CS461 Homework 3

Mannan Shukla

Due: Nov. 19, 11:59 PM
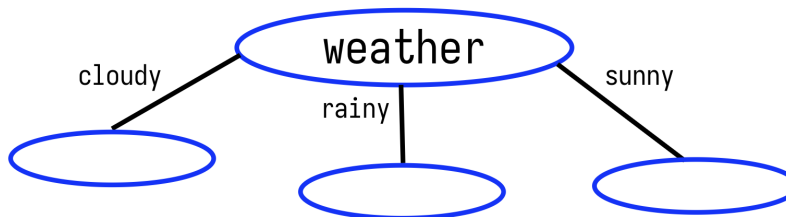
1. Decision Tree

1.1 Information Gain and Root Node Selection

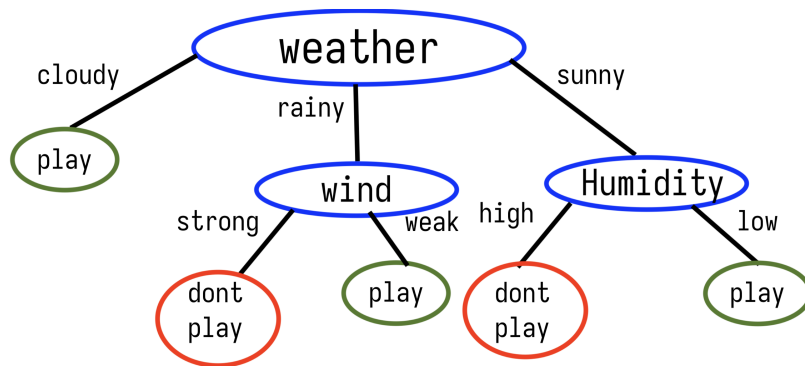The initial entropy of the Play is: $H(Play) = 1.0000$

- **Weather:** $IG(\text{Weather}) = 0.4000$

- **Temperature:** $IG(\text{Temperature}) = 0.1145$

- **Humidity:** $IG(\text{Humidity}) = 0.0349$

- **Wind:** $IG(\text{Wind}) = 0.1245$

Since **Weather** has the highest Information Gain 0.4000, it is selected as the root node of the decision tree.

1.2 Constructing the Decision Tree

The decision tree is shown below:

1.3 Pruning (Extra Points)

mrleebc1

2. Perceptron

2.1 Single Data Point

With a step size of 1, it will always take *one* iteration to classify the data point.

2.2 Random Initialization

Given a randomly initialized weight vector, if w_0 correctly classifies the point, then 0 iterations will be required, otherwise, we must consider another case:

2.3 Iterative Updates

3. Gaussian Discriminant Analysis (GDA)

3.1 Estimating Parameters

Class	Mean (μ)	Variance (σ^2)
Class +1	-0.0722	1.3031
Class -1	0.9402	1.9426

Table 1: Mean and variance for each class.

3.2 Test Accuracy

The test accuracy is 61%

3.3 Improving the Classifier

As currently constructed, we use MLE estimation. MLE only looks at likelihood and ignores prior probabilities. Using MAP rule, we can increase our effectiveness. With MAP, accuracy went up to 90%

3.4 2D GDA Statistics

Class	Mean (μ)	Covariance (Σ)
Class +1	$[0.0130754, 0.06295251]$	$\begin{bmatrix} 0.98285498 & 0.00612046 \\ 0.00612046 & 1.05782804 \end{bmatrix}$
Class -1	$[-0.02313942, -0.02114952]$	$\begin{bmatrix} 1.00329037 & -0.01142356 \\ -0.01142356 & 4.97693356 \end{bmatrix}$

Table 2: Mean and covariance for each class.

3.5 2D GDA Predictions

The test accuracy is 84%

3.6 Density-Based Accuracy (Extra Points)

The test accuracy is 85% which shows us that GDA is suitable. GDA provides a reasonable framework for classification because it balances robustness, simplicity, and efficiency, achieving near-optimal accuracy even when the data deviates slightly from Gaussian assumptions. A more complex mixture model wasn't that much of an improvement.

4. Logistic Regression

4.1 Data Preprocessing

4.2 Dimensionality Reduction

4.3 Gradient Descent

4.4 Train and Test Accuracy

4.5 Extra Points: Email Classification