Descripción de lo hecho para aplicar rmsprop

Después de buscar cómo funcionaba el optimizador, empecé a querer aplicarlo con las ecuaciones que dio. Primero pensaba que había que cambiar la función definida SGD pero luego vi que no era necesario ya que dentro de SGD se llama a update\_mini\_batch, y dentro de update\_mini\_batch se llamaba a backprop. A mi manera de entenderlo, los optimizadores son una mejora de SGD y no de backpropagation por lo que decidí moverle en update\_mini\_batch.

Salieron varios errores y fue pura prueba y error hasta que corrió. Algunos de los errores fueron que no tenía dónde guardar los promedios, que dividía entre cero (no había puesto épsilon), y no pude resolver completamente el error:

RuntimeWarning: overflow encountered in exp

return 1.0/(1.0+np.exp(-z))

Respecto a la capa soft-max y cross-entropy, creo que fue más fácil poner cross-entropy pero ya no supe cómo arreglarlo. Tratando de arreglar lo de soft-max probé definiendo soft-max dentro de la red y fuera de la red (como estaba definida sigmoid) pero ni eso funcionó (olvidé ponerlo en el commit)