You are given a training set , where each document , and the label . Consider the *negative logistic loss* function

where is an arbitrary weight vector (not necessarily the solution to logistic regression.

Prove that upper bounds the number of classification mistakes makes on the training set. Hint: makes a mistake on if .

By definition, makes a mistake if:

Since

Since

Since

Since makes at most classifications, we know that the number of classification mistakes . However, by above we know that . Thus, we know .