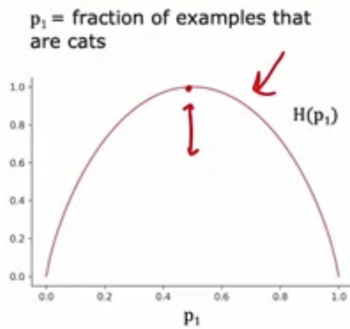


Entropy as a measure of impurity

1 / 1 punto



$$p_0 = 1 - p_1$$

$$\begin{aligned} H(p_1) &= -p_1 \log_2(p_1) - p_0 \log_2(p_0) \\ &= -p_1 \log_2(p_1) - (1 - p_1) \log_2(1 - p_1) \end{aligned}$$

Note: " $0 \log(0)$ " = 0

Recuerde que la entropía se definió en la conferencia como $H(p_1) = -p_1 \log_2(p_1) - p_0 \log_2(p_0)$, donde p_1 es la fracción de ejemplos positivos y p_0 la fracción de ejemplos negativos.

En un nodo dado de un árbol de decisión, 6 de 10 ejemplos son gatos y 4 de 10 no son gatos. ¿Qué expresión calcula la entropía? $H(p_1)$ de este grupo de 10 animales?

- ☐ $(0.6) \log_2(0.6) + (0.4) \log_2(0.4)$
- ☒ $-(0.6) \log_2(0.6) - (0.4) \log_2(0.4)$
- ☐ $(0.6) \log_2(0.6) + (1 - 0.4) \log_2(1 - 0.4)$
- ☐ $-(0.6) \log_2(0.6) - (1 - 0.4) \log_2(1 - 0.4)$

☒ **Correcto**

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

2.

1 / 1 punto

Information gain

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

Recall that information was defined as follows:

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

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?

- ☒ $H(0.5) - (\frac{7}{10} H(4/7) + \frac{3}{10} H(1/3))$
- ☐ $H(0.5) - (\frac{4}{7} * H(4/7) + \frac{4}{7} * H(1/3))$
- ☐ $H(0.5) - (7 * H(4/7) + 3 * H(1/3))$
- ☐ $H(0.5) - (H(4/7) + H(1/3))$











☒ **Correcto**

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

3.

1 / 1 punto

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?

- ☐ [1, 1, 0]
- ☐ [1,0,0]
- ☐ [0, 1, 0]
- ☒ [0, 0, 1]

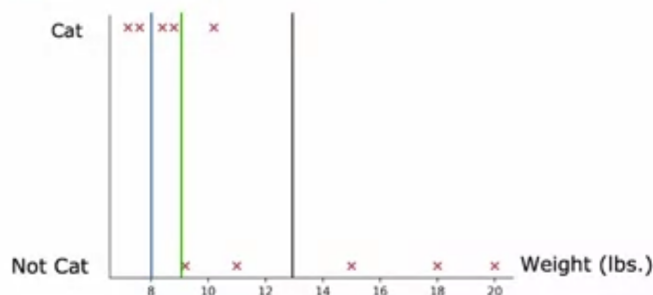
✓ **Correcto**

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 punto

Splitting on a continuous variable



$$H(0.5) - \left(\frac{2}{10} H\left(\frac{2}{2}\right) + \frac{8}{10} H\left(\frac{3}{8}\right) \right) = 0.24$$

$$H(0.5) - \left(\frac{4}{10} H\left(\frac{4}{4}\right) + \frac{6}{10} H\left(\frac{1}{6}\right) \right) = 0.61$$

$$H(0.5) - \left(\frac{7}{10} H\left(\frac{5}{7}\right) + \frac{3}{10} H\left(\frac{0}{3}\right) \right) = 0.40$$



Para una característica de valor continuo (como el peso del animal), hay 10 animales en el conjunto de datos. Según la lección, ¿cuál es la forma recomendada de encontrar la mejor división para esa característica?

- ☒ Elija los 9 puntos medios entre los 10 ejemplos como posibles divisiones y encuentre la división que proporcione la mayor ganancia de información.
- ☐ Use una codificación one-hot para convertir la característica en un vector de características discretas de 0 y 1, luego aplique el algoritmo que habíamos discutido para características discretas.
- ☐ Pruebe cada valor espaciado a intervalos regulares (p. ej., 8, 8,5, 9, 9,5, 10, etc.) y encuentre la división que proporcione la mayor ganancia de información.
- ☐ Utilice el descenso de gradiente para encontrar el valor del umbral dividido que proporciona la mayor ganancia de información.



Correcto

Correcto. Esto es lo que se propone en las conferencias.

5.

1 / 1 punto

¿Cuáles de estos son criterios comúnmente utilizados para decidir dejar de dividir? (Escoge dos.)

- ☐ Cuando un nodo es 50% de una clase y 50% de otra clase (valor de entropía más alto posible)
- ☒ Cuando el árbol ha alcanzado una profundidad máxima



Correcto

¡Sí!

- ☐ Cuando la ganancia de información de divisiones adicionales es demasiado grande
- ☒ Cuando el número de ejemplos en un nodo está por debajo de un umbral



Correcto

¡Sí!

