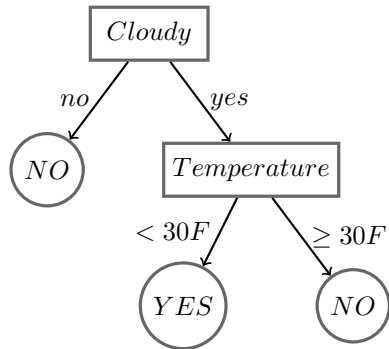


1 Definitions Oh My!

- (a) **Decision Trees:** Popular representation and technique for classification and prediction, in which each node represents an attribute that is tested (or a decision), each branch represents an outcome, and each leaf node represents a resulting label

Example: Will it rainy today?



- (b) **Entropy:** A measurement of the uncertainty in a random variable, hence with high entropy we are less confident about the outcome

$$H(Y) = - \sum_y P(Y = y) \log_2 P(Y = y)$$

- (c) **Conditional-Entropy:** The expected value of the entropy of Y given X , over all values of X . This lets us quantify the entropy of Y given that we know X .

$$H(Y | X) = \sum_x P(X = x) H(Y | X = x)$$

- (d) **Mutual Information:** Measurement of how much uncertainty about variable Y decreases if we know and have observed X .

$$I(Y; X) = H(Y) - H(Y | X = x)$$

2 Decision Trees

2.1 Mutual Information and Entropy Practice

1. Calculate the entropy of tossing a fair coin.

2. Calculate the entropy of tossing a coin that lands only on tails. note : $0 \cdot \log_2 0 = 0$.

3. Calculate the entropy of a fair dice roll.

4. When is the mutual information $I(Y;X) = 0$?

2.2 Decision Tree *Runthrough*

Fun fact: Pat loves to go on runs! However, sometimes things come up which deter him from running outside. The following attributes have affected his running in the past:

- How the sky looks the day of his run (sunny, overcast, rainy)
- The temperature outside (hot, mild, cool)
- The humidity 30 minutes before he plans on running (high, normal)

As a member of the 07-280 running club, you have gathered the following data and decided that you want to make a decision tree to predict whether Pat will go on a run.

Outlook (X_1)	Temperature (X_2)	Humidity (X_3)	Go on run? (Y)
sunny	hot	high	no
overcast	hot	high	yes
rainy	mild	high	yes
rainy	cool	normal	yes
sunny	mild	high	no
sunny	mild	normal	yes
rainy	mild	normal	yes
overcast	hot	normal	yes

To start, let's first determine the attribute for the root node of the decision tree. To do this, we need to determine the mutual information for each attribute.

1. Calculate $H(Y)$, the entropy of Pat going on a run.

2. Calculate $I(Y; X_1)$, the mutual information when splitting on the outlook.

3. Calculate $I(Y; X_2)$, the mutual information when splitting on temperature.

4. Calculate $I(Y; X_3)$, the mutual information when splitting on humidity.

5. Which attribute do we split on?

6. What does the decision tree look like so far?

Now that we've split on the first attribute, the remaining data points which still cannot be classified are shown in the table below.

Outlook (X_1)	Temperature (X_2)	Humidity (X_3)	Go on run? (Y)
sunny	hot	high	no
sunny	mild	high	no
sunny	mild	normal	yes

From now on, we will be strictly considering the branch of the decision tree where the outlook is sunny.

7. Calculate $H(Y)$, the entropy of going on a run given that it is sunny.

8. Calculate $I(Y; X_2)$.

9. Calculate $I(Y; X_3)$.

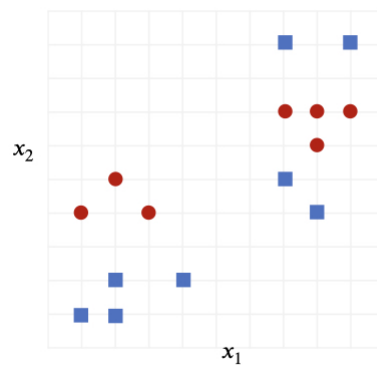
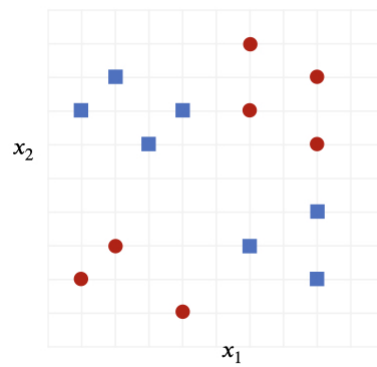
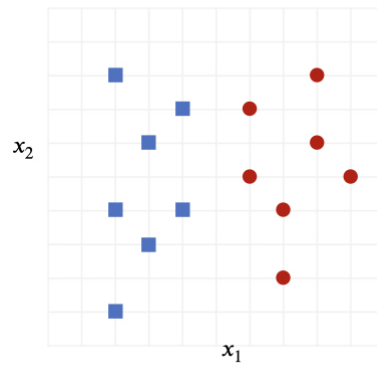
10. Which attribute do we split on now?

11. Draw the final decision tree.

3 Decision Boundaries

3.1 Decision Trees

What could the decision boundaries look like for the following datasets, assuming zero error?



4 Decision Trees

Explore different values for `max_depth` by running through the DT Jupiter Notebook (see recitation materials).

4.1 Discussion Questions

1. What is the best value of `max_depth` for this dataset and why?

2. Why do you think the error is large with very small `max_depth`?

3. How is the value of `max_depth` related to overfitting?

4. Why can we not display the decision boundaries like we saw in lecture?

5. For which values of `max_depth` would the decision boundaries be the most complex?