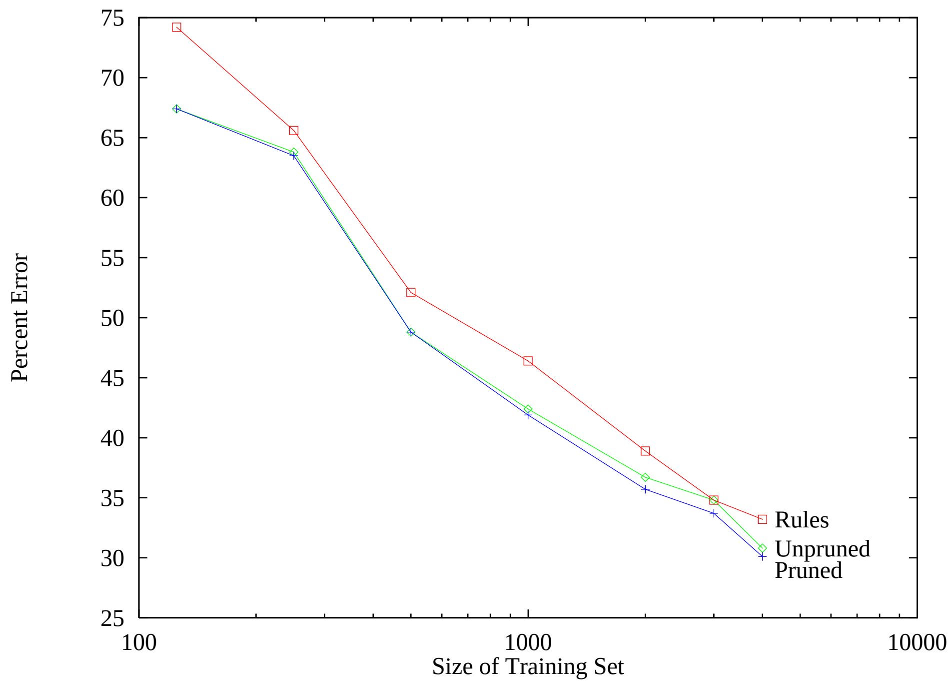
CS533 — Program 4 — Solution

1. Learning curves for C4.5:



The cross-validation numbers for training on 3000 points were 34.8 (unpruned), 33.7 (pruned), and 34.8 (rules), so I would make new predictions using the pruned trees.

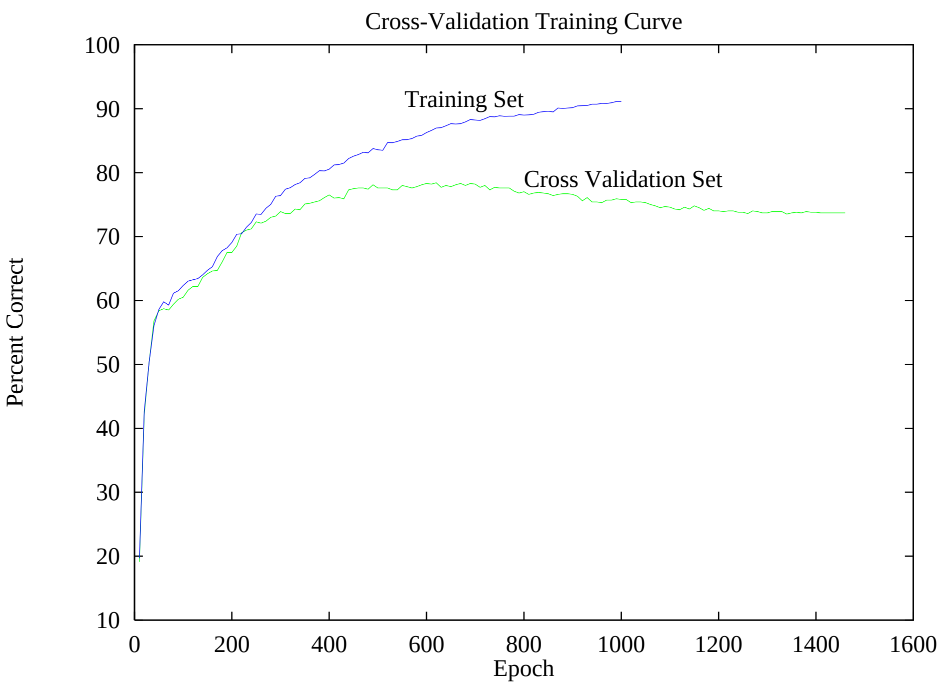
Confusion Matrix for Pruned Trees on test data:

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)	(m)	<-classified a
140	2	3	6					1		3	2	3	(a): class A
1	106		9	3	1	2	6	2	1	5	1	3	(b): class B
	2	110		4	6	1	1	2		4	2		(c): class C
1	5		102			2	6	1		2		1	(d): class D
1	2	3	1	103	1	5	2	4		3		1	(e): class E
	8	2	4	2	101		4	2	1		2		(f): class F
1	6	4	3	4		94	2	5	1	2	2	1	(g): class G
3	8	1	7	3	3		69	4	1	5		1	(h): class H
1	5		3		3	2	3	113	6	3	1		(i): class I
	3	2	5	1	8	1		3	107	2			(j): class J
5	1	2	1	3	1	1	5			88	4		(k): class K
	1	2	2		1	2		1	2		125	1	(l): class L
4	1		1		2	1	3	2		3	1	128	(m): class M
	2		1	1	1	3	2	2	2	3		6	(n): class N
1	8	2	5			5	5		1	2	1	1	(o): class O
1	4	1	4	2	7	1	4	5	2	1			(p): class P
1	2			2		4		2	1	4	4	1	(q): class Q
4	9		4	1		2	5		1	7		1	(r): class R
2	8		1	3	1	5		3		1		1	(s): class S
1			2	1	4		1	1		4	1	1	(t): class T
2	2	4	1	1		2			3	2		7	(u): class U
3	3	2		2	3		3	1		1	1		(v): class V
3	2	1			1		5	1	1		1	2	(w): class W

2	7	1	4	11	4	3	6	4		1	2	2	(x): class X
1		2	3		2		2	2			2		(y): class Y
	6		1	6	3	4			1			1	(z): class Z

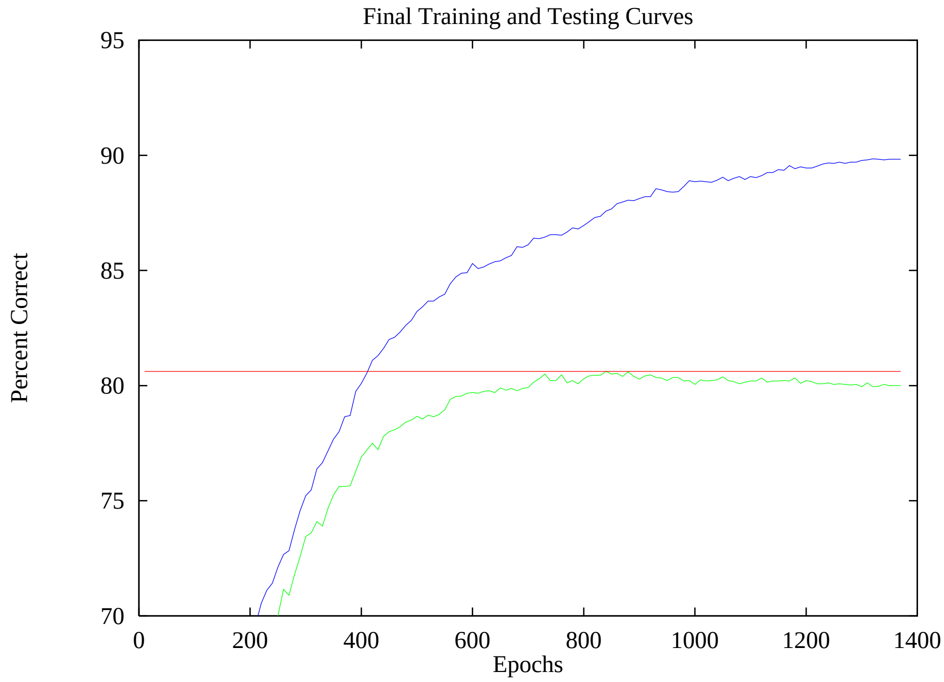
(n)	(o)	(p)	(q)	(r)	(s)	(t)	(u)	(v)	(w)	(x)	(y)	(z)	<-classified as
			1	3			2	1			1	1	(a): class A
1		1	1	6	10		1	1	1	1		2	(b): class B
1	5	1	5	4		3	1	1	1	2	1	1	(c): class C
2	5	3	1	5	4	3						2	(d): class D
2			7	2	2	3		1	1	1		3	(e): class E
		11	1	1	4	5			2	4	6	1	(f): class F
	5	3	3	4	1			1	2			6	(g): class G
	9	5	3	9	1		2	4	3	1			(h): class H
1	1	1	2	1	4	1	1				1		(i): class I
	1	1		1	1		2	1	1			1	(j): class J
	1			7	4	5	1		1	1	1	2	(k): class K
	1	2	5	4			1		1	5	1		(l): class L
7	2	1		3		1	4		7	1			(m): class M
115	1	1		1	1	1	3	1	1				(n): class N
1	94	5	7	3	1		1	1	3		1	1	(o): class O
	1	121	1	1	4	2		2	1	2	3	2	(p): class P
	8	1	98	1	1			7		2		3	(q): class Q
1	4	1	1	89		1	1	1	1				(r): class R
2		1	6	6	107		2	1		4		10	(s): class S
1				6	2	114		3	1	2	5	3	(t): class T
3	1	1	3	4		1	113	5	2	3	2		(u): class U
3		1				2	1	120	5		4	1	(v): class V
3		1					2	1	118	2	1	2	(w): class W
	2			5	5	1	3	1	1	110	2	4	(x): class X
	1	8	4	1	1	8	2	8	1		100		(y): class Y
	2	1	2		9	1	1	1	1			109	(z): class Z

2. Training curve for opti and optesti on the cross-validation set. There is clearly some overfitting.



From this, we see that the optimal stopping epoch was 620.

3. Training curve for opti on the final test set. Again, there was some overfitting, but it was not as severe. The performance at 620 epochs was 79.75 percent correct, which is less than the actual maximum of 80.62 (achieved at 840 epochs). At 1370 epochs (training to minimum), the performance was 80.00. So this suggests that overfitting was not a critical problem.



4. Here is a learning curve for opt.