Exercise Sheet on Evaluation of Hypotheses

For the questions in this topic, you may find some of the statistical tables in the Appendix useful.

1. Suppose you have used your favourite concept learning algorithm to learn a hypothesis h_1 from some training data. You are interested in knowing the accuracy that the hypothesis can be expected to achieve on the underlying population. To assess this accuracy, you apply the hypothesis to a test data set consisting of 45 instances that you had held back from the training data set. The error rate observed on the training data is 6.67%. Calculate the 95% confidence interval for the true error.
2. You now decide to change a few parameters within the learning algorithm used in Question 7.1 and learn two more hypotheses, h_2 and h_3. The error rates for these new hypotheses observed on the test data set of 45 instances were 8.89% and 13.3%, respectively. To what degree can you be confident that h_2 will perform worse than h_1 on the underlying population? Is your confidence higher or lower for h_3 performing worse than h_1 on the underlying population?
3. You now decide to try out a decision tree induction algorithm to see if it can outperform your favourite concept learning algorithm. You decide to use 10 10-fold cross-validation. The error rates for the 10 cross-validation folds for the two algorithms are shown in Table 1.

	Error Rates	
CV Fold	Favourite Algorithm	Decision Tree Induction
1	8.89%	9.3%
2	9.52%	9.48%
3	8.13%	9.12%
4	9.48%	9.13%
5	10.12%	9.98%
6	10.23%	11.01%
7	8.56%	9.02%
8	9.12%	8.56%
9	9.23%	9.23%
10	9.11%	9.08%

Table 1: Cross-Validation Error Rates

With what confidence level can you assume that your favourite concept description algorithm will outperform the decision tree induction algorithm in this domain?

4. You decide to further investigate the errors in classification generated by the hypotheses generated in Questions 7.1 and 7.2. The confusion matrices for the three hypotheses are shown below:

h_1		Actual		Marginal Sum
		Positive	Negative	
Predicted	Positive	29	1	30
	Negative	2	13	15
	Marginal Sum	31	14	45

h_2		Actual		Marginal Sum
		Positive	Negative	
Predicted	Positive	29	3	32
	Negative	1	12	13
	Marginal Sum	30	15	45

h_3		Actual		Marginal Sum
		Positive	Negative	
Predicted	Positive	27	3	30
	Negative	3	12	15
	Marginal Sum	30	15	45

Using the Euclidean distance on an ROC plot from the "perfect classifier" as the metric, choose the best classifier

- assuming equal costs for false positives and false negatives
- assuming that false positives cost 4 times as much as false negatives

Appendix: Useful Statistical Tables

	Degrees of Freedom			
Probability	1	2	3	4
0.95	0.004	0.10	0.35	0.71
0.90	0.02	0.21	0.58	1.06
0.80	0.06	0.45	1.01	1.65
0.70	0.15	0.71	1.42	2.2
0.50	0.46	1.39	2.37	3.36
0.30	1.07	2.41	3.66	4.88
0.20	1.64	3.22	4.64	5.99
0.10	2.71	4.60	6.25	7.78
0.05	3.84	5.99	7.82	9.49
0.01	6.64	9.21	11.34	13.28
0.001	10.83	13.82	16.27	18.47

Table 2: χ^2 Table

Confidence Level N%	30%	50%	68%	80%	90%	95%	98%	99%
z_N	0.41	0.67	1.00	1.28	1.64	1.96	2.33	2.58

Table 3: Values for z_N for two-sided N% confidence intervals

	Confidence Level N				
Degrees of Freedom	65%	90%	95%	98%	99%
2	1.21	2.92	4.30	6.96	9.92
5	1.031	2.02	2.57	3.36	4.03
9	0.985	1.83	2.26	2.82	3.25
10	0.984	1.81	2.23	2.76	3.17
20	0.957	1.72	2.09	2.53	2.84
30	0.949	1.70	2.04	2.46	2.75
120	0.938	1.66	1.98	2.36	2.62
∞	0.934	1.64	1.96	2.33	2.58

Table 4: Values of $t_{N,v}$ for two-sided confidence intervals