15.095: Machine Learning Under a Modern Optimization Lens **Fall 2021**

Homework 5: Due November 12

Hand in: **pdf** upload to Canvas. Please append any Julia code **at the end** of the whole pdf.

Question 1: Extensions of the Greedy CART Algorithm and their Use for Prescriptions (40 Points)

In this problem, you will greedily fit classification trees (using CART) to the data distribution given in Figure 5.1 (whereby $n = 37$ and $p = 2$); then, you will use the fitted tress for prescriptions.

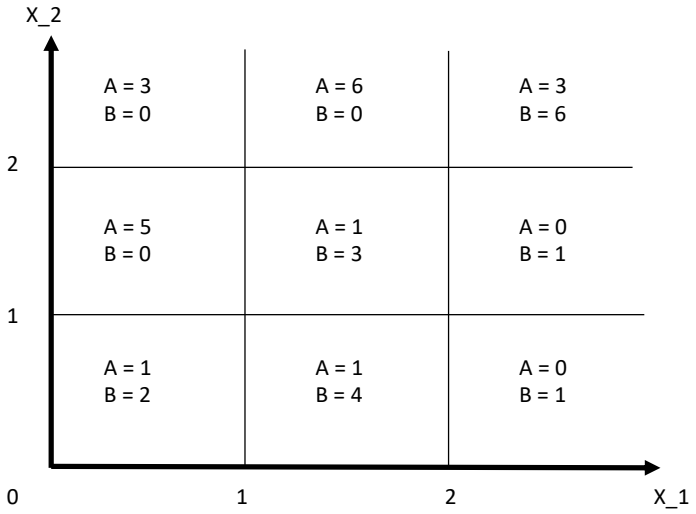

Figure 5.1: Distribution of data of two classes (A and B) in each of the regions.

For interpretability, your tree will need to satisfy the following requirements (some of which we studied in Problem 3 in the Midterm):

1. **Maximum depth:** Restrict the tree depth to 2.
2. **Integer splits:** Splits can only be performed at integer points.
3. **Variable ordering requirement:** In the Midterm, we asked you to incorporate in the OCT formulation the variable ordering requirement outlined below:
   "You are given sets of features $\mathcal{F}_1, \ldots, \mathcal{F}_\kappa,\ 1 < \kappa < D$, such that $\mathcal{F}_k \cap \mathcal{F}_l = \emptyset$ and $\cup_{l=1}^{\kappa} \mathcal{F}_l = \{1, \ldots, p\}$. All features from set $\mathcal{F}_k$ are not allowed to be selected at split nodes unless, for all $l < k$, there exists some feature $i \in F_l$ which has already been selected in a preceeding split node in the same path."
   In our simple, 2-dimensional example, let $\mathcal{F}_1 = \{1\}, \mathcal{F}_2 = \{2\}$. You need to satisfy this requirement in building your CART tree.

(a) (3 Points) Briefly describe how the resulting classification tree will look like based on the aforementioned requirements.

(b) (5 Points) Greedily build a tree using the misclassification error at each split.

(c) (9 Points) Greedily build a tree using the gini impurity at each split.
Recall that the formula to compute the Gini impurity for a node t is $I_t = 1 - \sum_i^J p_j^2$, where p_j is the probability of finding label j in node t. Provide an interpretation of the Gini impurity. *(Hint: Think of where the aforementioned formula comes from.)* Explain why the Gini impurity is considered a more appropriate metric for greedily fitting classification trees.

(d) (3 Points) Compare the two resulting trees. Are they the same? If not, provide a brief discussion about why different metrics can lead to different trees.

(e) (20 Points) Finally, you will use your trained CART trees to make prescriptions. Consider the binary variable $z \in \{0, 1\}$ indicating whether subject (i.e., data point) i received a specific treatment or not. Depending on a subject's class, the cost of giving them the treatment or not is as follows:

c(z;y)	*y=A*	*y=B*
z=0	0	3
z=1	10	1

Consider a situation in which 9 new subjects, one from each region in Figure 5.1 (recall that each subject's region is decided based on their features), are candidates to receive the treatment. However, you only have 3 treatments available.

Develop an integer linear optimization formulation to decide who will get the treatment. Solve (by hand) this formulation using each of the trained trees from parts (b) and (c). Who gets the treatment under each tree?

Question 2: From Predictions to Prescriptions: the Newsvendor Problem (60 Points)

In this problem, you will be working with data adapted from the Kaggle competition `https://www.kaggle.com/c/favorita-grocery-sales-forecasting`. The data came from a large South American grocery chain, and the prediction task was to forecast sales for over 200,000 products sold in 55 grocery stores in Ecuador.

Using this data set, we will consider the single-period Newsvendor problem with multiple items. The objective is to determine how much of each product 1-100 to stock in grocery store #1 for the next day, August 15, 2017. For this problem, we will make the following assumptions:

- We restock the grocery store daily.
- We can only stock integer number of products.
- The grocery store has a maximum capacity of Q units that we can stock.
- We can stock at most $\frac{1}{20}Q$ units of any particular product.
- If a product is perishable and we don't sell our entire quantity, then we lose money equal to the cost of product times the number of extra units.

- If a product is non-perishable and we don't sell our entire quantity, then we can save the extra units to sell another day and do not lose money. (For revenue purposes, you can assume we can sell the extra units back to the supplier at the end of the day at full cost).

We provide you with the following data files to use for the exercises:

- `items.csv`: The subset of 100 grocery store products that we will consider for this problem, with data fields including each product's ID, category, subcategory, base price, cost, and whether the product is perishable (1) or not (0).
- `sales.csv`: Historical sales data for these 100 products at grocery store #1 from January 2, 2017 to August 15, 2017. Data fields include date, product ID, number of products sold, whether the product was on promotion (1) or not (0), and whether the product was on display (1) or not (0).
- `sideinformation.csv`: Extra information from the past year which may help to predict sales data, including daily oil prices and national holidays of Ecuador. Some values for the daily oil prices that were missing have been imputed for you.

(a) (10 Points) Formulate an optimization problem to maximize profit for grocery store #1 for the next day (August 15, 2017) given the provided historical data and assumptions. Using the following table as a list of variables, write down the optimization problem.

Variable	Definition	Type
s_i	the number of units for non-perishable product i	decision variable
t_j	the number of units for perishable product j	decision variable
p_i	the price of non-perishable product i	constant
q_j	the price of perishable product j	constant
b_i	the cost of non-perishable product i	constant
c_j	the cost of perishable product j	constant
$\hat{d}_i$	the demand of non-perishable product i	uncertain parameters
$\hat{e}_j$	the demand of perishable product j	uncertain parameters

(b) In the following part, you will solve the optimization problem in Julia using three approaches:

(i) (15 Points) Baseline approach: suppose we would like to use demand from the previous day as estimated demand for each product. Describe how you would process the data to construct $\hat{d}_i$ and $\hat{e_j}$ in the above formulation? Implement this optimization model and report profit.

(ii) (15 Points) Oracle approach: suppose we have an "oracle" which can perfectly predict the demand. Describe how you would process the data to construct $\hat{d}_i$ and $\hat{e}_j$ in the above formulation? Implement this optimization model and report profit.

(iii) (15 Points) KNN approach: for each product, we estimate demand as the average of the K=5 nearest neighbors among the past 100 days of data, you may want to use side information data set to do so. Describe how you would process the data to construct $\hat{d}_i$ and $\hat{e_j}$ in the above formulation? Implement this optimization model and report profit.

(iv) (5 Points) Instead of KNN, we could also use other ML models to estimate demand. In this case, briefly comment on how the profit would be in comparison to the above three approaches.