CS 754 : Advanced Image ProcessingAssignment 4

Meet Udeshi - 14D070007
Arka Sadhu - 140070011

April 4, 2017

Q1

A1.1 Derivatives for Calculation of Gradient Descent

The given cost function is

$$E(W,h_i) = \sum_{i=1}^{N}\sum_{j=1}^{n}(-y_{ji}log(Wh_i)_j + (Wh_i)_j) + \lambda\sum_{k=1}^{K}\sum_{j=1}^{N}h_{kj} \tag{1}$$

- Gradient of W:
  We note to get the gradient of a matrix, we need to differentiate with respect to every element of the matrix. The gradient matrix will simply be derivative to of the cost function with respect to the correspoding element of W.

  Let us consider the derivative with respect to the a^{th} row and b^{th} column denoted by w_{ab}.

  $$[\nabla W]_{ab} = \frac{dE}{dw_{ab}}$$

  $$\frac{dE}{dw_{ab}} = \frac{d}{dw_{ab}}(\sum_{i=1}^{N}\sum_{j=1}^{n}(-y_{ji}log(Wh_i)_j + (Wh_i)_j) + \lambda\sum_{k=1}^{K}\sum_{j=1}^{N}h_{kj})$$

  We note that

  $$(Wh_i)_j = \sum_{l=1}^{K}W_{jl}h_{li}$$

  Clearly the second term of the cost function is not dependent on W and hence goes to zero. For the first term we need to consider only the a^{th} row

  $$\frac{dE}{dw_{ab}} = \frac{d}{dw_{ab}}(\sum_{i=1}^{N}(-y_{ai}log(Wh_i)_a + (Wh_i)_a)$$

  Consider the first term

  $$\frac{d}{dw_{ab}}\sum_{i=1}^{N}(-y_{ai}log(Wh_i)_a) = \sum_{i=1}^{N}\frac{d}{dw_{ab}}(-y_{ai}log(\sum_{l=1}^{K}W_{al}h_{li})) = \sum_{i=1}^{N}-\frac{y_{ai}}{(Wh_i)_a}(h_{bi})$$

  Now consider the second term

  $$\frac{d}{dw_{ab}}\sum_{i=1}^{N}(Wh_i)_a = \sum_{i=1}^{N}h_{bi}$$

  Therefore

  $$\frac{dE}{dw_{ab}} = \sum_{i=1}^{N}-\frac{y_{ai}}{(Wh_i)_a}(h_{bi}) + h_{bi} = \sum_{i=1}^{N}(1-\frac{y_{ai}}{(Wh_i)_a})h_{bi}$$

  We vectorize the equaiton for matlab as follows

  $$[\nabla(W)] = (1 - Y./WH)H^{'}$$

- Gradient of H:
  We follow similar strategy to get the gradient of H. We differentiate the cost with respect