✅ **Congratulations! You passed!**

Grade received 100%	**Latest Submission Grade** 100%	**To pass** 80% or higher	**Go to next item**

1.

Entropy as a measure of impurity

p_1 = fraction of examples that are cats

$$p_0 = 1 - p_1$$

$$H(p_1) = -p_1 log_2(p_1) - p_0 log_2(p_0)$$
$$= -p_1 \ log_2(p_1) - (1 - p_1)log_2(1 - p_1)$$

Note: "0 log(0)" = 0

1 / 1 point
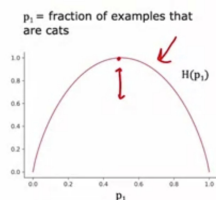
Recall that entropy was defined in lecture as H(p_1) = - p_1 log_2(p_1) - p_0 log_2(p_0), where p_1 is the fraction of positive examples and p_0 the fraction of negative examples.

At a given node of a decision tree, , 6 of 10 examples are cats and 4 of 10 are not cats. Which expression calculates the entropy $H(p_1)$ of this group of 10 animals?

- ◉ $-(0.6)log_2(0.6) - (0.4)log_2(0.4)$
- ○ $-(0.6)log_2(0.6) - (1 - 0.4)log_2(1 - 0.4)$
- ○ $(0.6)log_2(0.6) + (1 - 0.4)log_2(1 - 0.4)$
- ○ $(0.6)log_2(0.6) + (0.4)log_2(0.4)$

✓ **Correct**
Correct. The expression is $-(p_1)log_2(p_1) - (p_0)log_2(p_0)$

2.

1 / 1 point

Information gain

$$= H(p_1^{root}) - \left(w^{left} H\left(p_1^{left}\right) + w^{right} H\left(p_1^{right}\right)\right)$$

Recall that information was defined as follows:

$$H(p_1^{root}) - \left(w^{left} H(p_1^{left}) + w^{right} H(p_1^{right})\right)$$

Before a split, the entropy of a group of 5 cats and 5 non-cats is $H(5/10)$. After splitting on a particular feature, a group of 7 animals (4 of which are cats) has an entropy of $H(4/7)$. The other group of 3 animals (1 is a cat) and has an entropy of $H(1/3)$. What is the expression for information gain?

- ○ $H(0.5) - (7 * H(4/7) + 3 * H(1/3))$
- ○ $H(0.5) - \left(\frac{4}{7} * H(4/7) + \frac{4}{7} * H(1/3)\right)$
- ○ $H(0.5) - (H(4/7) + H(1/3))$
- ◉ $H(0.5) - \left(\frac{7}{10} H(4/7) + \frac{3}{10} H(1/3)\right)$

✓ **Correct**
Correct. The general expression is $H(p_1^{root}) - \left(w^{left} H(p_1^{left}) + w^{right} H(p_1^{right})\right)$

3.

1 / 1 point

One hot encoding

Ear-shape	Pointy ears	Floppy ears	Oval ears	Face shape	Whiskers	Cat
Pointy	1	0	0	Round	Present	1
Oval	0	0	1	Not round	Present	1
Oval	0	0	1	Round	Absent	0
Pointy	1	0	0	Not round	Present	0
Oval	0	0	1	Round	Present	1
Pointy	1	0	0	Round	Absent	1
Floppy	0	1	0	Not round	Absent	0
Oval	0	0	1	Round	Absent	1
Floppy	0	1	0	Round	Absent	0
Floppy	0	1	0	Round	Absent	0

To represent 3 possible values for the ear shape, you can define 3 features for ear shape: pointy ears, floppy ears,

oval ears. For an animal whose ears are not pointy, not floppy, but are oval, how can you represent this information as a feature vector?

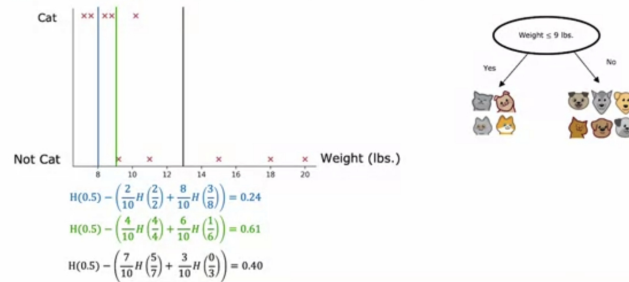- ○ [0, 1, 0]
- ○ [1, 1, 0]
- ○ [1,0,0]
- ◉ [0, 0, 1]

  ✓ **Correct**
  Yes! 0 is used to represent the absence of that feature (not pointy, not floppy), and 1 is used to represent the presence of that feature (oval).

4.   1 / 1 point

Splitting on a continuous variable

For a continuous valued feature (such as weight of the animal), there are 10 animals in the dataset. According to the lecture, what is the recommended way to find the best split for that feature?

- ○ Use gradient descent to find the value of the split threshold that gives the highest information gain.
- ◉ Choose the 9 mid-points between the 10 examples as possible splits, and find the split that gives the highest information gain.
- ○ Use a one-hot encoding to turn the feature into a discrete feature vector of 0's and 1's, then apply the algorithm we had discussed for discrete features.
- ○ Try every value spaced at regular intervals (e.g., 8, 8.5, 9, 9.5, 10, etc.) and find the split that gives the highest information gain.

  ✓ **Correct**
  Correct. This is what is proposed in the lectures.

5.   1 / 1 point

Which of these are commonly used criteria to decide to stop splitting? (Choose two.)

- ☐ When the information gain from additional splits is too large
- ☑ When the tree has reached a maximum depth

  ✓ **Correct**
  Yes!

- ☑ When the number of examples in a node is below a threshold

  ✓ **Correct**
  Yes!

- ☐ When a node is 50% one class and 50% another class (highest possible value of entropy)