

## 实验 2:

### PartI: 构建决策树、使用决策树

#### 1.预测隐形眼镜类型

使用 `lenses.txt` 中的隐形眼镜数据集，采用第三章中介绍的 ID3 算法构建决策树。

使用决策树，输入几组隐形眼镜特征数据，

例如: `'young','hyper','no','reduced', 'pre','hyper','no','normal'`; 等进行测试，预测隐形眼镜类型。

2 利用机器学习库 `sklearn` 中的决策树分类器 `DecisionTreeClassifier` 对 `Iris` 数据集进行交叉验证，测试其准确率。

3.利用机器学习库 `sklearn` 中的随机森林分类器 `RandomForestClassifier` 对 `Iris` 数据集进行交叉验证，测试其准确率。

### PartII: Logistic 回归；使用 Logistic 回归预测病马的死亡率

4.使用 `horseColicTraining.txt` 文件作为训练集（每行包含了病马的 20 个特征和是否死亡的标签），`horseColicTest.txt` 作为测试集，利用 Logistic 回归预测病马的死亡率。计算多次迭代后的平均错误率。

5.采用 `sklearn.linear_model.LogisticRegression` 实现上述数据集 Logistic 回归预测病马死亡率。

6.通过访问：

<http://archive.ics.uci.edu/ml/machine-learning-databases/horse-colic/horse-colic.data> 中提供的病马原始数据，采用 sklearn.impute 中 SimpleImputer 对原始缺失数据进行处理（处理策略不限定，如：特殊值、均值等）。

```
2 1 530101 38.50 66 28 3 3 ? 2 5 4 4 ? ? ? 3 5 45.00 8.40 ? ? 2 2 11300 00000 00000 21 1 534817 39.2 88 20 ? ? 4 1 3 4 2 ? ? ? 4 2 50 85
00000 2 1 1 528031 37.20 42 12 2 1 1 1 3 3 3 1 ? 4 5 ? 7.00 ? ? 1 2 04124 00000 00000 2 2 9 5291329 38.00 92 28 1 1 2 1 1 3 2 3 ? 7.20
00000 00000 2 1 1 533692 38.1 60 12 3 3 3 1 ? 4 3 3 2 2 ? ? 51 65 ? ? 1 1 03111 00000 00000 2 2 1 529518 37.80 60 42 ? ? ? 1 ? ? ? ? ?
00000 00000 2 2 1 529493 38.30 112 16 ? 3 5 2 ? ? 1 1 2 ? ? 5 51.00 6.00 2 1.00 3 2 05205 00000 00000 1 1 1 533847 37.8 72 ? ? 3 ? 1 5 3
0 00000 2 1 1 528890 38.90 80 44 3 3 3 2 2 3 3 2 2 7.00 3 1 54.00 6.50 3 ? 2 1 07111 00000 00000 2 2 1 529642 37.20 84 48 3 3 5 2 4 1 2 1
00000 2 2 1 530438 37.80 48 32 1 1 3 1 2 1 ? 1 1 ? 4 5 37.00 6.70 ? ? 1 2 02124 00000 00000 2 2 1 527957 38.50 60 ? ? 2 2 1 1 1 2 2 2 1 ? 1
2 00000 00000 00000 1 1 1 5299253 37.70 56 ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? 2 1 02113 00000 00000 2 1 1 528469 38.10 52 24 1 1 5 1 4 3 1
000 00000 2 1 1 529849 37.80 60 80 1 3 2 2 2 3 3 ? 2 5.50 4 ? 40.00 4.50 2 ? 1 1 05206 00000 00000 1 2 1 529304 38.00 54 30 2 3 3 3 3 1 2
3205 00000 00000 2 2 1 530319 ? 84 36 ? ? 3 1 ? 3 1 2 1 ? 3 2 44.00 8.50 ? ? 1 1 03111 00000 00000 2 1 1 534145 38.1 88 32 3 3 4 1 2 3 3
00000 2 1 1 533696 38.7 60 32 4 3 2 2 4 4 4 ? ? ? 4 5 53 64 3 2 3 1 03205 00000 00000 2 1 9 5297379 38.40 84 40 3 3 2 1 3 3 3 1 1 ? ? ? 3
3 2 1 1 ? 3 4 39.00 6.00 2 ? 1 1 02206 00000 00000 1 1 1 534556 38.3 80 40 ? ? 6 2 4 3 1 ? 2 ? 1 4 67 10.2 2 1 3 1 02208 00000 00000 2 2
4 3 1 2 1 ? 3 5 43 73 2 1 ? 1 03111 00000 00000 2 2 1 530294 37.90 40 24 1 1 1 1 2 3 1 ? ? ? ? 3 40.00 5.70 ? ? 1 1 00400 00000 00000 2 1
5 75 81 ? ? 3 2 01400 00000 00000 2 2 1 535246 ? 80 32 3 3 2 1 2 3 3 2 1 ? 3 ? 50 80 ? ? 1 1 00000 00000 00000 2 1 1 534115 38.2 48 ? 1 3
5.00 1 2 00000 00000 00000 1 1 1 533928 37.8 60 24 1 ? 3 2 ? 4 4 2 3 2 ? 5 52 75 ? ? 3 1 02205 00000 00000 2 2 1 528248 38.00 42 40 3 1 1
535085 ? 65 24 ? ? ? 2 5 ? 4 3 1 ? ? 5 ? ? ? 3 1 03205 00000 00000 2 2 1 532985 37.5 44 20 1 1 3 1 ? 1 1 ? ? ? 1 ? 35 7.2 ? ? 1 2 00000
00 7.50 ? ? 1 2 03111 00000 00000 1 1 1 5279822 38.00 ? 24 3 3 6 2 5 ? 4 1 1 ? ? ? 68.00 7.80 ? ? 2 1 02205 00000 00000 2 2 1 534857 37.8
? ? 2 1 03205 00000 00000 2 2 1 534163 37.9 88 24 1 1 2 1 2 2 1 ? ? ? 4 1 37 56 ? ? 1 2 00000 00000 00000 2 2 1 533697 38.0 44 12 3 1 1 ?
24 00000 00000 2 1 1 535137 38.5 96 30 2 3 4 2 4 4 3 2 1 ? 3 5 50 65 ? ? 1 1 06111 03112 00000 2 1 1 530297 ? ? ? ? ? ? ? ? ? ? ? ?
02209 00000 00000 1 1 1 535240 38.1 40 36 1 2 2 1 2 2 ? ? ? ? ? ? ? ? ? ? 3 1 03112 00000 00000 2 1 1 529736 ? 52 28 3 3 4 1 3 4 3 2 1 ?
1 1 534280 38.5 104 40 3 3 ? 1 4 3 4 ? ? ? ? ? ? ? ? ? 1 1 02124 00000 00000 2 2 1 530028 39.50 92 28 3 3 6 1 5 4 1 ? 3 ? 4 ? 72.00 6.40
00000 00000 1 1 1 530354 ? 36 32 1 1 4 1 5 3 3 2 3 4.00 ? 4 41.00 5.90 ? ? 2 1 02205 00000 00000 2 1 1 5281091 38.40 92 20 1 ? ? 2 ? 3 3
00000 00000 1 1 1 533738 ? 80 ? 3 3 3 1 4 4 4 ? ? ? 4 5 43 70 ? ? 1 1 03111 00000 00000 2 1 9 5294539 38.80 184 84 1 ? 1 1 4 1 3 ? ? ? 2
00000 00000 2 2 1 529865 38.60 68 20 2 1 3 1 3 3 2 1 1 ? 1 5 38.00 6.50 1 ? 1 2 05124 00000 00000 1 2 1 527829 38.30 54 18 3 1 2 1 2 3 2
00000 00000 2 2 1 528019 37.80 40 16 1 1 1 1 1 1 ? ? 1 1 37.00 6.80 ? ? 1 2 00000 00000 00000 2 2 1 529172 38.20 56 40 4 3 1 1 2 4 3
08400 00000 00000 12 1 535054 38.6 45 16 2 1 2 1 1 1 ? ? ? ? 1 1 43 58 ? ? 1 2 00000 00000 00000 21 1 528890 38.90 80 44 3 3 3 1 2 3 3 2
```