VG441 Homework 1

Submit your problem set solutions as a PDF file with python code in the Appendix.

Problem 1

Prove that the derivative of $\theta^T X^T X \theta$ with respect to θ is $2X^T X\theta$.

Problem 2

Age	Home Owner	Car Owner	Having Kids	Salary
40	Yes	Yes	Yes	10000
20	No	No	No	500
50	Yes	No	Yes	8000
30	Yes	No	No	5000

Tasks

- Run GBM on paper for two iterations (i.e., stopping at F2 and PR2). No more than 4 leaves. Use learning rate $\gamma = 0.1$. Features can be re-used in DT.
- Run XGBoost on paper for two iterations (i.e., stopping at F2 and PR2). No more than 4 leaves. Use regularizer $\lambda = 1$ and pruning $\gamma = 0$ and learning rate $\mu = 0.1$.

Problem 3 (Open-Ended)

Dataset

California housing price data in the 1990-2000. 1–9 are the features and 10 is the target.

1. longitude: A measure of how far west a house is; a higher value is farther west
2. latitude: A measure of how far north a house is; a higher value is farther north
3. housingMedianAge: Median age of a house within a block; a lower number is a newer building
4. totalRooms: Total number of rooms within a block
5. totalBedrooms: Total number of bedrooms within a block
6. population: Total number of people residing within a block
7. households: Total number of households, a group of people residing within a home unit, for a block
8. medianIncome: Median income for households within a block of houses (measured in tens of thousands of US Dollars)
9. oceanProximity: Location of the house w.r.t ocean/sea
10. medianHouseValue: Median house value for households within a block (measured in US Dollars)

Tasks

- Build a Linear Regression Model using 80% training set and 20% testing set. Interpret your results as much as you can.
- Build a GBM using 80% training set and 20% testing set. Interpret your results as much as you can.
- Build a XGBoost Model using 80% training set and 20% testing set. Interpret your results as much as you can.