

# Assignment 1

## Data mining

Prashanth K (08305006), Prashant S. Borole (08305022)  
Sriram Kashyap (08305028), Anup Kulkarni (08305022)  
IIT Bombay

September 06, 2008

### Loss functions used

#### Square loss

$S = (Y - XW)^T(Y - XW)$   
and the gradient is  $G = \frac{\partial}{\partial W} S = -2X^T(Y - XW)$

#### Additive logarithm loss

$F = \sum_i [\ln(1 + e^{(k(Y_i - X_i W))}) + \ln(1 + e^{(-k(Y_i - X_i W))})]$   
gradient  $G = \frac{\partial}{\partial W} F = \sum_i [\frac{-kX_i}{1+e^{(k(Y_i - X_i W))}} + \frac{kX_i}{1+e^{(-k(Y_i - X_i W))}}]$   
where  $k$  is the parameter.

say  $P_i = (Y_i - X_i W)$ , then for large values of  $P_i$ ,  $\frac{1}{1+e^{-kP}} \approx 1$  and  $\frac{1}{1+e^{kP}} \approx 0$