Logistic Regression – Loss function & Gradient Descent

where .

***Note***: for the multinomial logistic regression, the loss function is therefore

It can be written in the special form when category and

is the cost/loss function; while is the log likelihood function (can be derived from Bernoulli). To minimize the cost/loss function is equivalent to maximize the likelihood function. The gradient descent method can be used to minimize the cost/loss function; while Newton’s method can be used to find the maximum arguments for parameters by solving the 1st derivative of .

where we have

where , and are vectors. Also the constant term is