

ComS 573 Machine Learning

Problem Set 3

Note: Please do not hesitate to contact the instructor or the TA if you have difficulty understanding or getting started with solving any of the problems.

1. (20 pts.) Consider the multiclass logistic regression model.

(a) (5 pts.) Show that the derivatives of the softmax activation function are given by

$$\frac{\partial y_k}{\partial a_j} = y_k(\delta_{kj} - y_j).$$

(b) (15 pts.) *Derive* the batch and single sample gradient descent weight update rules for minimizing the cross-entropy error function.