

# Solution: Homework Assignment 3

## COGS 118A: Introduction to Machine Learning I

1. Convex: A, C

4.1: B

4.2:

A. The inverse is on the square matrix that has size  $d \times d$ , where  $d$  is the dimension of each data point. Therefore, the matrix inverse would be expensive when the dimension is large, not when the number of data points is large.

B. Gradient descent could be slower since it does computation in serial; however, matrix inverse could be slow as well. It's hard to compare in general.

C. Yes with an appropriate learning rate and for a convex problem.

D. Yes, but it's not that hard for a convex problem, as you did for problem 3.

4.3: A, D

4.4:

A. Local optimal will not give you the minimal as the global optimal does; however, it might be better as it might lead to higher testing accuracy.

B. Yes

C. Yes for a convex problem with an appropriate learning rate. No for a non-convex problem in general, but might it find the global optimal by chance with some particular initial condition and learning rate.

D. Least square solution finds the global optimal for the least square problem, but not for any other more general optimization problems.