Machine Learning - Exercise 2  
Theoretical Part

|  |  |
| --- | --- |
| Name: | Tomer Gill (תומר גיל) |
| ID: | 318459450 |
| U2 Username: | gilltom |
| Date: | 31/03/2018 |

# Multiclass Logistic Regression

1. , where indicates the i-th index of the vector / matrix, **W** the weights matrix and **b** the bias vector.
2. I'll calculate the gradients of the average loss for all the training examples w.r.t each weight () and the bias ():  
     
   **Weights:**   
   We'll call the derivative of the loss of example t as , so in total . For (where y is the correct tag), **.** For : .  
   In conclusion, ,   
   where .  
     
   **Bias**: , so like before we'll look at the case for a single example t: , and for it's the same without the -1.  
   For conclusion, ,   
   where .

Now, for the update rules: the general SGD update rule is for each row in the weight matrix , and for each element in the bias vector . So the update rules are: