Jeremy Calvert
CSEP 573
Assignment 4
December 5, 2007

1.1 Experiment with at least 2 different initial settings of cluster means, covariance and prior probabilities of data. For each initial setting, answer:

(a) Compare the clusters you get with the ground truth: what is the accuracy? The accuracy should not be too low if your algorithm works properly. What can you conclude from the results?

Each run of the EM algorithm with randomly selected points chosen for initial mean, and from which the initial standard deviation is derived consistently yields the same results, with 9 errors, or 94% accuracy. The results suggest that the setosa cluster is easier to differentiate from the virginica and versicolor clusters, and that these two clusters are somewhat harder to differentiate. Perhaps this implementation fails in this regard because the covariance matrix is not in fact diagonal.

(b) Include a table that has all data points, with mis-clustered data points clearly marked.

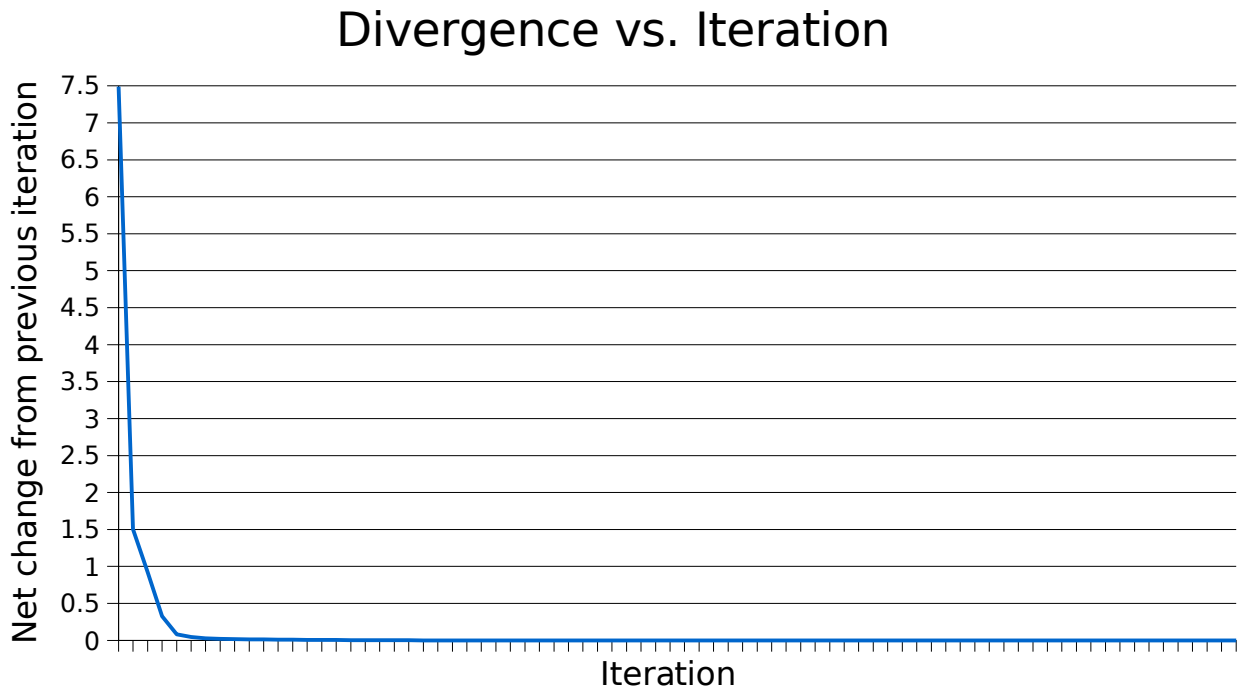
This is included in the following three pages.

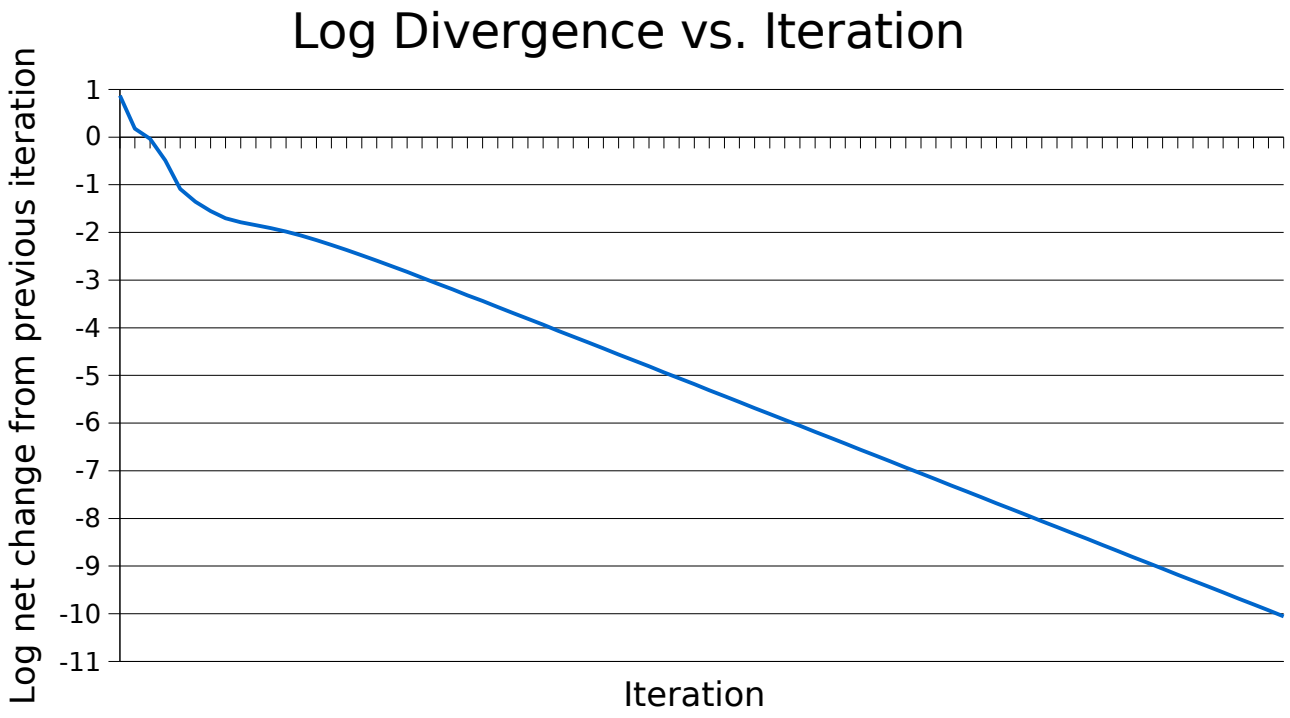
A	B	C	D	Correct	Run 1	Run 2
5.1	3.5	1.4	0.2	Iris-setosa	Iris-setosa	Iris-setosa
4.9	3	1.4	0.2	Iris-setosa	Iris-setosa	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa	Iris-setosa	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5	3.6	1.4	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5.4	3.9	1.7	0.4	Iris-setosa	Iris-setosa	Iris-setosa
4.6	3.4	1.4	0.3	Iris-setosa	Iris-setosa	Iris-setosa
5	3.4	1.5	0.2	Iris-setosa	Iris-setosa	Iris-setosa
4.4	2.9	1.4	0.2	Iris-setosa	Iris-setosa	Iris-setosa
4.9	3.1	1.5	0.1	Iris-setosa	Iris-setosa	Iris-setosa
5.4	3.7	1.5	0.2	Iris-setosa	Iris-setosa	Iris-setosa
4.8	3.4	1.6	0.2	Iris-setosa	Iris-setosa	Iris-setosa
4.8	3	1.4	0.1	Iris-setosa	Iris-setosa	Iris-setosa
4.3	3	1.1	0.1	Iris-setosa	Iris-setosa	Iris-setosa
5.8	4	1.2	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5.7	4.4	1.5	0.4	Iris-setosa	Iris-setosa	Iris-setosa
5.4	3.9	1.3	0.4	Iris-setosa	Iris-setosa	Iris-setosa
5.1	3.5	1.4	0.3	Iris-setosa	Iris-setosa	Iris-setosa
5.7	3.8	1.7	0.3	Iris-setosa	Iris-setosa	Iris-setosa
5.1	3.8	1.5	0.3	Iris-setosa	Iris-setosa	Iris-setosa
5.4	3.4	1.7	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5.1	3.7	1.5	0.4	Iris-setosa	Iris-setosa	Iris-setosa
4.6	3.6	1	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5.1	3.3	1.7	0.5	Iris-setosa	Iris-setosa	Iris-setosa
4.8	3.4	1.9	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5	3	1.6	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5	3.4	1.6	0.4	Iris-setosa	Iris-setosa	Iris-setosa
5.2	3.5	1.5	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5.2	3.4	1.4	0.2	Iris-setosa	Iris-setosa	Iris-setosa
4.7	3.2	1.6	0.2	Iris-setosa	Iris-setosa	Iris-setosa
4.8	3.1	1.6	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5.4	3.4	1.5	0.4	Iris-setosa	Iris-setosa	Iris-setosa
5.2	4.1	1.5	0.1	Iris-setosa	Iris-setosa	Iris-setosa
5.5	4.2	1.4	0.2	Iris-setosa	Iris-setosa	Iris-setosa
4.9	3.1	1.5	0.1	Iris-setosa	Iris-setosa	Iris-setosa
5	3.2	1.2	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5.5	3.5	1.3	0.2	Iris-setosa	Iris-setosa	Iris-setosa
4.9	3.1	1.5	0.1	Iris-setosa	Iris-setosa	Iris-setosa
4.4	3	1.3	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5.1	3.4	1.5	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5	3.5	1.3	0.3	Iris-setosa	Iris-setosa	Iris-setosa
4.5	2.3	1.3	0.3	Iris-setosa	Iris-setosa	Iris-setosa
4.4	3.2	1.3	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5	3.5	1.6	0.6	Iris-setosa	Iris-setosa	Iris-setosa
5.1	3.8	1.9	0.4	Iris-setosa	Iris-setosa	Iris-setosa
4.8	3	1.4	0.3	Iris-setosa	Iris-setosa	Iris-setosa
5.1	3.8	1.6	0.2	Iris-setosa	Iris-setosa	Iris-setosa
4.6	3.2	1.4	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5.3	3.7	1.5	0.2	Iris-setosa	Iris-setosa	Iris-setosa
5	3.3	1.4	0.2	Iris-setosa	Iris-setosa	Iris-setosa

A	B	C	D	Correct	Run 1	Run 2
7	3.2	4.7	1.4	Iris-versicolor	Iris-virginica	Iris-virginica
6.4	3.2	4.5	1.5	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.9	3.1	4.9	1.5	Iris-versicolor	Iris-virginica	Iris-virginica
5.5	2.3	4	1.3	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.5	2.8	4.6	1.5	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.7	2.8	4.5	1.3	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.3	3.3	4.7	1.6	Iris-versicolor	Iris-virginica	Iris-virginica
4.9	2.4	3.3	1	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.6	2.9	4.6	1.3	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.2	2.7	3.9	1.4	Iris-versicolor	Iris-versicolor	Iris-versicolor
5	2	3.5	1	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.9	3	4.2	1.5	Iris-versicolor	Iris-versicolor	Iris-versicolor
6	2.2	4	1	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.1	2.9	4.7	1.4	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.6	2.9	3.6	1.3	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.7	3.1	4.4	1.4	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.6	3	4.5	1.5	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.8	2.7	4.1	1	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.2	2.2	4.5	1.5	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.6	2.5	3.9	1.1	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.9	3.2	4.8	1.8	Iris-versicolor	Iris-virginica	Iris-virginica
6.1	2.8	4	1.3	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.3	2.5	4.9	1.5	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.1	2.8	4.7	1.2	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.4	2.9	4.3	1.3	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.6	3	4.4	1.4	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.8	2.8	4.8	1.4	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.7	3	5	1.7	Iris-versicolor	Iris-virginica	Iris-virginica
6	2.9	4.5	1.5	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.7	2.6	3.5	1	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.5	2.4	3.8	1.1	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.5	2.4	3.7	1	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.8	2.7	3.9	1.2	Iris-versicolor	Iris-versicolor	Iris-versicolor
6	2.7	5.1	1.6	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.4	3	4.5	1.5	Iris-versicolor	Iris-versicolor	Iris-versicolor
6	3.4	4.5	1.6	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.7	3.1	4.7	1.5	Iris-versicolor	Iris-virginica	Iris-virginica
6.3	2.3	4.4	1.3	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.6	3	4.1	1.3	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.5	2.5	4	1.3	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.5	2.6	4.4	1.2	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.1	3	4.6	1.4	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.8	2.6	4	1.2	Iris-versicolor	Iris-versicolor	Iris-versicolor
5	2.3	3.3	1	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.6	2.7	4.2	1.3	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.7	3	4.2	1.2	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.7	2.9	4.2	1.3	Iris-versicolor	Iris-versicolor	Iris-versicolor
6.2	2.9	4.3	1.3	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.1	2.5	3	1.1	Iris-versicolor	Iris-versicolor	Iris-versicolor
5.7	2.8	4.1	1.3	Iris-versicolor	Iris-versicolor	Iris-versicolor

A	B	C	D	Correct	Run 1	Run 2
6.3	3.3	6	2.5	Iris-virginica	Iris-virginica	Iris-virginica
5.8	2.7	5.1	1.9	Iris-virginica	Iris-virginica	Iris-virginica
7.1	3	5.9	2.1	Iris-virginica	Iris-virginica	Iris-virginica
6.3	2.9	5.6	1.8	Iris-virginica	Iris-virginica	Iris-virginica
6.5	3	5.8	2.2	Iris-virginica	Iris-virginica	Iris-virginica
7.6	3	6.6	2.1	Iris-virginica	Iris-virginica	Iris-virginica
4.9	2.5	4.5	1.7	Iris-virginica	Iris-versicolor	Iris-versicolor
7.3	2.9	6.3	1.8	Iris-virginica	Iris-virginica	Iris-virginica
6.7	2.5	5.8	1.8	Iris-virginica	Iris-virginica	Iris-virginica
7.2	3.6	6.1	2.5	Iris-virginica	Iris-virginica	Iris-virginica
6.5	3.2	5.1	2	Iris-virginica	Iris-virginica	Iris-virginica
6.4	2.7	5.3	1.9	Iris-virginica	Iris-virginica	Iris-virginica
6.8	3	5.5	2.1	Iris-virginica	Iris-virginica	Iris-virginica
5.7	2.5	5	2	Iris-virginica	Iris-virginica	Iris-virginica
5.8	2.8	5.1	2.4	Iris-virginica	Iris-virginica	Iris-virginica
6.4	3.2	5.3	2.3	Iris-virginica	Iris-virginica	Iris-virginica
6.5	3	5.5	1.8	Iris-virginica	Iris-virginica	Iris-virginica
7.7	3.8	6.7	2.2	Iris-virginica	Iris-virginica	Iris-virginica
7.7	2.6	6.9	2.3	Iris-virginica	Iris-virginica	Iris-virginica
6	2.2	5	1.5	Iris-virginica	Iris-versicolor	Iris-versicolor
6.9	3.2	5.7	2.3	Iris-virginica	Iris-virginica	Iris-virginica
5.6	2.8	4.9	2	Iris-virginica	Iris-virginica	Iris-virginica
7.7	2.8	6.7	2	Iris-virginica	Iris-virginica	Iris-virginica
6.3	2.7	4.9	1.8	Iris-virginica	Iris-virginica	Iris-virginica
6.7	3.3	5.7	2.1	Iris-virginica	Iris-virginica	Iris-virginica
7.2	3.2	6	1.8	Iris-virginica	Iris-virginica	Iris-virginica
6.2	2.8	4.8	1.8	Iris-virginica	Iris-virginica	Iris-virginica
6.1	3	4.9	1.8	Iris-virginica	Iris-virginica	Iris-virginica
6.4	2.8	5.6	2.1	Iris-virginica	Iris-virginica	Iris-virginica
7.2	3	5.8	1.6	Iris-virginica	Iris-virginica	Iris-virginica
7.4	2.8	6.1	1.9	Iris-virginica	Iris-virginica	Iris-virginica
7.9	3.8	6.4	2	Iris-virginica	Iris-virginica	Iris-virginica
6.4	2.8	5.6	2.2	Iris-virginica	Iris-virginica	Iris-virginica
6.3	2.8	5.1	1.5	Iris-virginica	Iris-versicolor	Iris-versicolor
6.1	2.6	5.6	1.4	Iris-virginica	Iris-virginica	Iris-virginica
7.7	3	6.1	2.3	Iris-virginica	Iris-virginica	Iris-virginica
6.3	3.4	5.6	2.4	Iris-virginica	Iris-virginica	Iris-virginica
6.4	3.1	5.5	1.8	Iris-virginica	Iris-virginica	Iris-virginica
6	3	4.8	1.8	Iris-virginica	Iris-virginica	Iris-virginica
6.9	3.1	5.4	2.1	Iris-virginica	Iris-virginica	Iris-virginica
6.7	3.1	5.6	2.4	Iris-virginica	Iris-virginica	Iris-virginica
6.9	3.1	5.1	2.3	Iris-virginica	Iris-virginica	Iris-virginica
5.8	2.7	5.1	1.9	Iris-virginica	Iris-virginica	Iris-virginica
6.8	3.2	5.9	2.3	Iris-virginica	Iris-virginica	Iris-virginica
6.7	3.3	5.7	2.5	Iris-virginica	Iris-virginica	Iris-virginica
6.7	3	5.2	2.3	Iris-virginica	Iris-virginica	Iris-virginica
6.3	2.5	5	1.9	Iris-virginica	Iris-virginica	Iris-virginica
6.5	3	5.2	2	Iris-virginica	Iris-virginica	Iris-virginica
6.2	3.4	5.4	2.3	Iris-virginica	Iris-virginica	Iris-virginica
5.9	3	5.1	1.8	Iris-virginica	Iris-virginica	Iris-virginica

(c) How many iterations does it take to converge? Draw a curve of EM iterations-accuracy. What can you conclude from the curve?





Based on the linear shape of the graph of the log of the convergence, it converges exponentially, which is to say, on each iteration, the divergence is some fixed multiple (smaller than 1) of what it was in the previous iteration.

2. (10 points) Suppose there is a decision tree algorithm that always learns a canonical tree, which has some nice properties and is often very useful. A canonical tree requires that all nodes with the same depth split the same variable (attribute). Assume on a given dataset the size (number of nodes) of the learned canonical tree is T, and the size of the tree learned by ID3 is D.
Use your intuition to tell which of the following are correct.

(a) T is at least as large as D.
(b) T is larger than D up to a constant C.
(c) T is larger than D up to a constant factor C.
(d) T can be exponentially larger than D.
(e) T and D are not comparable in general.

Intuitively, since we know nothing about the decision tree algorithm that always learns a canonical tree, (e) is true. For example, it could be that the algorithm always learns a decision stump (which is canonical), in which case (a) – (d) are (almost) always false.

It does seem, however, that if the success rates are comparable, that at least (a) will hold in most cases. This is because whenever attributes are related (not entirely independent), the choice of attribute that minimizes entropy will depend on the value represented by the branch of a parent, meaning that on one branch ID3 will opt to split on one attribute, and on another branch, another attribute. And since ID3 does this in an effort to minimize the size of the decision tree, the resulting tree will be smaller than any canonical tree with a comparable success rate.

3. (10 points) A dataset has 10 attributes X1 to X10. We have 9 data points D1 to D9 and we use ID3 to learn a decision tree. Can you give an example of D1 to D9 that maximizes the depth of the tree learned by ID3?

Basically, ID3 grows a large tree when little progress is made toward successfully predicting the goal class. Since we're asked to give an example that maximizes the depth (rather than the overall size) we can take the degenerate case where the attributes values are all identical, and the goal class is evidentially independent of these attributes, and as noisy as possible, as in the following:

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Goal
1	T	T	T	T	T	T	T	T	T	T	F
2	T	T	T	T	T	T	T	T	T	T	T
3	T	T	T	T	T	T	T	T	T	T	F
4	T	T	T	T	T	T	T	T	T	T	T
5	T	T	T	T	T	T	T	T	T	T	F
6	T	T	T	T	T	T	T	T	T	T	T
7	T	T	T	T	T	T	T	T	T	T	F
8	T	T	T	T	T	T	T	T	T	T	T
9	T	T	T	T	T	T	T	T	T	T	F

Then in each stage of ID3, one attribute will be selected, and will have one branch descending down (since each attribute has a single realized value) to a child, where another attribute will be selected (with equally poor entropy).

4. (10 points) What is the asymptotic worst-case time complexity of the propositional rule induction algorithm in the lecture notes, as a function of the number of examples and number of attributes? (Assume the evaluation function is accuracy.)

First we analyze the worst case complexity of learning a single rule using the algorithm:

```
head ← y
body ← Ø
repeat
     for each literal x
          rx ← r with x added to body
          Eval(rx)
     body ← body ^ best x
until no x improves Eval(r)
return r
```

The inner loop repeats up to once for each literal tested. In each iteration of the outer loop, we consider one less literal (since one has just been added), so, letting n denote the number of literals the max number of total iterations is $n + (n – 1) + ... + 1 = n(n-1)/2$ for a complexity of $O(n^2)$. There are actually two literals per attribute (assuming the attributes are boolean), but this doesn't affect the order of the complexity. So if m denotes the number of attributes, the complexity is still $O(m^2)$.

Next we analyze the complexity of learning a set of rules using the algorithm:

```
R ← Ø
S ← examples
repeat
     learn a single rule r
     R ← R U {r}
     S ← S - positive examples covered by r
until S = Ø
return R
```

Here we learn a single rule for each example, and in worst case we run through all examples. So if s denotes the number of samples, the total overall worst case complexity is $O(s * m^2)$.