

# SAAS CX Homework 10 Solutions

JONATHAN PAN AND MICHAEL WANG

Spring 2022

## §1 Decision Trees

**Problem 1.1.** We can use entropy to gauge how good a split is. The lower the weighted entropy after a split, the better.

**Problem 1.2.** We first find the best split for every feature by trying to split on every data point value and calculating the weighted entropy. Then we pick the best feature and its corresponding split and repeat.

## §2 Ensemble Learning

**Problem 2.1.** This will decrease the chances of overfitting. If we didn't do this, all the trees may end up looking very similar to each other.

## §3 k-Nearest Neighbors

**Problem 3.1.**

- $k = 1$ : We only ask our closest neighbor. We will always get a training error of 0. Low bias, high variance.
- $k = N$ : We ask every data point. All points will be assigned to a single class. High bias, low variance.