Vanessa Le

Professor Brooks

CS 386-02

13 November 2024

Assignment 5: Working with Uncertainty

PROBLEM 1: Decision Trees in scikit-learn.

n_estimators	criterion	average
10	gini	0.9526005278683435
10	entropy	0.9543238627542306
25	gini	0.9490607048594939
25	entropy	0.9631268436578171
50	gini	0.9508150908244062
50	entropy	0.9508461419034312

GridSearchCV:

[{'model': 'Random Forest', 'cv_results': mean_fit_time std_fit_time ... mean_train_score std_train_score

0 0.057387 0.027228 ... 0.970563 0.005660

1 0.067599 0.024306 ... 0.980226 0.005208

2 0.050567 0.005781 ... 0.981986 0.001647

3 0.054298 0.004768 ... 0.982866 0.002143

[4 rows x 21 columns]}, {'model': 'Hist Gradient Boosting', 'cv_results': mean_fit_time std_fit_time ... mean_train_score std_train_score

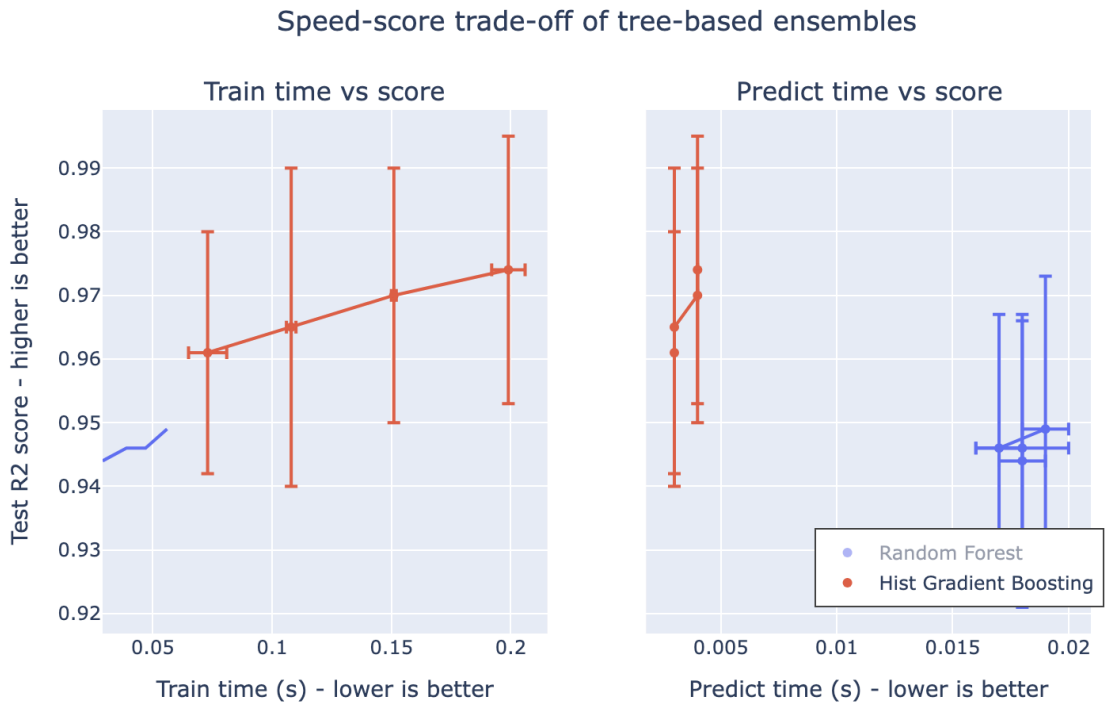
0 0.076161 0.017742 ... 0.989456 0.002557

1 0.133158 0.027054 ... 1.000000 0.000000

2 0.181014 0.027376 ... 1.000000 0.000000

3 0.224724 0.013797 ... 1.000000 0.000000

Plot:



PART 4: UTILITY

1) It knows that over **rocky terrain it can go 2 km/h**. Over **sandy terrain it can go 3 km/h**, and over **smooth terrain it can go 5 km/h**. There are three routes it might choose from. Unfortunately, our terrain data for the three routes is incomplete, so we only have estimates.

- Route 1 is 5 km long. There is a 20% chance it is sandy, 30% chance it is smooth, and a 50% chance it is rocky.
- Route 2 is 7 km long. There is a 40% chance it is sandy, a 20% chance it is smooth, and a 40 % chance it is rocky.
- Route 3 is 6 km long. There is a 50% chance it is sandy, a 40% chance it is smooth, and a 10% chance it is rocky.

Which route should we pick?

Rocky 2 km/hr = 0.033 km/min

1hr/2km * 60 min/1hr = 60min/2km = **30min/km**

Sandy 3km/hr = 20min/km

1hr/3km * 60min/1hr = 60min/3km = **20min/km**

Smooth 5km/1hr = 12min/km

1hr/2km * 60min/1hr = 60min/2km = **30min/km**

Route 1: 5 km long. 20% chance it is sandy, 30% chance it is smooth, and a 50% chance it is rocky.

(0.20 * 20min/km) + (0.30 * 12min/km) + (0.50 * 30min/km) = 4 + 3.6 + 15 = 22.6min/km * 5km) = **113 min**

Route 2: 7 km long. 40% chance it is sandy, a 20% chance it is smooth, and a 40 % chance it is rocky.

(0.40 * 20min/km) + (0.20 * 12min/km) + (0.40 * 30min/km) = 8 + 2.4 + 12 = 22.4min/km * 7km) = **156.8 min**

Route 3: 6 km long. 50% chance it is sandy, a 40% chance it is smooth, and a 10% chance it is rocky.

(0.50 * 20min/km) + (0.40 * 12min/km) + (0.10 * 30min/km) = 10 + 4.8 + 17.8 = 17.8min/km * 6km) = **106.8 min**

We should pick Route 3 because it is the fastest route.

2) Route 1 contains a crater. If the wall of the crater is **intact**, we can take a shortcut through the crater, which will **save 20 minutes**. If the wall has been **damaged**, we will need to go around, which will **add 15 minutes** to our journey. There is a **30% chance that the wall is damaged**. Route 3 contains a bridge. If that **bridge is damaged**, we will need to repair it, which will **add 40 minutes** to our time. There is a **60% chance that the bridge is damaged**.
Update your estimates for the travel time for each route. Now which route seems best?

Route 1: Contains a crater

0.70 * 20min = **Saving 14min**

0.30 * 15min = **Adding 4.5min**

Route 3: Contains a bridge

0.60 * 40 = **Adding 24min**

Updated estimates:

Route 1: 113 - 14 + 4.5 = **103.5 min**

Route 3: 106.8 + 24 = **130.8 min**

We should choose Route 1 because it is now the fastest route.

3) Now we have an additional piece of potential information. There is an orbiting satellite that can tell us whether Route 2 is rocky or not. If not, that would be great news, and would make it much more appealing! The only problem is that the satellite is not yet in position. How long should we wait for the satellite?

- First: If the satellite said that Route 2 was not rocky, how long would we expect it to take?
    - Not rocky = Route 2 is 66.6% chance it's sandy, 33.3% chance it's smooth
    - Sandy = 0.66 * 20min/km = 13.2min/km * 7km = **92.4 min**
    - Smooth = 0.33 * 12min/km = 3.96min/km * 7km = **27.72 min**
    - 92.4 + 27.72 = **120.12 min**
    - **The rover will take 120.12 minutes if Route 2 wasn't rocky.**
- Second: What's the probability that the satellite will tell us this?
    - **40% chance that it's sandy, 20% chance that it's smooth, and 40% chance that it's rocky. 40% + 20% = 60%, so there's a 60% chance the satellite will tell us that it's not rocky.**
- Third: If the satellite tells us route 2 is in fact rocky, what do we do? How long will that take?
    - **There is a 40% chance it's rocky, so we would choose Route 1 since it takes 103.5 min.**
- Last: given all of this, how long should we wait for the satellite?

- Route 1: 113 - 14 + 4.5 **= 103.5 min**
- Route 2: 92.4 + 27.72 **= 120.12 min**
- Route 3: 106.8 + 24 **= 130.8 min**
- **I do not think that we should wait for the satellite at all because Route 1 is the better option, thus we should take that route instead of waiting.**