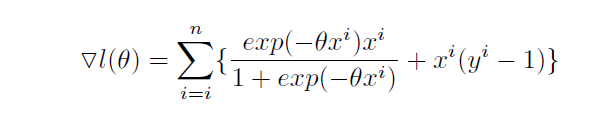
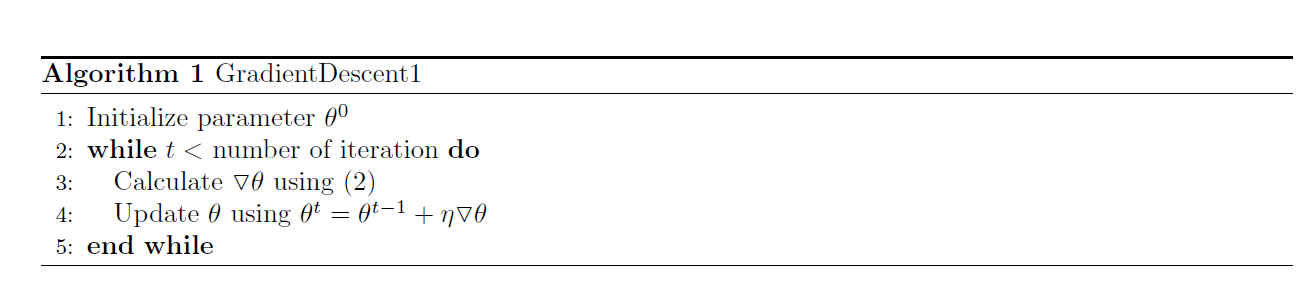
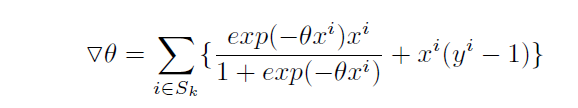
Problem 2:

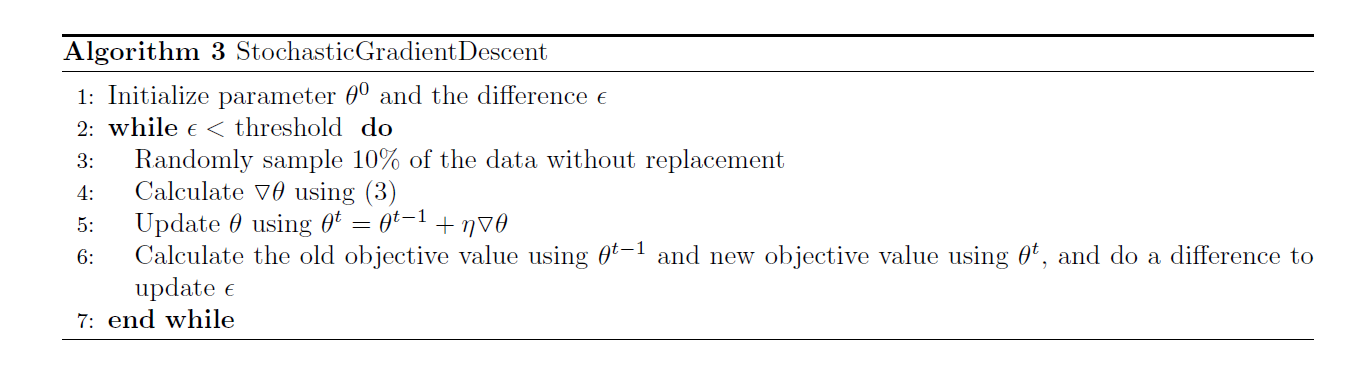
1. 

Pseudo-Code

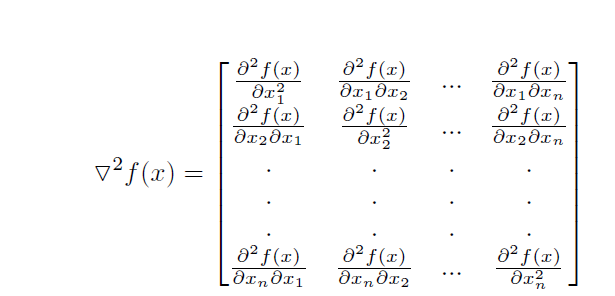


1. Stochastic Gradient Descent

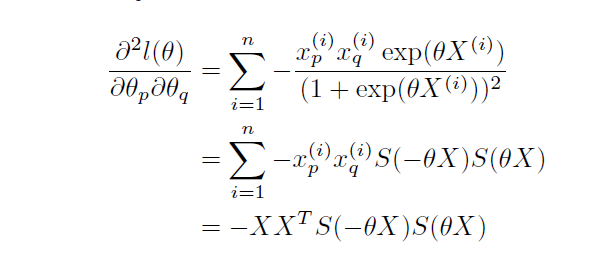




Hessian Matrix



Second Derivative of Log-likelihood function



Where



As S is the sigmoid function, S is always positive, as is XXT. Therefore, the training problem is concave. Thus, it does not has local minima, and with sufficiently small training rate, the gradient descent will converges to global minima.