Malcolm Snyder/John Santaguida

Projects in ML Assignment 3

Dr. Kunda

1. Description of dataset:
   1. We researched the most popular general genomes of fish, and found that 94% of fish are categorized under:
      1. Salmon
      2. Tuna
      3. Shark
      4. Catfish
      5. Flounder
      6. Bass
      7. Trout
      8. Eel
   2. Having decided to use the most popular genomes of fish as our initial target concept, we used Google to search for images of the desired fish. We then found 11-13 pictures of each, ranging from casual to scientific, gathering 100 in total. We looked for pictures that were clear, even choosing some drawings instead of photographs to maximize feature clarity.
   3. We analyzed the photos and found that there was variance between species in:
      1. Color
      2. Secondary Color
      3. Dorsal fin shape
      4. Number of lower fins
      5. Tailfin shape
      6. Number of eyes visible
      7. Visible Spots
      8. Visible Stripes
      9. Visible Teeth
      10. Visible Whiskers
   4. We manually looked at the 100 picture of fish and assigned the appropriate values for the input vectors.
   5. We used random number generation to select a test set of 20 pictures from the previous 100.
2. Our ID3 algorithm outputs a decision tree in the form of:
   1. {Root : { RootVal : Output, RootVal: {Nextlabel: {NextVal: Output, NextVal: Output}}, RootVal: Output}}
      1. Any *Output* can be switched with a *{Nextlabel: {NextVal: Output, NextVal: Output}}* and vice versa
   2. The input for our algorithm is currently received as a CSV – a comma-delimited file with each line representing an input vector, exported directly from our excel sheet.
3. Running on the PlayTennis example dataset we get the output:

{'Outlook': {'Overcast': 'Yes', 'Sunny': {'Humidity': {'High': 'No', 'Normal': 'Yes'}}, 'Rain': {'Wind ': {'Strong': 'No', 'Weak': 'Yes'}}}}

1. Running on our own provided dataset with manually created input vectors, (emailed to instructor) we get the output:

{'Color': {'pink': 'Salmon', 'brown': {'Tail Shape': {'flat': {'# Eyes Visible': {'1': 'Trout', '2': 'Flounder'}}, 'single': 'Eel', 'none': 'Eel', 'round': 'Flounder', 'point': {'Visible Teeth?': {'T': 'Salmon', 'F': 'Trout'}}}}, 'blue': 'Tuna', 'grey': {'Tail Shape': {'flat': 'Catfish', 'single': 'Eel', 'round': {'Dorsal Fin Type': {'single': 'Flounder', 'trigger': 'Catfish'}}, 'none': 'Eel', 'point': {'Whiskers?': {'T': 'Catfish', 'F': {'Secondary Color': {'white': 'Shark', 'black': 'Tuna'}}}}}}, 'yellow': {'Secondary Color': {'brown': 'Bass', 'yellow': 'Eel', 'red': 'Trout', 'silver': 'Trout'}}, 'green': {'Dorsal Fin Type': {'single': 'Bass', 'trigger': 'Trout', 'split': 'Bass'}}, 'black': {'Dorsal Fin Type': {'single': 'Eel', 'trigger': 'Salmon'}}, 'white': {'Dorsal Fin Type': {'single': 'Eel', 'trailing': 'Catfish'}}, 'silver': {'Spots?': {'T': {'Stripes?': {'T': 'Trout', 'F': 'Salmon'}}, 'F': 'Tuna'}}}}

* + 1. Further breakdown and analysis of the accuracy of these results is presented within the excel sheet attached to and detailing our datasets.