Assignment 2 (Hands on experience with Classification)

Use http://www.cse.msu.edu/~ptan/dmbook/tutorials/tutorial6/tutorial6.html

The following question use the vertebrate data set.

1. Modify the classifier to classify "reptile" and "not-a-reptile". Find the best classifier for this task using the training data in Appendix 1. Maximum depth is 2.

   Plot the classifier in the format shown in Appendix 2. What is its training accuracy? Use the testing data in Appendix 3. What is its testing accuracy? (30%)

2. Repeat your experiment with un-restricted maximum depth. What are your findings and why?

   (20%)

The following question use the model overfitting data set.

3. The depth of the tree is a hyper-parameter of the classification method. One method to find the best set of hyper-parameters is to split the training set into training subset and validation subset.

   Use 50% of the data as training set, 25% of the data as validation set and the remaining 25% of the data as the test set. Find the best depth of the tree. What is the testing accuracy using this depth? (50%)

What you should hand in

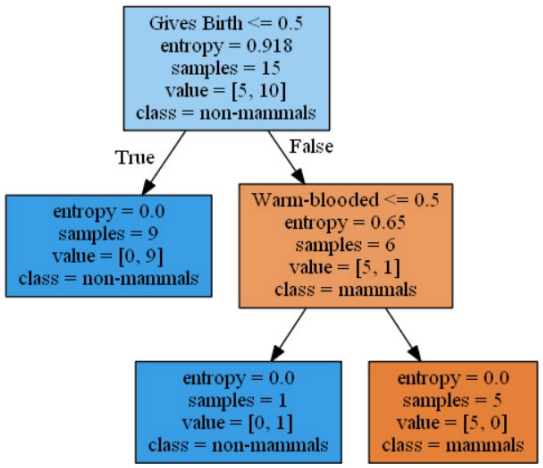
One Jupyter notebook containing your code, output, and answers.

Appendix 1

The vertebrate data set can be found in http://www.cse.msu.edu/~ptan/dmbook/software/

	Name	Warm-blooded	Gives Birth	Aquatic Creature	Aerial Creature	Has Legs	Hibernates	Class
0	human	1	1	0	0	1	0	mammals
1	python	0	0	0	0	0	1	reptiles
2	salmon	0	0	1	0	0	0	fishes
3	whale	1	1	1	0	0	0	mammals
4	frog	0	0	1	0	1	1	amphibians
5	komodo	0	0	0	0	1	0	reptiles
6	bat	1	1	0	1	1	1	mammals
7	pigeon	1	0	0	1	1	0	birds
8	cat	1	1	0	0	1	0	mammals
9	leopard shark	0	1	1	0	0	0	fishes
10	turtle	0	0	1	0	1	0	reptiles
11	penguin	1	0	1	0	1	0	birds
12	porcupine	1	1	0	0	1	1	mammals
13	eel	0	0	1	0	0	0	fishes
14	salamander	0	0	1	0	1	1	amphibians

Appendix 2



Appendix 3

	Name	Warm-blooded	Gives Birth	Aquatic Creature	Aerial Creature	Has Legs	Hibernates
0	gila monster	0	0	0	0	1	1
1	platypus	1	0	0	0	1	1
2	owl	1	0	0	1	1	0
3	dolphin	1	1	1	0	0	0