List of Tables

Table 2.1: The Single- and Three-Letter Amino Acid Codes 10
Table 2.2: The four types of kernel functions frequently used with SVM 21
Table 2.3: Set Operations 26
Table 2.4: Sunburn Dataset [Wroblewski, 1995] 29
Table 2.5: A sample database 31
Table 2.6: Discernibility matrix for data in Table 2.5 31
Table 2.7: Weather dataset 33
Table 2.8: Binary matrix Representation 36
Table 2.9: Major GPCR classification based on GPCRDB (as of Nov’ 2003) 40
Table 3.1: Format for sequence stored in database 53
Table 3.2: STRLIB01 interface methods 56
Table 3.3: A-Set, P-Set and D-Set for GPCR database 57
Table 4.1a: A-Table for Myoglobin Dataset 62
Table 4.1b: D-Table for Myoglobin Dataset 63
Table 4.1c: P-Table for Myoglobin Dataset 64
Table 4.2a: A-Table for GPCR 65
Table 4.2b: D-Table for GPCR 65
Table 4.2c: P-Table for GPCR 65
Table 4.3: Sorted data of Table 2.5 based on E 69
Table 4.4: Generation of Boolean Table 69
Table 4.5: Final iteration of Reduct Computation Algorithm 70
Table 4.6: Reduced table with only predominant attributes \{b,c\} 70
Table 4.7: Reduced table with only predominant attributes \{b,d\} 71
Table 4.8: Comparison of RDT with other Classification Techniques 74
Table 5.1: Frequency Table for the given Sequence’s 79
Table 5.2: Neighborhood Associations for different characters with A as center and distance ‘d’ 80
Table 5.3: Neighborhood Matrix (NM) for distance ‘d’=1 80
Table 5.4a: Neighborhood Associations with centers at S (Last character). 81
Table 5.4b: Neighborhood Associations with centers at M (First character). 81
Table 5.4c: Neighborhood Associations with centers at W (LFC). 82
Table 5.5: BAM for Distance d=1 for different Threshold  
Table 5.6: Training set with character W as center (LFC)  
Table 5.7: Binarized values of Table 5.6 with Threshold T=0.05  
Table 5.8: Reduced table containing PA  
Table 5.9: Reduct for the LFC for different training sets  
Table 5.10: Results of Five-fold test of GPCR dataset  
Table 5.11: Comparison of Five-fold test on GPCR dataset  
Table 5.12: Comparison of classification techniques using Kappa Statistics  
Table 6.1: BAM for Class E sequences of GPCR family  
Table 6.2: Frequency distribution table of string ‘y’  
Table 6.3: Neighborhood analysis of sequence ‘y’ at distance ‘d’ =5 with center ‘W’  
Table 6.4: Binarized vector of Table 6.3 data with Threshold ‘T’ = 0.05  
Table 6.5: P-Table for s1  
Table 6.6: Frequency distribution of sequence s1  
Table 6.7: Binary Association Matrix with ‘W’ as center  
Table 7.1: P-Table for s2  
Table 7.2: Frequency distribution for sequence s2  
Table 7.3: BAM of sequence s2  
Table 7.4: Sets generated by Sequence Arithmetic Module  
Table 7.5: Characteristics of datasets  
Table 7.6: Comparison of RDT with other Classification Methods  
Table 7.7: Accuracy rate of the cross-validation for classifying the GPCR