W = vector of size #features

b = scalar of size 1

initialize W and b with all zeros

alpha = 0.00001

n = #features

m = #training samples

for i in range(500) :

predictions = W.X\_train + b

error\_grad = predictions – y\_train

b = b – alpha \* sum(error\_grad)

for j in range(n):

W[j] = W[j] – alpha \*sum( (error\_grad\*jth column of X\_train))