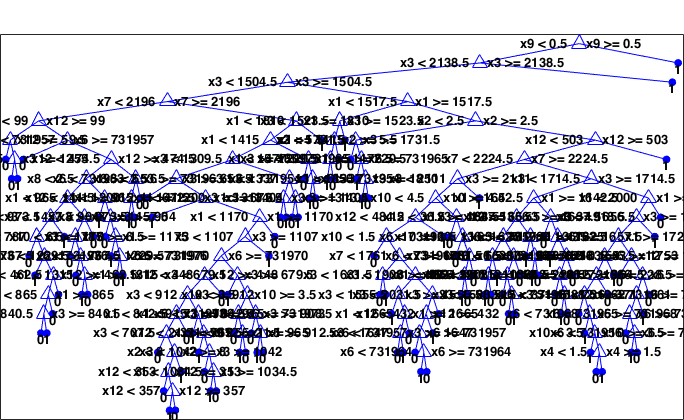
# Classification with Decision Tree on Flight Delays Dataset

## Transformation of the Dataset

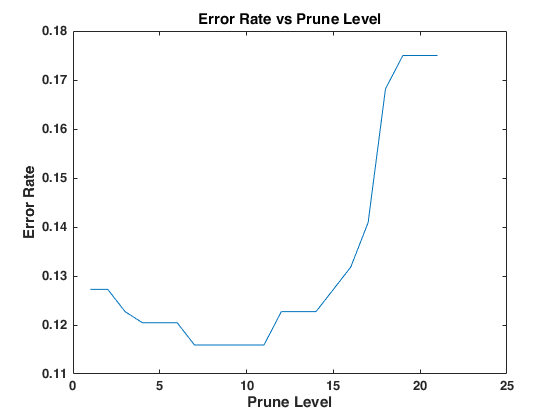
I wrote the codes to transform categorical data from string and date formats into integer format and then I partitioned the dataset into 1761 training data points and 440 validation data points.

## Using Gini Index to Create a Full Classification Tree

I fitted a full classification tree using “fitctree” by using Gini Index. Error rate of the tree is 0.1318 for my run. Please note that there may be differences between each run as in each run the validation and training data is re-shuffled. Here is the graph of the tree:

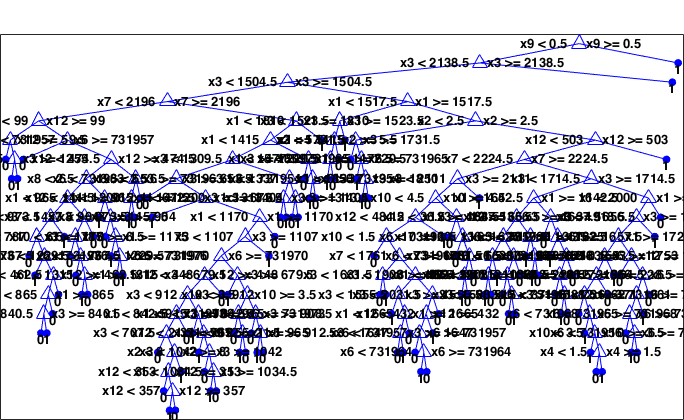


Here is the graph of the error rates for different prunes using Gini Index:



Minimum error rate is 0.1159 for prune level 7, 8, 9, 10 and 11 and they are in the defined limits. Thus, I picked 11 as the best prune level since it has the minimum error rate and pruned more.

Here is the graph of this tree:



### Click the arrow on the left to see rules of the tree as an output of MATLAB’s view function.

1 if x9<0.5 then node 2 elseif x9>=0.5 then node 3 else 0

2 if x3<2138.5 then node 4 elseif x3>=2138.5 then node 5 else 0

3 class = 1

4 if x3<1504.5 then node 6 elseif x3>=1504.5 then node 7 else 0

5 class = 1

6 if x7<2196 then node 8 elseif x7>=2196 then node 9 else 0

7 if x1<1517.5 then node 10 elseif x1>=1517.5 then node 11 else 0

8 class = 0

9 if x1<1810 then node 12 elseif x1>=1810 then node 13 else 0

10 if x3<1523.5 then node 14 elseif x3>=1523.5 then node 15 else 1

11 if x2<2.5 then node 16 elseif x2>=2.5 then node 17 else 0

12 if x1<1415 then node 18 elseif x1>=1415 then node 19 else 0

13 class = 1

14 if x2<5.5 then node 20 elseif x2>=5.5 then node 21 else 1

15 class = 1

16 if x3<1731.5 then node 22 elseif x3>=1731.5 then node 23 else 1

17 if x12<503 then node 24 elseif x12>=503 then node 25 else 0

18 if x3<1309.5 then node 26 elseif x3>=1309.5 then node 27 else 0

19 class = 0

20 class = 1

21 class = 0

22 class = 0

23 if x6<731965 then node 28 elseif x6>=731965 then node 29 else 1

24 if x7<2224.5 then node 30 elseif x7>=2224.5 then node 31 else 0

25 class = 1

26 if x3<653.5 then node 32 elseif x3>=653.5 then node 33 else 0

27 if x1<1337 then node 34 elseif x1>=1337 then node 35 else 0

28 class = 1

29 if x3<1850 then node 36 elseif x3>=1850 then node 37 else 0

30 class = 0

31 if x3<1714.5 then node 38 elseif x3>=1714.5 then node 39 else 0

32 class = 0

33 if x1<672.5 then node 40 elseif x1>=672.5 then node 41 else 0

34 class = 1

35 if x3<1404 then node 42 elseif x3>=1404 then node 43 else 0

36 class = 1

37 class = 0

38 if x1<1642.5 then node 44 elseif x1>=1642.5 then node 45 else 0

39 if x1<2000 then node 46 elseif x1>=2000 then node 47 else 0

40 class = 1

41 if x1<1170 then node 48 elseif x1>=1170 then node 49 else 0

42 class = 0

43 class = 1

44 if x3<1637.5 then node 50 elseif x3>=1637.5 then node 51 else 0

45 class = 0

46 if x3<1918.5 then node 52 elseif x3>=1918.5 then node 53 else 0

47 class = 0

48 if x3<1107 then node 54 elseif x3>=1107 then node 55 else 0

49 class = 0

50 if x12<136.5 then node 56 elseif x12>=136.5 then node 57 else 0

51 if x1<1625 then node 58 elseif x1>=1625 then node 59 else 1

52 if x1<1722.5 then node 60 elseif x1>=1722.5 then node 61 else 0

53 class = 1

54 if x6<731970 then node 62 elseif x6>=731970 then node 63 else 0

55 class = 1

56 if x3<1587.5 then node 64 elseif x3>=1587.5 then node 65 else 0

57 class = 0

58 class = 1

59 if x3<1638.5 then node 66 elseif x3>=1638.5 then node 67 else 0

60 if x3<1753 then node 68 elseif x3>=1753 then node 69 else 1

61 class = 0

62 if x12<448 then node 70 elseif x12>=448 then node 71 else 0

63 if x3<679.5 then node 72 elseif x3>=679.5 then node 73 else 0

64 class = 0

65 class = 1

66 class = 1

67 class = 0

68 if x5<228.5 then node 74 elseif x5>=228.5 then node 75 else 0

69 class = 1

70 if x3<912 then node 76 elseif x3>=912 then node 77 else 0

71 class = 0

72 class = 1

73 if x10<3.5 then node 78 elseif x10>=3.5 then node 79 else 0

74 if x6<731961 then node 80 elseif x6>=731961 then node 81 else 0

75 class = 0

76 class = 0

77 if x1<915 then node 82 elseif x1>=915 then node 83 else 0

78 if x6<731973 then node 84 elseif x6>=731973 then node 85 else 0

79 if x3<908.5 then node 86 elseif x3>=908.5 then node 87 else 0

80 class = 0

81 if x6<731968 then node 88 elseif x6>=731968 then node 89 else 0

82 class = 1

83 class = 0

84 class = 0

85 if x1<965 then node 90 elseif x1>=965 then node 91 else 1

86 class = 0

87 if x1<912.5 then node 92 elseif x1>=912.5 then node 93 else 0

88 class = 1

89 if x10<3.5 then node 94 elseif x10>=3.5 then node 95 else 0

90 class = 1

91 class = 0

92 class = 1

93 class = 0

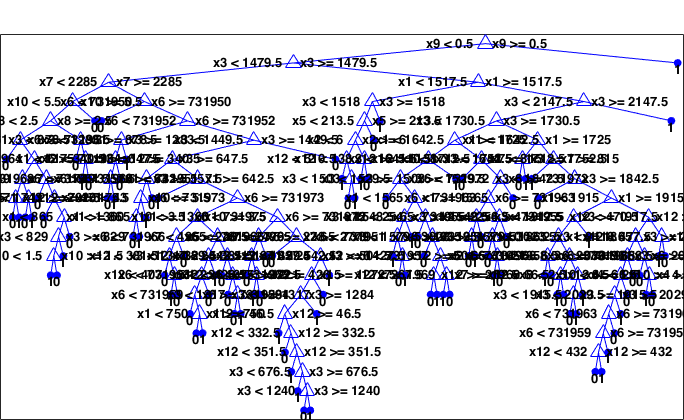
94 class = 1

95 class = 0

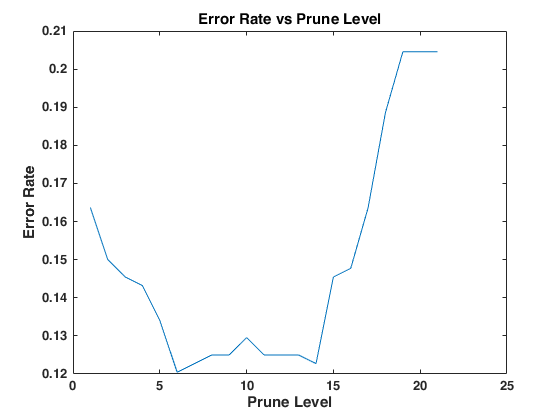
I would choose the best pruned tree since it is faster to compute and less complex if there is no great difference of miscalculation error between the minimum error tree.

## Using Entropy to Create a Full Classification Tree

Error rate of the tree is 0.1659 for my run. Here is the graph of the tree:

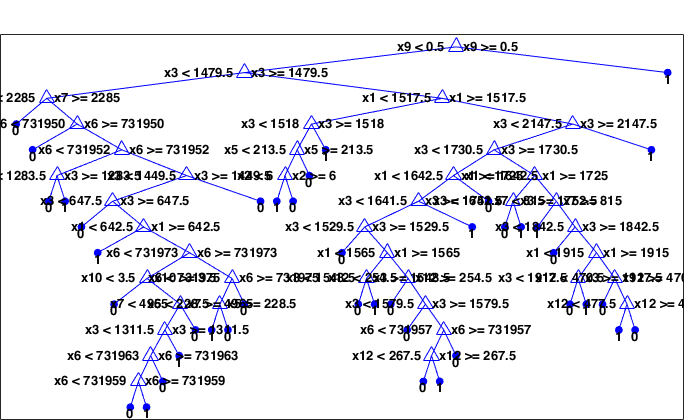


Here is the graph of Error Rates vs Prune Levels using Entropy:



Minimum error rate is 0.1205 for prune level 6 and it is in the defined limits.

Here is the graph of this tree:



### Click the arrow on the left to see rules of the tree as an output of MATLAB’s view function.

1 if x9<0.5 then node 2 elseif x9>=0.5 then node 3 else 0

2 if x3<1479.5 then node 4 elseif x3>=1479.5 then node 5 else 0

3 class = 1

4 if x7<2285 then node 6 elseif x7>=2285 then node 7 else 0

5 if x1<1517.5 then node 8 elseif x1>=1517.5 then node 9 else 0

6 class = 0

7 if x6<731950 then node 10 elseif x6>=731950 then node 11 else 0

8 if x3<1518 then node 12 elseif x3>=1518 then node 13 else 1

9 if x3<2147.5 then node 14 elseif x3>=2147.5 then node 15 else 0

10 class = 0

11 if x6<731952 then node 16 elseif x6>=731952 then node 17 else 0

12 if x5<213.5 then node 18 elseif x5>=213.5 then node 19 else 0

13 class = 1

14 if x3<1730.5 then node 20 elseif x3>=1730.5 then node 21 else 0

15 class = 1

16 if x3<1283.5 then node 22 elseif x3>=1283.5 then node 23 else 0

17 if x3<1449.5 then node 24 elseif x3>=1449.5 then node 25 else 0

18 if x2<6 then node 26 elseif x2>=6 then node 27 else 1

19 class = 0

20 if x1<1642.5 then node 28 elseif x1>=1642.5 then node 29 else 0

21 if x1<1725 then node 30 elseif x1>=1725 then node 31 else 0

22 class = 0

23 class = 1

24 if x3<647.5 then node 32 elseif x3>=647.5 then node 33 else 0

25 class = 0

26 class = 1

27 class = 0

28 if x3<1641.5 then node 34 elseif x3>=1641.5 then node 35 else 0

29 class = 0

30 if x3<1752.5 then node 36 elseif x3>=1752.5 then node 37 else 1

31 if x7<815 then node 38 elseif x7>=815 then node 39 else 0

32 class = 0

33 if x1<642.5 then node 40 elseif x1>=642.5 then node 41 else 0

34 if x3<1529.5 then node 42 elseif x3>=1529.5 then node 43 else 0

35 class = 1

36 class = 0

37 class = 1

38 class = 1

39 if x3<1842.5 then node 44 elseif x3>=1842.5 then node 45 else 0

40 class = 1

41 if x6<731973 then node 46 elseif x6>=731973 then node 47 else 0

42 class = 0

43 if x1<1565 then node 48 elseif x1>=1565 then node 49 else 0

44 class = 0

45 if x1<1915 then node 50 elseif x1>=1915 then node 51 else 0

46 if x10<3.5 then node 52 elseif x10>=3.5 then node 53 else 0

47 if x6<731975 then node 54 elseif x6>=731975 then node 55 else 0

48 if x3<1548.5 then node 56 elseif x3>=1548.5 then node 57 else 1

49 if x12<254.5 then node 58 elseif x12>=254.5 then node 59 else 0

50 if x3<1917.5 then node 60 elseif x3>=1917.5 then node 61 else 0

51 if x12<470.5 then node 62 elseif x12>=470.5 then node 63 else 0

52 class = 0

53 if x7<4965 then node 64 elseif x7>=4965 then node 65 else 0

54 if x5<228.5 then node 66 elseif x5>=228.5 then node 67 else 1

55 class = 0

56 class = 0

57 class = 1

58 class = 0

59 if x3<1579.5 then node 68 elseif x3>=1579.5 then node 69 else 0

60 class = 0

61 class = 1

62 class = 0

63 if x12<477.5 then node 70 elseif x12>=477.5 then node 71 else 0

64 if x3<1311.5 then node 72 elseif x3>=1311.5 then node 73 else 0

65 class = 0

66 class = 1

67 class = 0

68 class = 0

69 if x6<731957 then node 74 elseif x6>=731957 then node 75 else 0

70 class = 1

71 class = 0

72 if x6<731963 then node 76 elseif x6>=731963 then node 77 else 0

73 class = 1

74 if x12<267.5 then node 78 elseif x12>=267.5 then node 79 else 1

75 class = 0

76 if x6<731959 then node 80 elseif x6>=731959 then node 81 else 0

77 class = 0

78 class = 0

79 class = 1

80 class = 0

81 class = 1

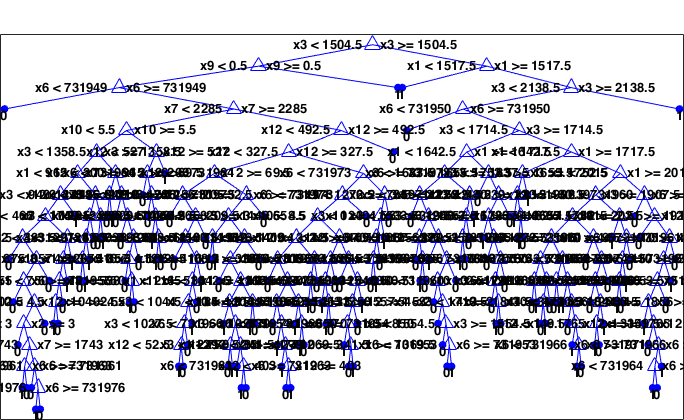
### 

## Assuming Misclassification Costs

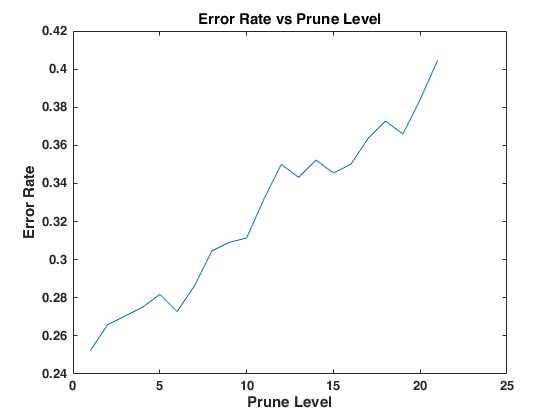
Assume misclassification costs as the following:

* When true label is “Delayed” - $50
* When true label is “Ontime” - $5

I used Gini Index to calculate error rates.  
Error rate of the tree is 0.2659. Here is the graph of the tree:

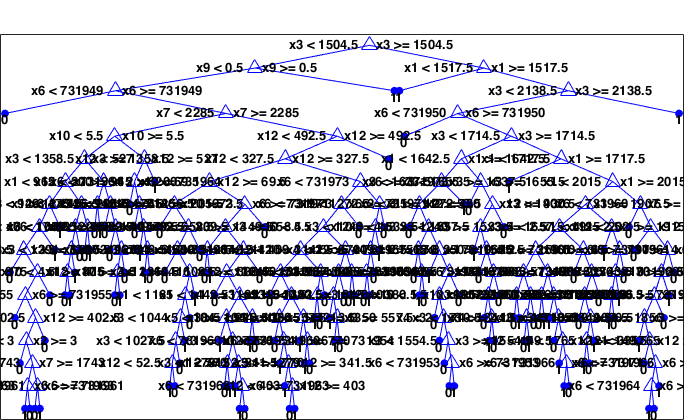


Here is the graph of the error rates vs prune levels:



Minimum error rate is 0.2523 for prune level 1 but it is not in the defined limits. Thus the best prune level is stated as 2 with 0.2659 error rate.

Here is the graph of this tree:



### Click the arrow on the left to see rules of the tree as an output of MATLAB’s view function.

1 if x3<1504.5 then node 2 elseif x3>=1504.5 then node 3 else 1

2 if x9<0.5 then node 4 elseif x9>=0.5 then node 5 else 1

3 if x1<1517.5 then node 6 elseif x1>=1517.5 then node 7 else 1

4 if x6<731949 then node 8 elseif x6>=731949 then node 9 else 1

5 class = 1

6 class = 1

7 if x3<2138.5 then node 10 elseif x3>=2138.5 then node 11 else 1

8 class = 0

9 if x7<2285 then node 12 elseif x7>=2285 then node 13 else 1

10 if x6<731950 then node 14 elseif x6>=731950 then node 15 else 1

11 class = 1

12 if x10<5.5 then node 16 elseif x10>=5.5 then node 17 else 0

13 if x12<492.5 then node 18 elseif x12>=492.5 then node 19 else 1

14 class = 0

15 if x3<1714.5 then node 20 elseif x3>=1714.5 then node 21 else 1

16 if x3<1358.5 then node 22 elseif x3>=1358.5 then node 23 else 1

17 if x12<527 then node 24 elseif x12>=527 then node 25 else 0

18 if x12<327.5 then node 26 elseif x12>=327.5 then node 27 else 1

19 class = 0

20 if x1<1642.5 then node 28 elseif x1>=1642.5 then node 29 else 1

21 if x1<1717.5 then node 30 elseif x1>=1717.5 then node 31 else 1

22 if x1<965 then node 32 elseif x1>=965 then node 33 else 0

23 if x12<200 then node 34 elseif x12>=200 then node 35 else 1

24 if x6<731964 then node 36 elseif x6>=731964 then node 37 else 0

25 class = 1

26 if x12<69.5 then node 38 elseif x12>=69.5 then node 39 else 1

27 if x6<731973 then node 40 elseif x6>=731973 then node 41 else 1

28 if x3<1637.5 then node 42 elseif x3>=1637.5 then node 43 else 1

29 if x3<1655.5 then node 44 elseif x3>=1655.5 then node 45 else 0

30 class = 1

31 if x1<2015 then node 46 elseif x1>=2015 then node 47 else 1

32 if x3<949 then node 48 elseif x3>=949 then node 49 else 1

33 if x12<476 then node 50 elseif x12>=476 then node 51 else 0

34 if x12<45 then node 52 elseif x12>=45 then node 53 else 1

35 if x3<1456.5 then node 54 elseif x3>=1456.5 then node 55 else 1

36 class = 0

37 if x12<205.5 then node 56 elseif x12>=205.5 then node 57 else 0

38 if x10<2.5 then node 58 elseif x10>=2.5 then node 59 else 1

39 if x6<731973 then node 60 elseif x6>=731973 then node 61 else 0

40 if x6<731971 then node 62 elseif x6>=731971 then node 63 else 1

41 if x1<1272.5 then node 64 elseif x1>=1272.5 then node 65 else 1

42 if x2<3.5 then node 66 elseif x2>=3.5 then node 67 else 1

43 class = 1

44 class = 0

45 if x12<430 then node 68 elseif x12>=430 then node 69 else 0

46 if x3<1907.5 then node 70 elseif x3>=1907.5 then node 71 else 1

47 if x6<731960 then node 72 elseif x6>=731960 then node 73 else 1

48 if x12<466 then node 74 elseif x12>=466 then node 75 else 1

49 class = 1

50 if x7<1749 then node 76 elseif x7>=1749 then node 77 else 0

51 if x3<1057.5 then node 78 elseif x3>=1057.5 then node 79 else 1

52 class = 0

53 class = 1

54 class = 0

55 if x10<2.5 then node 80 elseif x10>=2.5 then node 81 else 1

56 class = 0

57 if x12<209.5 then node 82 elseif x12>=209.5 then node 83 else 0

58 class = 0

59 if x3<1449.5 then node 84 elseif x3>=1449.5 then node 85 else 1

60 if x3<658.5 then node 86 elseif x3>=658.5 then node 87 else 1

61 class = 0

62 if x10<4.5 then node 88 elseif x10>=4.5 then node 89 else 1

63 class = 0

64 if x3<1240 then node 90 elseif x3>=1240 then node 91 else 1

65 class = 0

66 class = 0

67 if x3<1533.5 then node 92 elseif x3>=1533.5 then node 93 else 1

68 if x3<1657.5 then node 94 elseif x3>=1657.5 then node 95 else 1

69 class = 0

70 if x4<2.5 then node 96 elseif x4>=2.5 then node 97 else 1

71 if x1<1915 then node 98 elseif x1>=1915 then node 99 else 1

72 class = 0

73 if x12<202.5 then node 100 elseif x12>=202.5 then node 101 else 1

74 if x3<728.5 then node 102 elseif x3>=728.5 then node 103 else 0

75 class = 1

76 if x3<1291 then node 104 elseif x3>=1291 then node 105 else 0

77 class = 0

78 class = 0

79 class = 1

80 class = 0

81 if x3<1500.5 then node 106 elseif x3>=1500.5 then node 107 else 1

82 class = 1

83 if x6<731967 then node 108 elseif x6>=731967 then node 109 else 0

84 if x3<648.5 then node 110 elseif x3>=648.5 then node 111 else 1

85 class = 0

86 if x1<1410 then node 112 elseif x1>=1410 then node 113 else 0

87 if x12<314.5 then node 114 elseif x12>=314.5 then node 115 else 1

88 if x12<409 then node 116 elseif x12>=409 then node 117 else 1

89 if x4<2.5 then node 118 elseif x4>=2.5 then node 119 else 1

90 if x1<870 then node 120 elseif x1>=870 then node 121 else 1

91 class = 1

92 if x6<731967 then node 122 elseif x6>=731967 then node 123 else 0

93 if x1<1565 then node 124 elseif x1>=1565 then node 125 else 1

94 class = 1

95 if x12<250.5 then node 126 elseif x12>=250.5 then node 127 else 0

96 if x10<3.5 then node 128 elseif x10>=3.5 then node 129 else 1

97 if x6<731961 then node 130 elseif x6>=731961 then node 131 else 0

98 class = 1

99 if x10<4.5 then node 132 elseif x10>=4.5 then node 133 else 1

100 class = 0

101 if x6<731975 then node 134 elseif x6>=731975 then node 135 else 1

102 class = 0

103 if x1<875 then node 136 elseif x1>=875 then node 137 else 1

104 if x10<4.5 then node 138 elseif x10>=4.5 then node 139 else 0

105 class = 1

106 if x12<416 then node 140 elseif x12>=416 then node 141 else 1

107 class = 0

108 class = 1

109 class = 0

110 class = 0

111 if x3<1364.5 then node 142 elseif x3>=1364.5 then node 143 else 1

112 class = 0

113 class = 1

114 if x12<108.5 then node 144 elseif x12>=108.5 then node 145 else 1

115 class = 0

116 if x12<336.5 then node 146 elseif x12>=336.5 then node 147 else 1

117 if x10<1.5 then node 148 elseif x10>=1.5 then node 149 else 0

118 if x6<731967 then node 150 elseif x6>=731967 then node 151 else 1

119 if x3<684.5 then node 152 elseif x3>=684.5 then node 153 else 1

120 if x3<645 then node 154 elseif x3>=645 then node 155 else 1

121 class = 0

122 class = 0

123 if x6<731970 then node 156 elseif x6>=731970 then node 157 else 1

124 class = 1

125 if x10<3.5 then node 158 elseif x10>=3.5 then node 159 else 1

126 class = 0

127 if x6<731968 then node 160 elseif x6>=731968 then node 161 else 1

128 if x1<1780 then node 162 elseif x1>=1780 then node 163 else 1

129 if x3<1743.5 then node 164 elseif x3>=1743.5 then node 165 else 1

130 class = 0

131 if x6<731969 then node 166 elseif x6>=731969 then node 167 else 0

132 if x10<2.5 then node 168 elseif x10>=2.5 then node 169 else 1

133 class = 0

134 if x3<2130 then node 170 elseif x3>=2130 then node 171 else 1

135 class = 0

136 if x6<731955 then node 172 elseif x6>=731955 then node 173 else 1

137 class = 0

138 class = 0

139 class = 1

140 class = 0

141 class = 1

142 if x1<1165 then node 174 elseif x1>=1165 then node 175 else 1

143 class = 1

144 class = 0

145 if x1<1442.5 then node 176 elseif x1>=1442.5 then node 177 else 1

146 class = 0

147 if x3<1315 then node 178 elseif x3>=1315 then node 179 else 1

148 if x3<840.5 then node 180 elseif x3>=840.5 then node 181 else 1

149 class = 0

150 if x12<380.5 then node 182 elseif x12>=380.5 then node 183 else 1

151 class = 0

152 class = 0

153 class = 1

154 class = 0

155 class = 1

156 class = 1

157 class = 0

158 if x10<1.5 then node 184 elseif x10>=1.5 then node 185 else 1

159 class = 0

160 if x10<2.5 then node 186 elseif x10>=2.5 then node 187 else 1

161 class = 0

162 if x3<1722.5 then node 188 elseif x3>=1722.5 then node 189 else 1

163 if x6<731966 then node 190 elseif x6>=731966 then node 191 else 1

164 if x12<208.5 then node 192 elseif x12>=208.5 then node 193 else 0

165 if x1<1760 then node 194 elseif x1>=1760 then node 195 else 1

166 if x10<3.5 then node 196 elseif x10>=3.5 then node 197 else 1

167 class = 0

168 if x6<731952 then node 198 elseif x6>=731952 then node 199 else 0

169 class = 1

170 if x5<228.5 then node 200 elseif x5>=228.5 then node 201 else 1

171 class = 1

172 class = 0

173 if x12<402.5 then node 202 elseif x12>=402.5 then node 203 else 1

174 if x3<1044 then node 204 elseif x3>=1044 then node 205 else 1

175 class = 0

176 if x5<184 then node 206 elseif x5>=184 then node 207 else 1

177 if x10<1.5 then node 208 elseif x10>=1.5 then node 209 else 0

178 if x5<213.5 then node 210 elseif x5>=213.5 then node 211 else 1

179 if x1<1350 then node 212 elseif x1>=1350 then node 213 else 1

180 class = 1

181 class = 0

182 class = 1

183 if x3<557.5 then node 214 elseif x3>=557.5 then node 215 else 1

184 class = 0

185 if x4<2 then node 216 elseif x4>=2 then node 217 else 1

186 class = 0

187 if x3<1710.5 then node 218 elseif x3>=1710.5 then node 219 else 1

188 class = 0

189 class = 1

190 if x3<1844.5 then node 220 elseif x3>=1844.5 then node 221 else 1

191 class = 0

192 class = 0

193 class = 1

194 class = 1

195 if x10<4.5 then node 222 elseif x10>=4.5 then node 223 else 1

196 class = 0

197 if x3<1856 then node 224 elseif x3>=1856 then node 225 else 1

198 class = 1

199 class = 0

200 if x3<2030.5 then node 226 elseif x3>=2030.5 then node 227 else 1

201 class = 0

202 if x2<3 then node 228 elseif x2>=3 then node 229 else 1

203 class = 0

204 if x3<1027.5 then node 230 elseif x3>=1027.5 then node 231 else 1

205 class = 1

206 class = 1

207 if x6<731960 then node 232 elseif x6>=731960 then node 233 else 1

208 class = 1

209 class = 0

210 class = 0

211 if x1<770 then node 234 elseif x1>=770 then node 235 else 1

212 class = 1

213 if x6<731954 then node 236 elseif x6>=731954 then node 237 else 1

214 class = 1

215 class = 0

216 class = 0

217 if x3<1554.5 then node 238 elseif x3>=1554.5 then node 239 else 1

218 class = 1

219 class = 0

220 class = 0

221 class = 1

222 class = 0

223 if x12<149.5 then node 240 elseif x12>=149.5 then node 241 else 1

224 if x1<1765 then node 242 elseif x1>=1765 then node 243 else 0

225 class = 1

226 class = 1

227 if x12<335 then node 244 elseif x12>=335 then node 245 else 1

228 class = 0

229 if x7<1743 then node 246 elseif x7>=1743 then node 247 else 1

230 if x12<52.5 then node 248 elseif x12>=52.5 then node 249 else 1

231 class = 0

232 if x3<1279 then node 250 elseif x3>=1279 then node 251 else 0

233 if x1<770 then node 252 elseif x1>=770 then node 253 else 1

234 class = 1

235 if x12<341.5 then node 254 elseif x12>=341.5 then node 255 else 1

236 class = 1

237 class = 0

238 class = 0

239 if x6<731953 then node 256 elseif x6>=731953 then node 257 else 1

240 class = 1

241 if x6<731966 then node 258 elseif x6>=731966 then node 259 else 1

242 class = 1

243 class = 0

244 class = 0

245 if x6<731971 then node 260 elseif x6>=731971 then node 261 else 1

246 if x6<731963 then node 262 elseif x6>=731963 then node 263 else 1

247 if x6<731961 then node 264 elseif x6>=731961 then node 265 else 1

248 class = 1

249 class = 0

250 class = 0

251 class = 1

252 class = 1

253 if x6<731963 then node 266 elseif x6>=731963 then node 267 else 1

254 class = 1

255 if x12<403 then node 268 elseif x12>=403 then node 269 else 0

256 class = 0

257 class = 1

258 class = 1

259 class = 0

260 if x6<731964 then node 270 elseif x6>=731964 then node 271 else 1

261 class = 1

262 class = 1

263 class = 0

264 class = 0

265 class = 1

266 class = 1

267 class = 0

268 class = 0

269 class = 1

270 class = 1

271 class = 0

### 

Misclassification error cost is: $4110