CSE547: Machine Learning for Big Data
Homework 4

Academic Integrity We take academic integrity extremely seriously. We strongly encourage students to form study groups. Students may discuss and work on homework problems in groups. However, each student must write down the solutions and the code independently. In addition, each student should write down the set of people whom they interacted with.

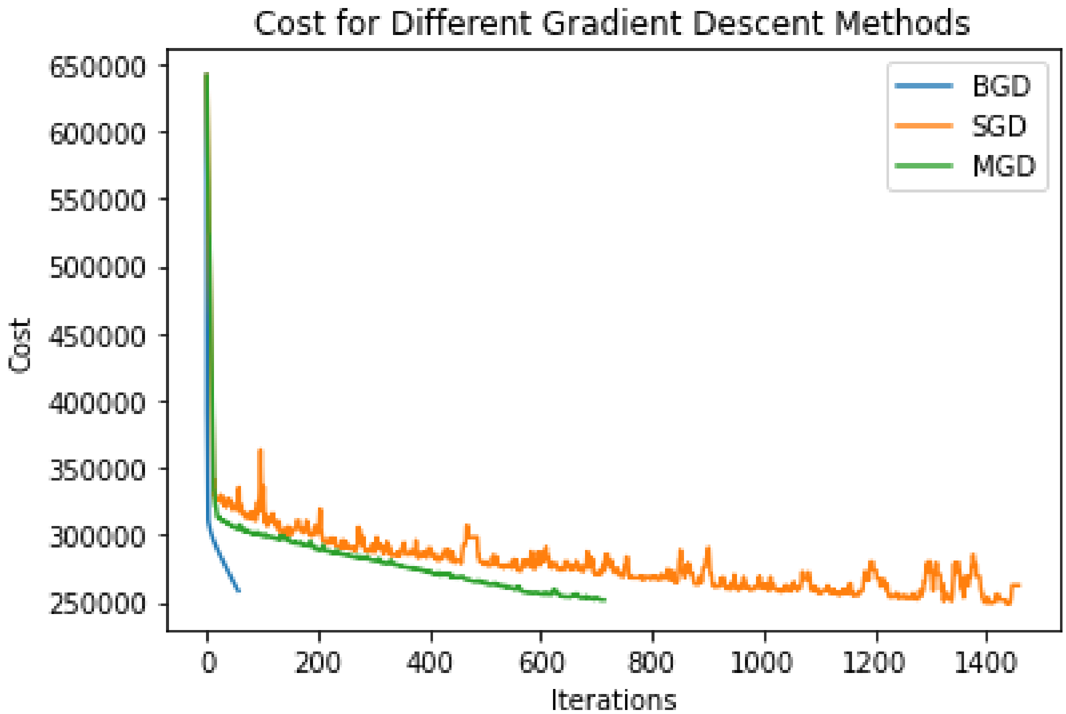
Discussion Group (People with whom you discussed ideas used in your answers):

On-line or hardcopy documents used as part of your answers:

I acknowledge and accept the Academic Integrity clause.

*(Dongyang Wang)*____

Answer to Question 1



The plot is shown below.
Batch Gradient Descent takes 0.3035109043121338 seconds.
Stochastic Gradient Descent takes 0.7520344257354736 seconds.
Stochastic Gradient Descent takes 0.4414517879486084 seconds.
The convergence times meet the expectation because it is related to how fast the algorithms are. As discussed in class, the stochastic methods are faster but require more iterations. On the other hand, the batch method takes longer while requiring fewer iterations. The combination is somethwere in between. The stochatic method is faster thanks to its sampling of the data and random update.

Answer to Question 2(a)

smallgraph Number of Nodes: 4274 smallgraph Number of Edges: 87148

Answer to Question 2(b)

Answer to Question 2(c)

Answer to Question 3(a)

The complexity is $O(\log 1/\delta)$ since that's the number of hash functions.

Answer to Question 3(b)

By defniition, $\hat{F}[i] = min_j c_{j,h_j(i)}$, it will appear as least as many times as the number of times i has appeared in S, which is the definition of F[i]. Therefore, $\hat{F}[i] \geq F[i]$

Answer to Question 3(c)

With t being the length of stream, $E(c_{j,h_j(i)})$ is at most the true the number of times i has appeared in S plus the probablitiy that j shows up multiplied the number of times. The probability is $1/range(h_j) = \epsilon/e$. Therefore, $E(c_{j,h_j(i)}) \leq F[i] + \epsilon/e * t$.

Answer to Question 3(d)

$P(\hat{F}[i] \leq F[i]+\epsilon t) = 1-\Pi_{j=1}^{log(1/\delta)} P(\hat{F}[i] \geq F[i]+\epsilon t) = 1-\Pi_{j=1}^{log(1/\delta)} P(min_j c_{j,h_j(i)} \geq F[i]+\epsilon t) = 1 - \Pi_{j=1}^{log(1/\delta)} P(c_{j,h_j(i)} \geq F[i]+\epsilon t)$.

By Markov inequality, $P(c_{j,h_j(i)} \geq F[i]+\epsilon t) \leq \frac{E(c_{j,h_j(i)} - F[i])}{\epsilon t}$. Along with result from c, we have $P(c_{j,h_j(i) \geq F[i]+\epsilon t} \leq \frac{1}{e}$.

Therefore, $P(\hat{F}[i] \leq F[i]+\epsilon t) = 1 - \Pi_{j=1}^{log(1/\delta)} P(c_{j,h_j(i)} \geq F[i]+\epsilon t) = 1 - (\frac{1}{e})^{log(1/\delta)} = 1 - \delta$

Answer to Question 3(e)