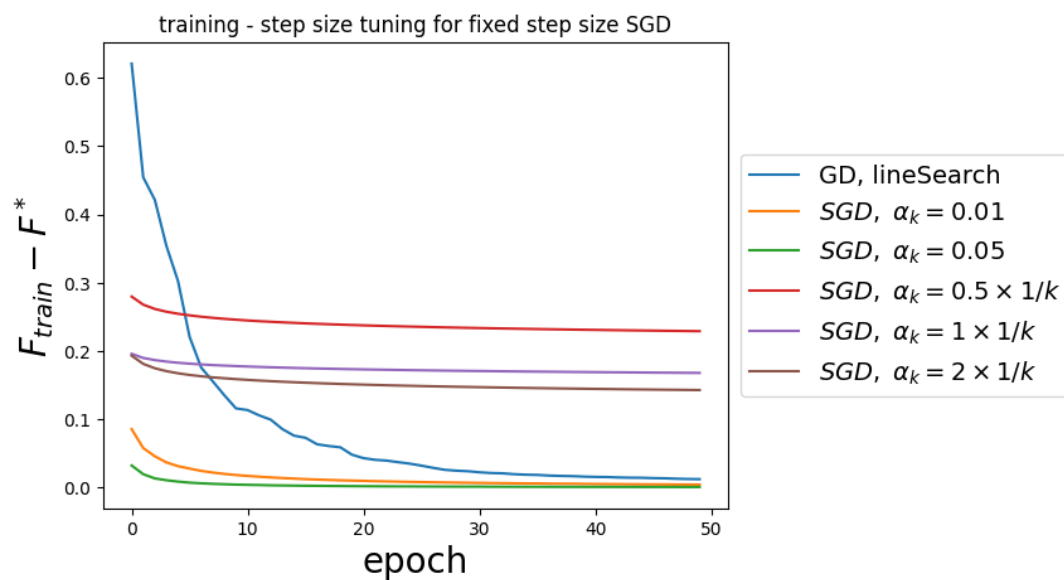


1. Data: “australian.mat”

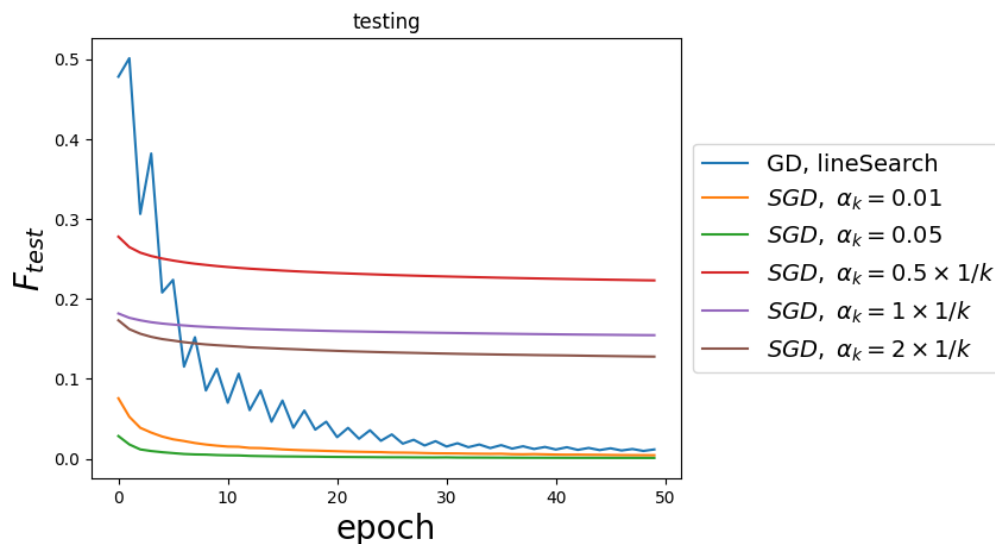
1.a) Loss Function: Linear Least Squares

From the plots below, I could find that all of the methods that are illustrated in the plots are converging. To be specific, Gradient Descent with Backtracking Line Search and Stochastic Gradient Descent with fixed step size converge with the rate of linear convergence and high accuracy. Otherwise, for Stochastic Gradient Descent with diminishing step size does converge but in sublinear, not in a linear convergence rate.

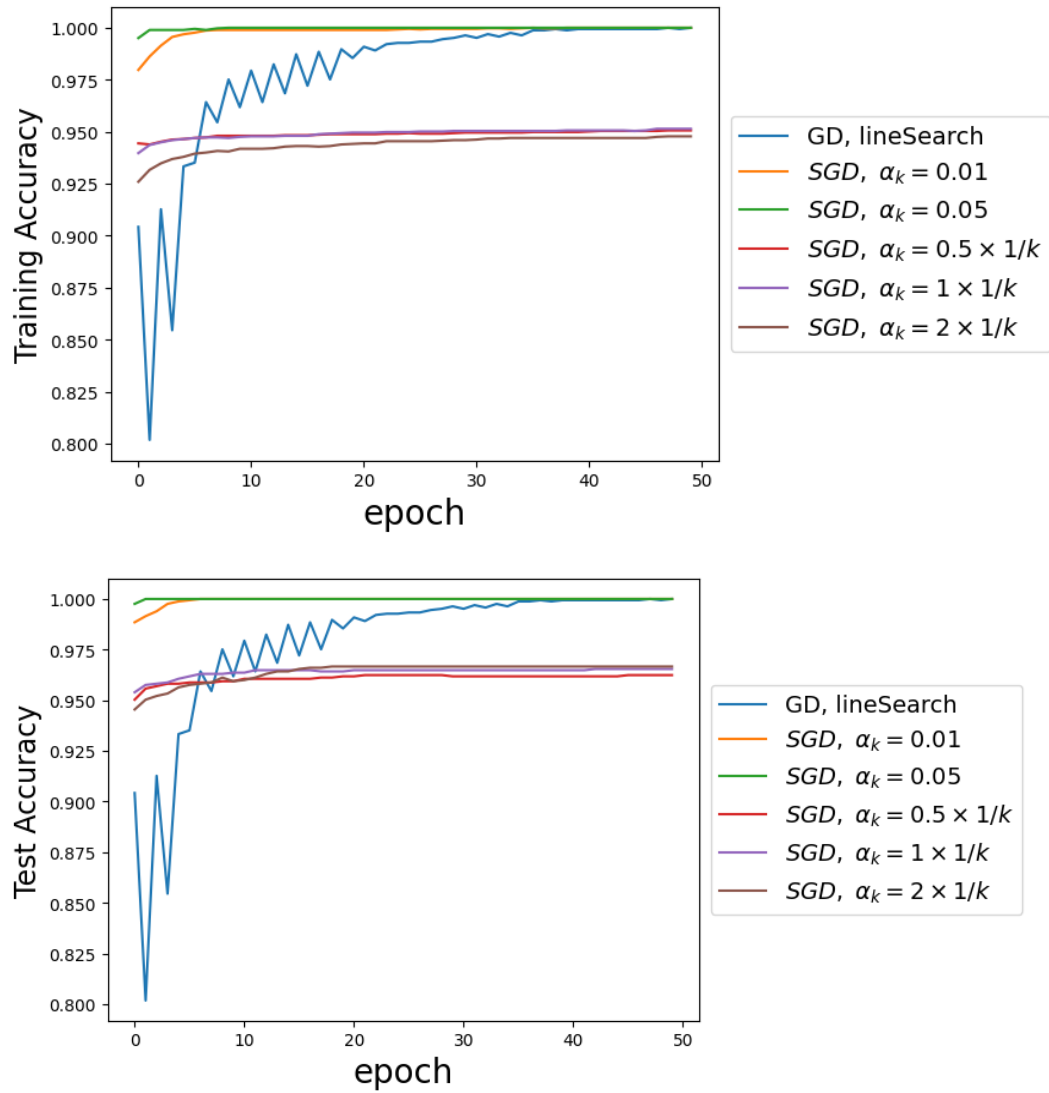
i. Optimality Gap



ii. Testing Loss

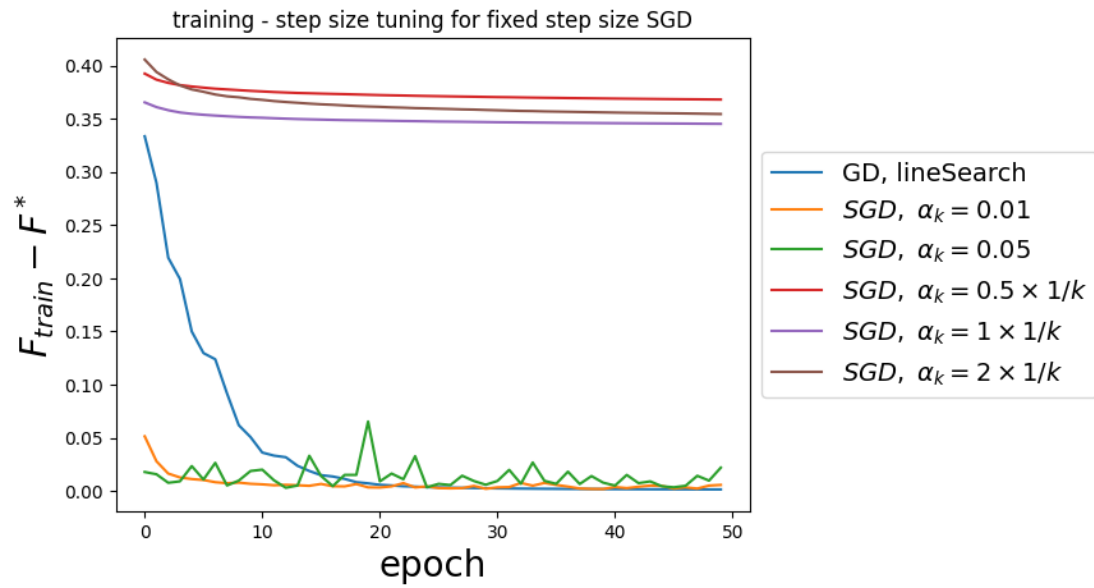


iii. Training and Testing Accuracy

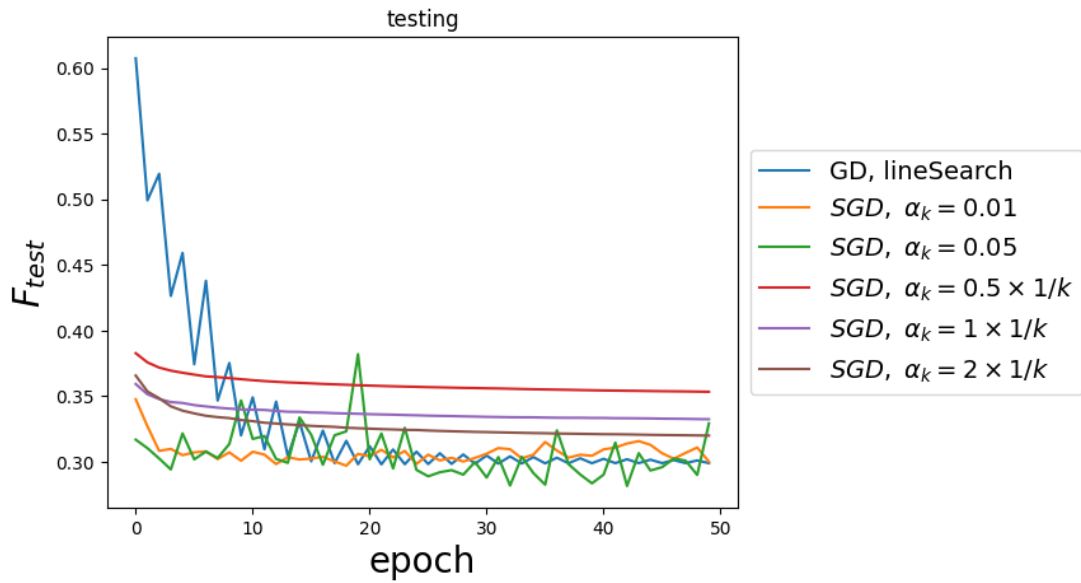


1.b) Loss Function: Logistic Regression

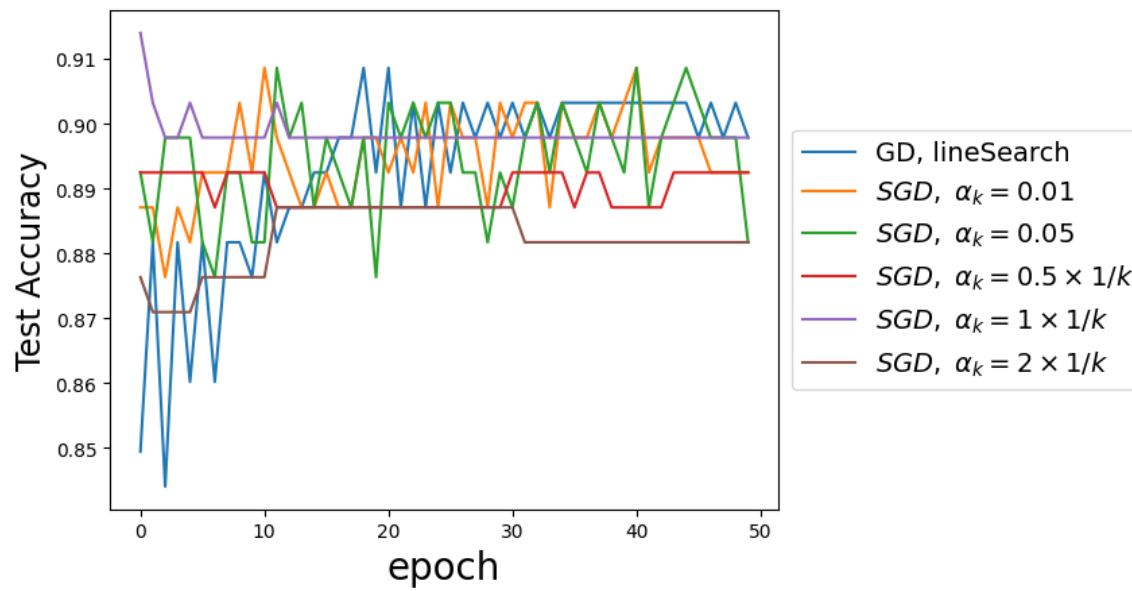
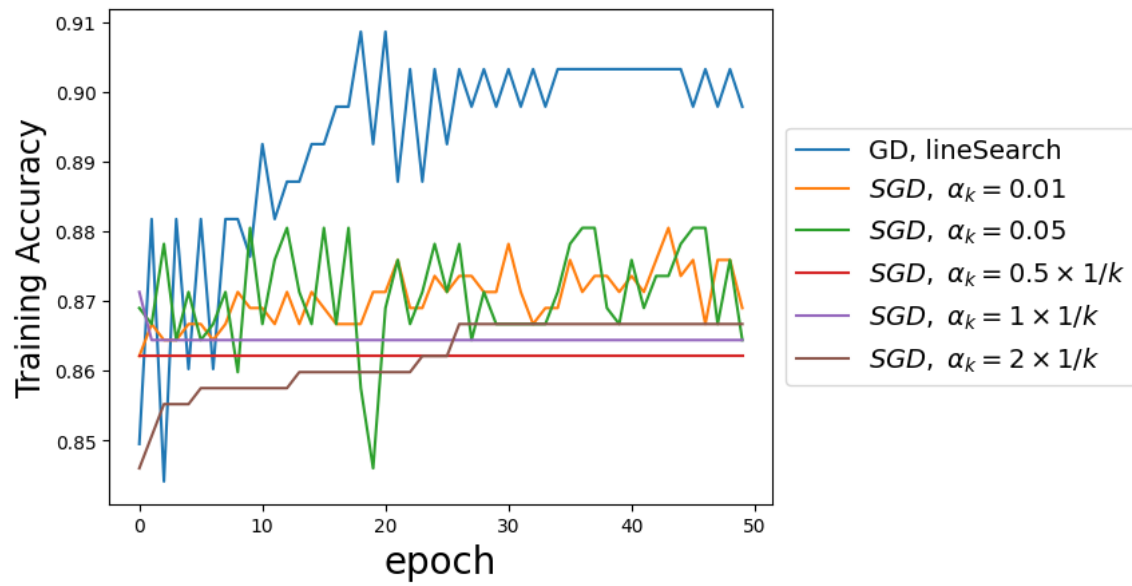
i. Optimality Gap



ii. Testing Loss



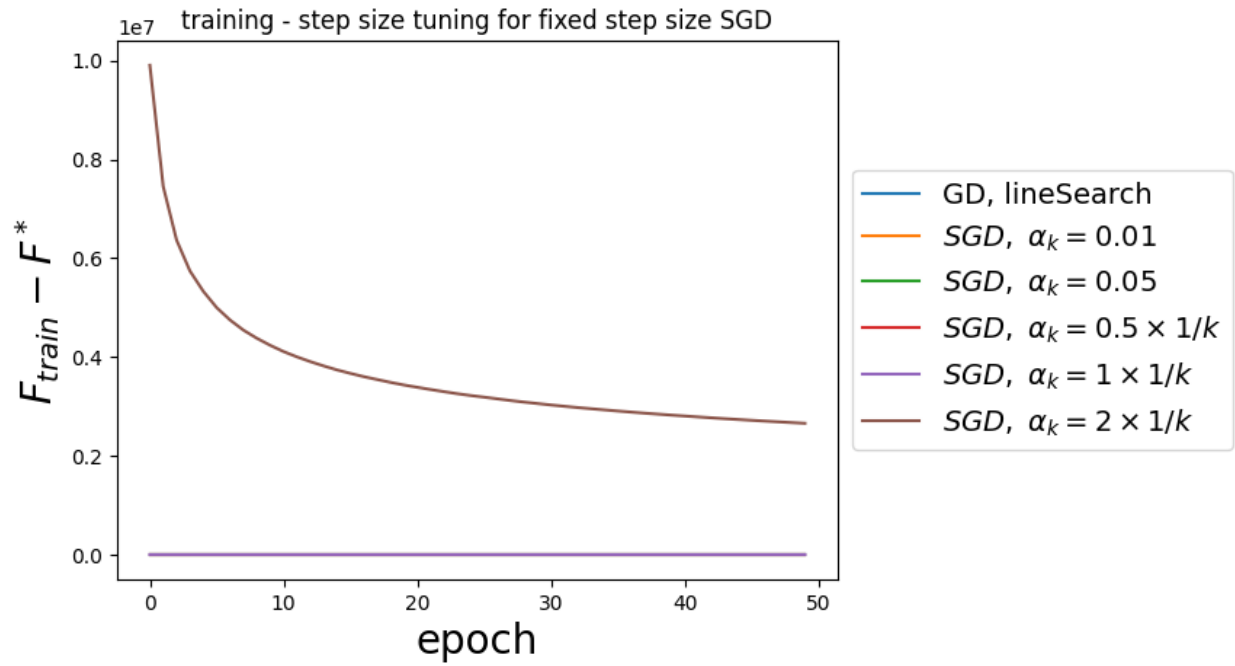
iii. Training and Testing Accuracy



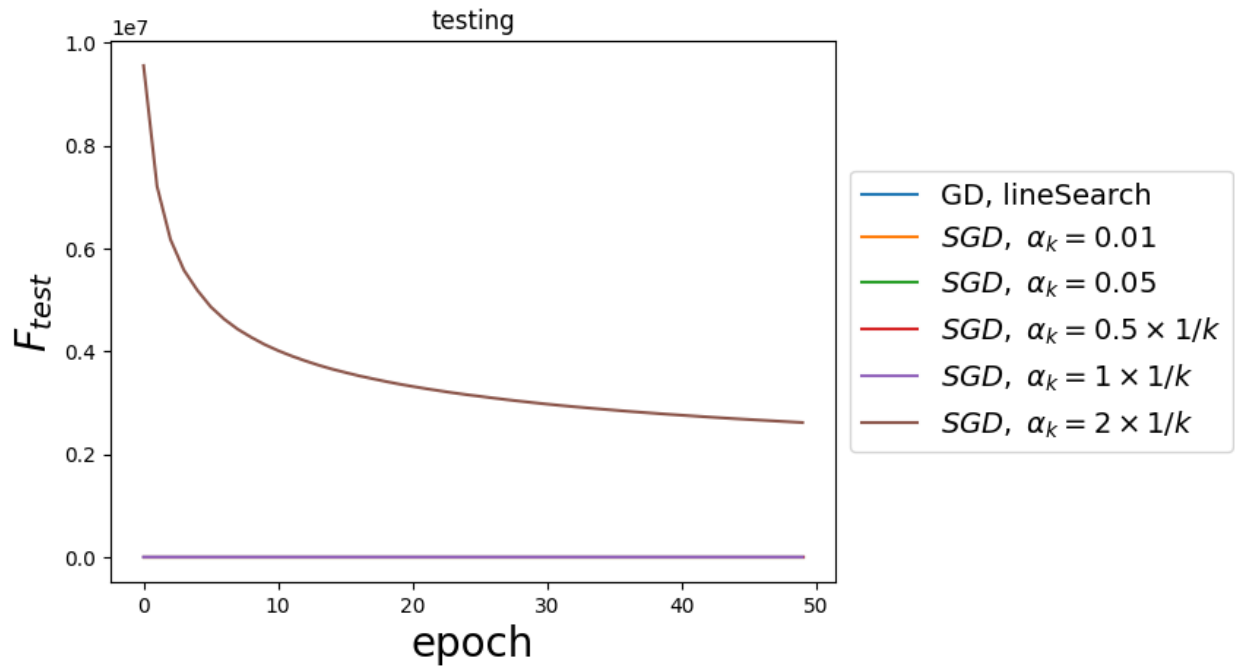
2. Data: “mushroom.mat”

1.a) Loss Function: Linear Least Squares

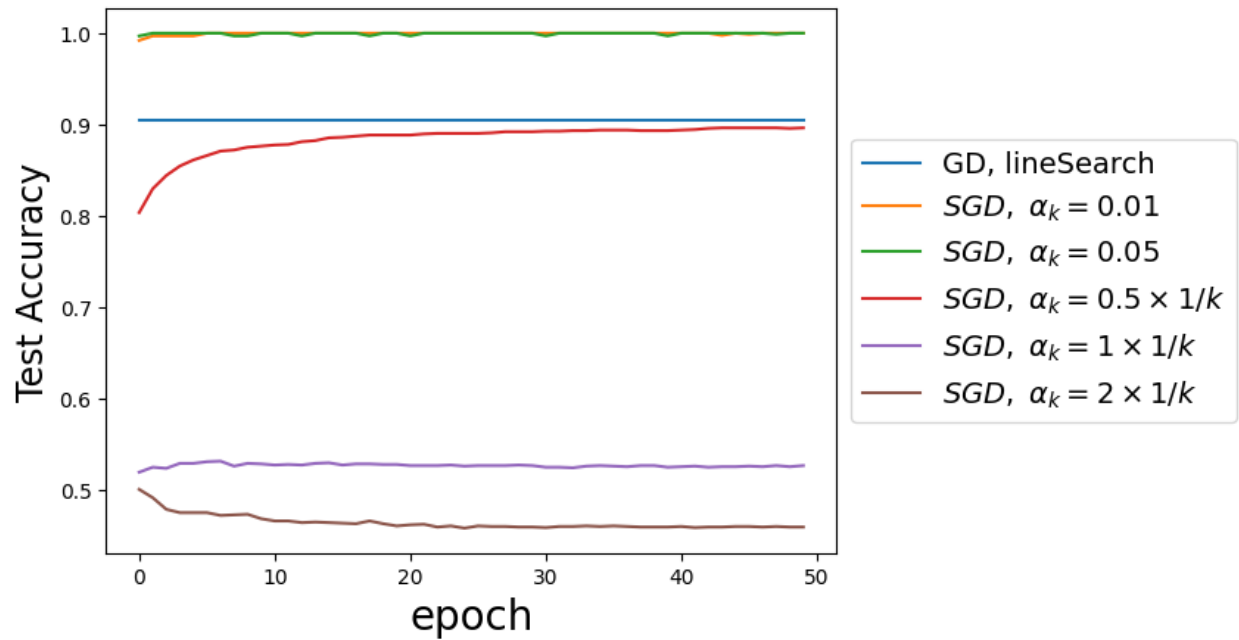
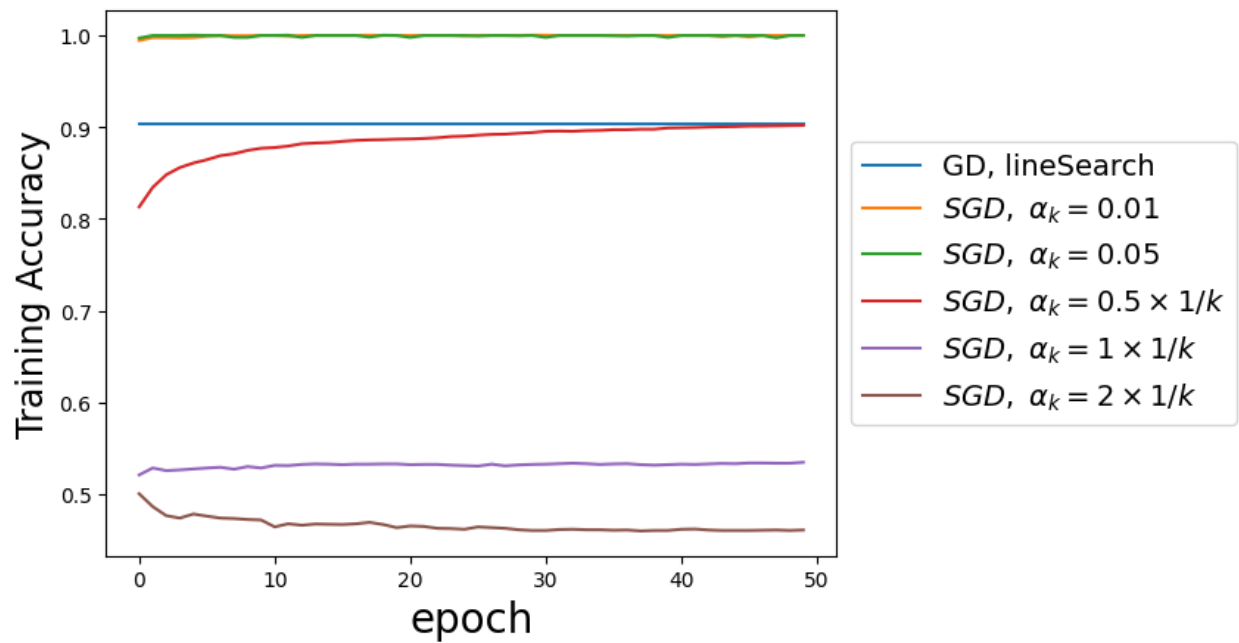
- i. Optimality Gap



ii. Testing Loss

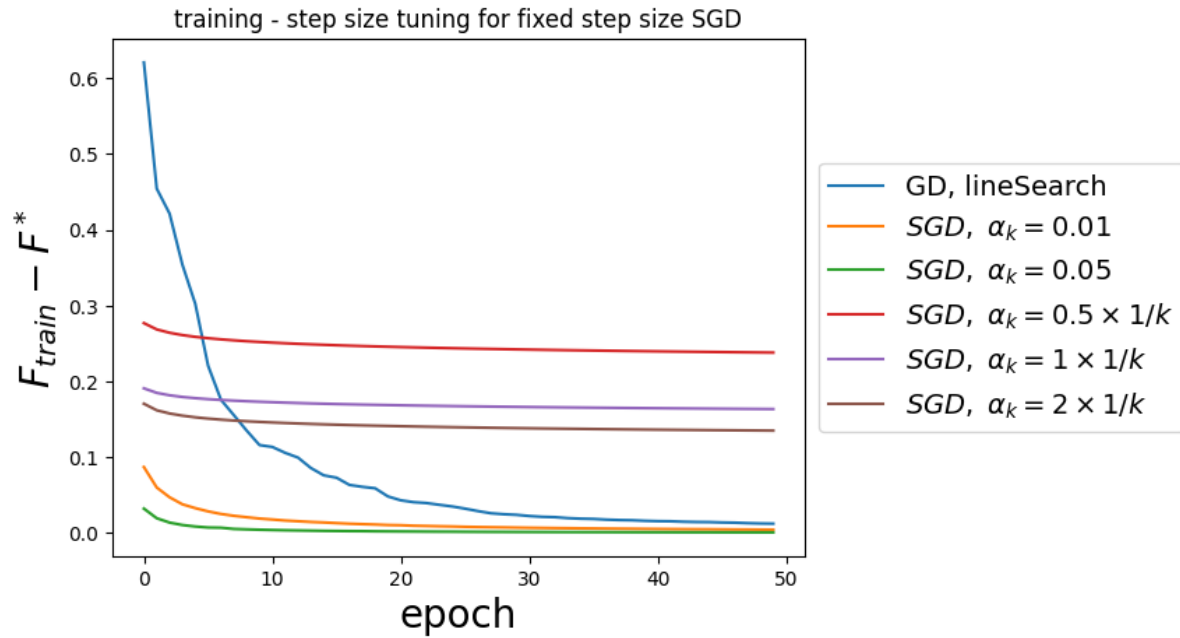


iii. Training and Testing Accuracy

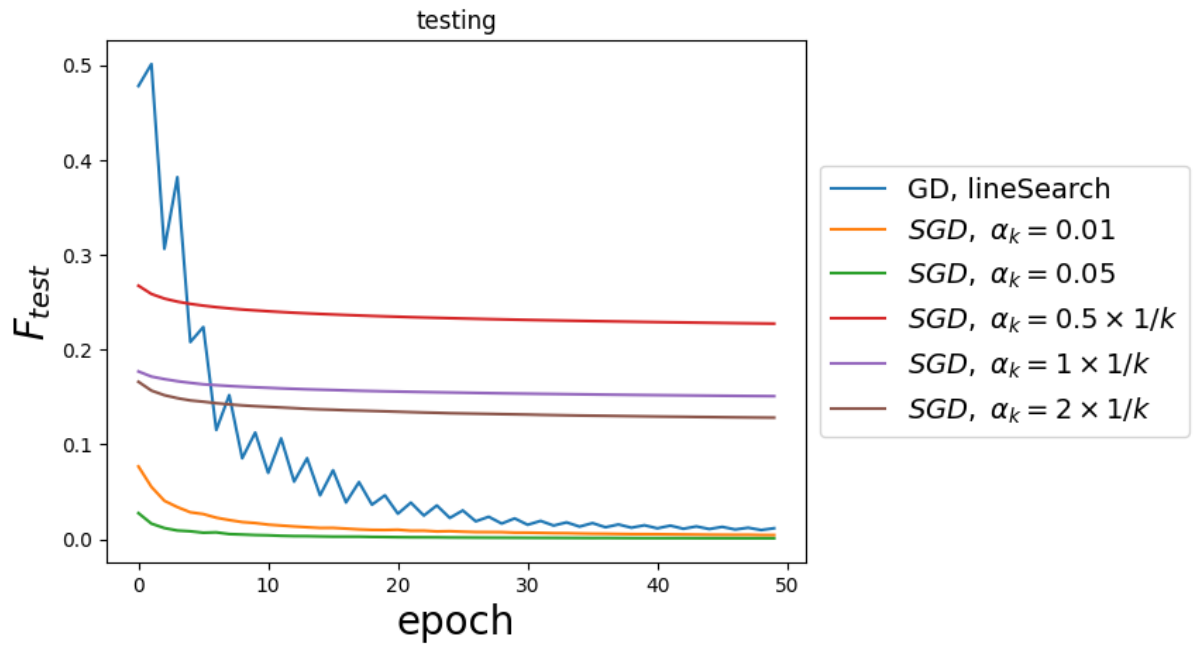


1.b) Loss Function: Logistic Regression

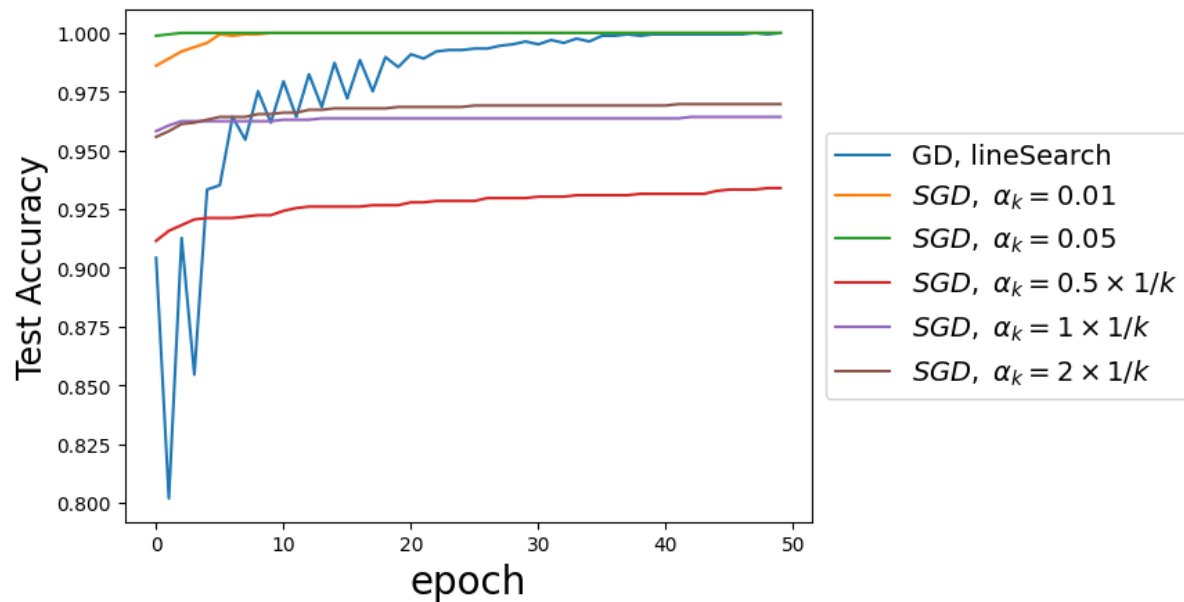
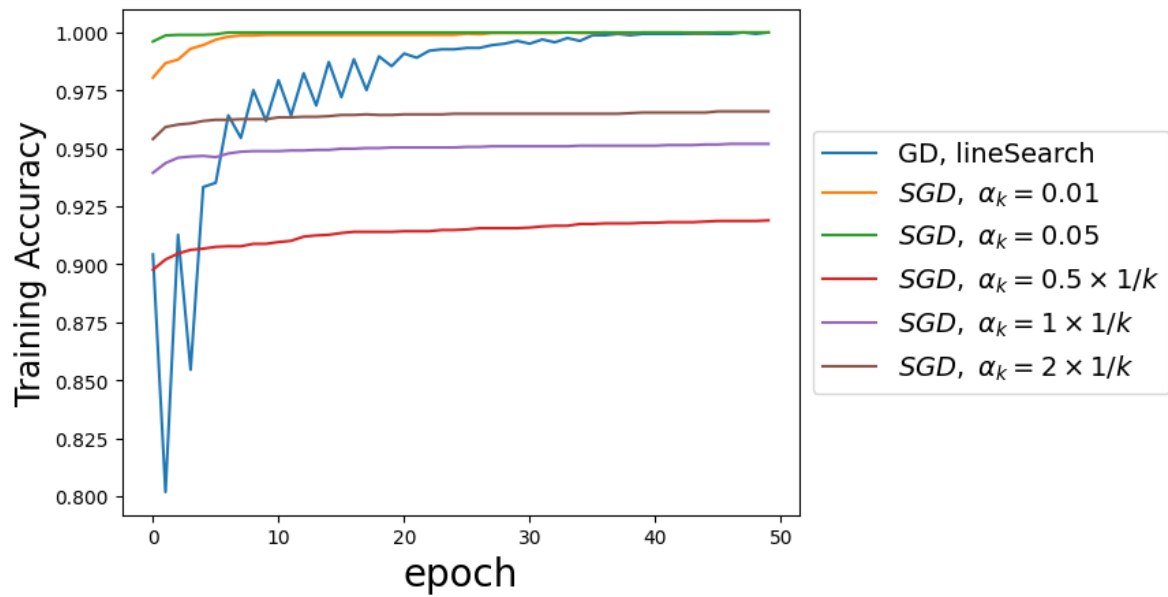
i. Optimality Gap



ii. Testing Loss

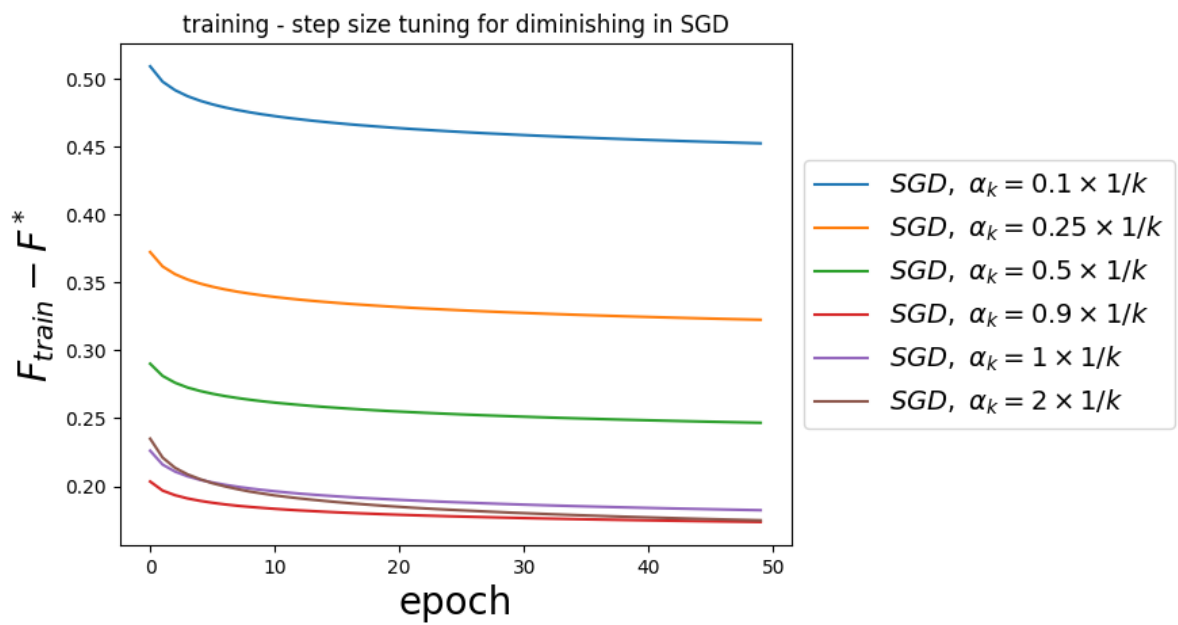
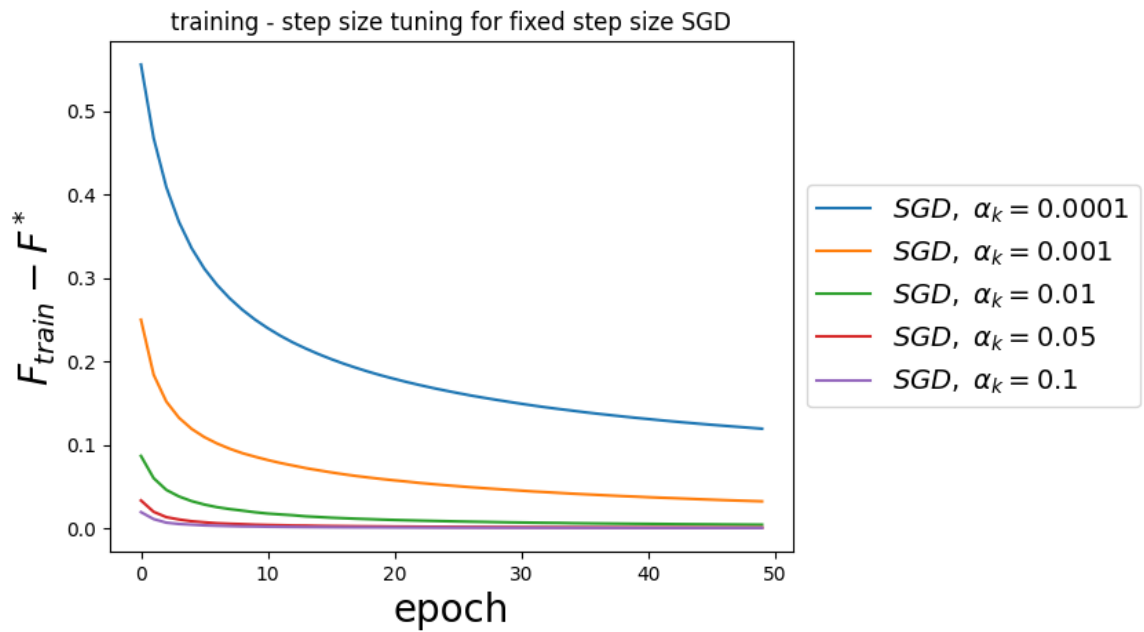


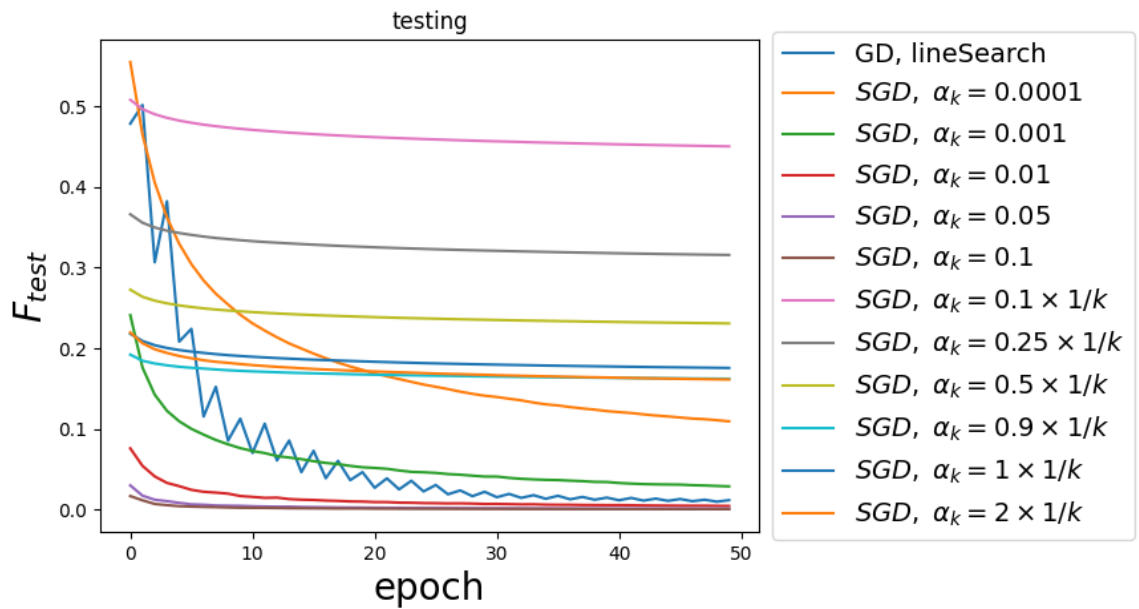
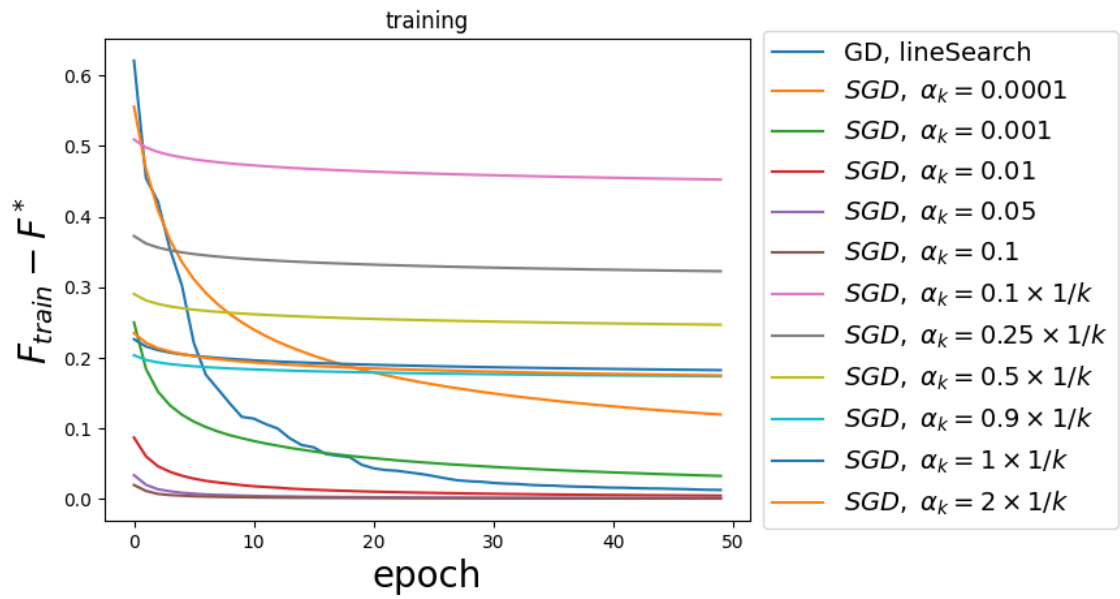
iii. Training and Testing Accuracy

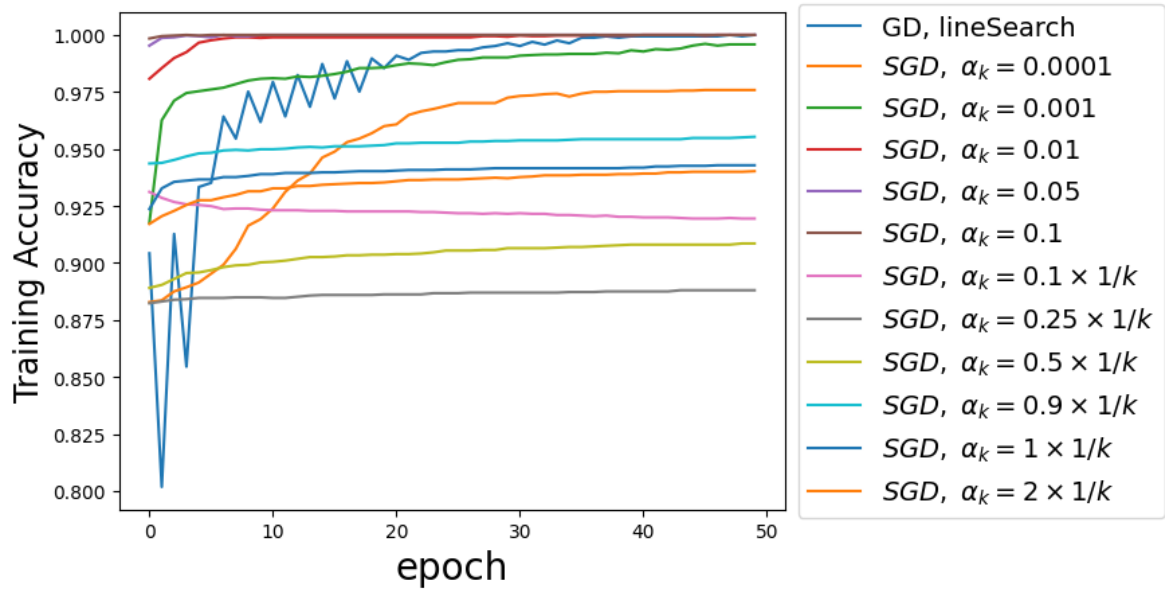


- **Step size tuning:**

With mushroom data and the model of fixed step size SGD with linear least squares loss function, I have test with multiple alphas and roughly could conclude that greater alpha values (in 0.01, 0.05, or 0.1) converge faster, which means that we can select comparatively bigger alpha for better performance. However, it should also need to be within a specific range of smaller than 1 so that it is not very easy/clear to find good parameters straightforwardly.







- Batch Size**

With using this mushroom dataset and using different mini batch sizes as displayed in the plot below, we can find that bigger batch sized model converges slower than the training models that have smaller batch size. This does affect the model convergence but does not have a serious impact on it as the step size plays its role. (I used the fixed $\alpha = 1$ for this model)

