<Decision Tree - Learning Process>

* Process of building a decision tree

① Decide what feature to use at root node (first node at the very top of the decision tree)

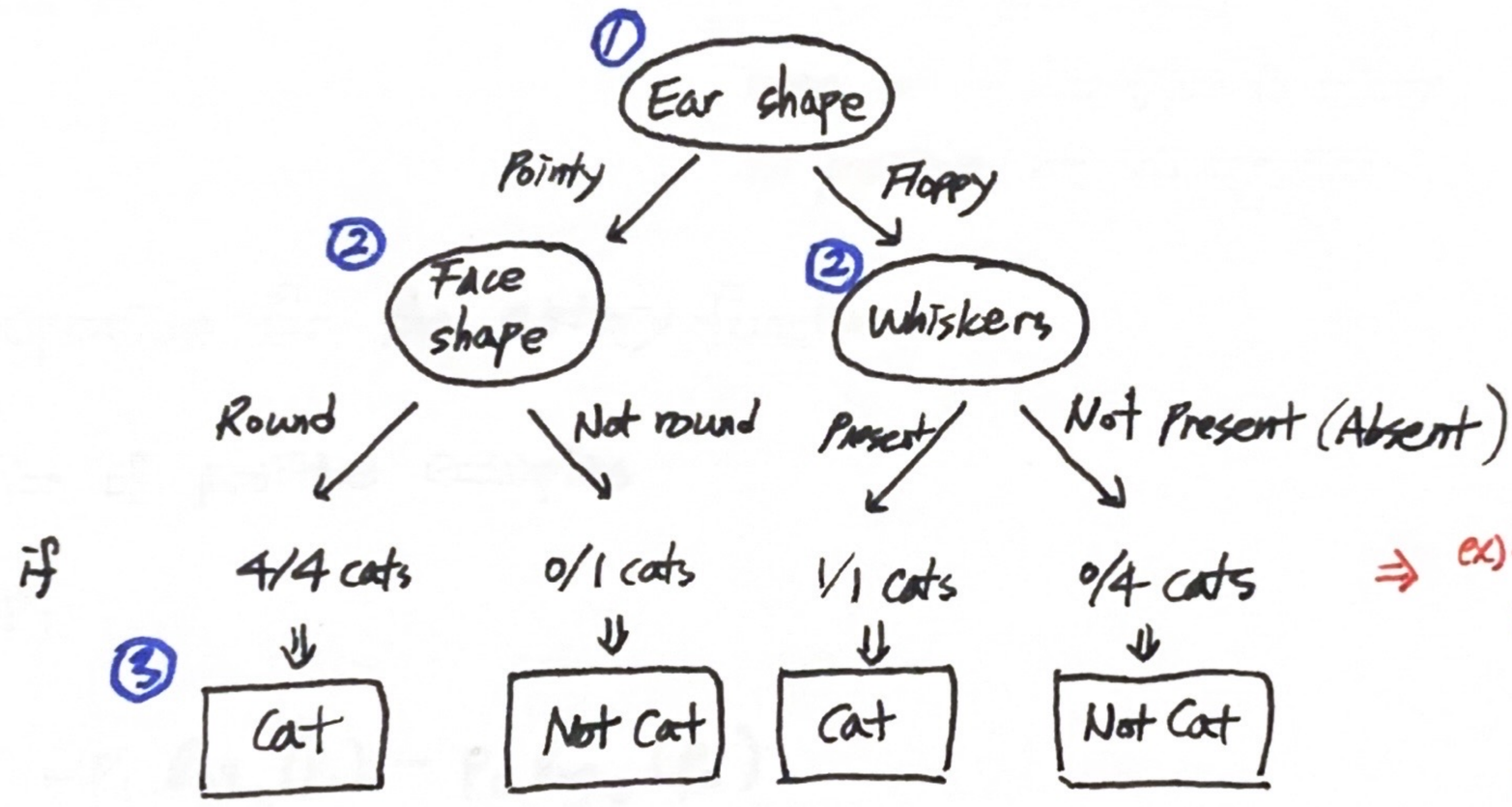
ex) Ear shape feature for root node

⇒ means looking at all of training examples (5 cats, 5 dogs) and split them according to the value of the ear shape feature

② Decide what feature to use at next node (decision node)

ex) Face shape for left branch, whiskers for right branch

③ If the node has no longer a mix of some classes (completely pure), then create that node as leaf node making a prediction



⇒ ex) 4/4 cats
: 현재 노드에 4개의 example이 있는데 4개모두 고양이임
예) Cat
: 현재 노드에 1개의 example이 있는데 1개 모두 고양이가 아님

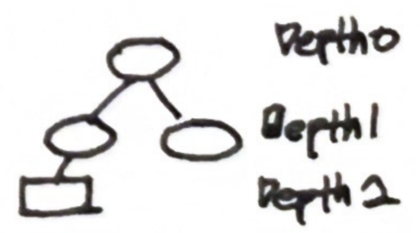
* Key decisions that I have to make at various steps during algorithm

① Decision 1: How to choose what feature to split on at each node?

⇒ decision trees will choose what feature to split on in order to try to "maximize purity" ⇒ "entropy" (minimize impurity)

② Decision 2: When do you stop splitting?

* criteria
- When a node is 100% one class
- When splitting a node will result in the tree exceeding a maximum depth
- When improvements in purity score are below a threshold
- When number of examples in a node is below a threshold.



→ Reason of limiting depth
- tree doesn't get too big
- small tree has less prone to overfitting