1.3

O método de gradient descente é usado para minimizar o risco empírico para uma determinada função de custo(POR ESSAS EQS), isto é feito através da adaptacção dos segundo as seguintes equações para o modo de treino em *batch* e usando momentum :

Onde corresponde ao peso do ramo com nó inical i e final j, (i=0 para DC)

Para calular o é necessário primeiro fazer na rede uma propagação da rede para a frente de forma a obtermos os e posteriormente uma propagação para trás obtendo os

(ESTA CURTA DEMAIS)

StepSize = 0.3

Momentum term = 0.5

(Qual a diferenca entre epoch e tempo? Nao sao lineares?)

Mediana dos epch =152

(E’ preciso justificar alguma coisa)

1.4)

A expressão permanece igual à representada na alínea 1.3 com uma pequena adição, o é agora também modificado a cada iteração sendo que:

Recalcular epochs <-params do prof?

Estranha

Para o mesmo momentum term () como este algoritmo é mais resistente a variações dos parâmetros, a variação do número de epochs é bastante baixo, da ordem das unidades,

No caso em que não existe passo adaptativo ocorrem grandes alteração o número de epochs ao variar minimamente os parametros. Como neste caso estamos a usar passo adaptativo, o valor inicial de tem pouca importância

Matriz confusao (é suposto ser cerca de 40% de accuracy?)

Os percentagem de vezes que a rede acerta em determinado padrão à entrada é relativamente baixa 40%. Sendo que obtém diferentes percentagens para dígitos diferentes. É também de notar que ao retreinar a rede, com o mesmo conjunto de treino, as percentagens com que a rede acerta no dígito alteram-se drasticamente devido aos pesos iniciais serem agora outros. Estavamos à espera de uma test set accuraccy de cerca de 80%.

Training error <- Corresponde ao MSE final? Probably not= recalcular 0.009629

Test accuraccy ~40%

O Test accuracy é um parâmetro mais relevante na avaliação da rede pois estamos a passar por esta padrões novos que esta não viu ainda. Além disso pode ocorrer overfitting que nao é verificável no MSE mas que entende-se que existe ao observar a test accuracy.