

1. The log-likelihood function of is:
2. After ignoring all of the unnecessary constants, the Maximum Likelihood Estimate of is , then

where

Therefore,

In conclusion, finding MLE of is equivalent to solving .

1. Let be a non-zero vector, we have

Moreover, , hence , which implies that is positive definite.

1. First simplify the objective function,

Then, find the partial derivative with respect to and set it equal to 0.

Solve this equation, we can get

Then find the second order partial derivative with respect to

Therefore, is the solution of

1. According to Bayes’ formula

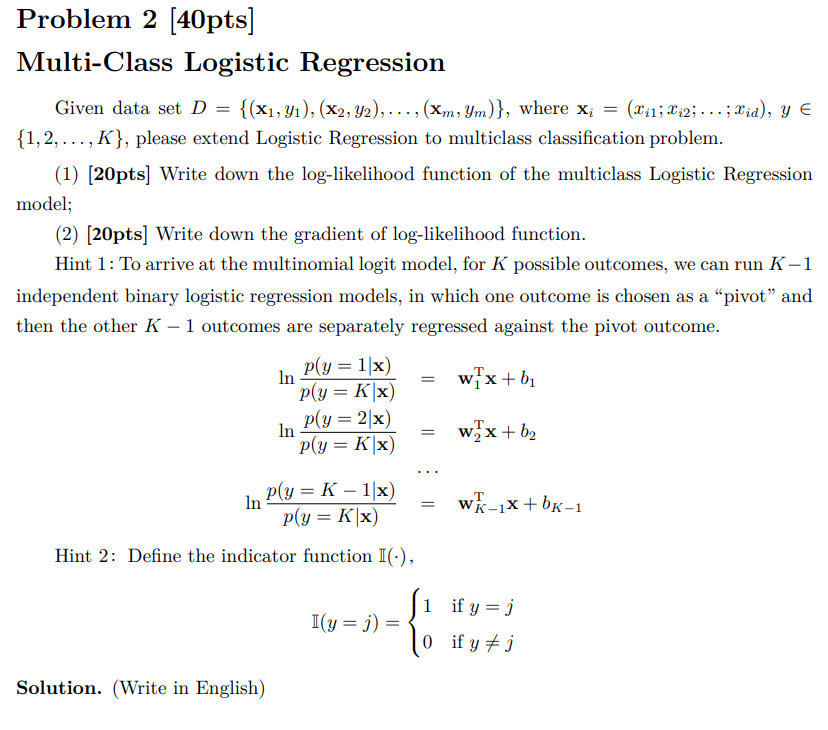
Write the distribution of in vector form

Then

Let be the log-likelihood function of , and logarithmization

Find the partial derivative with respect to and set it equal to 0, solve the equation, we get

Compare with the solution in (4), we find that is equal to when .



1. Let’s start from logistic function , where . After logarithmization, we can get

According to the Hint 1, we can run independent binary logistic regression models, in which one output is chosen as a main category and other outputs are separately regressed against the main category.

Treat in the formula as a class posterior probability estimate, and then rewrite the above formula to get

For all categories, we have , hence

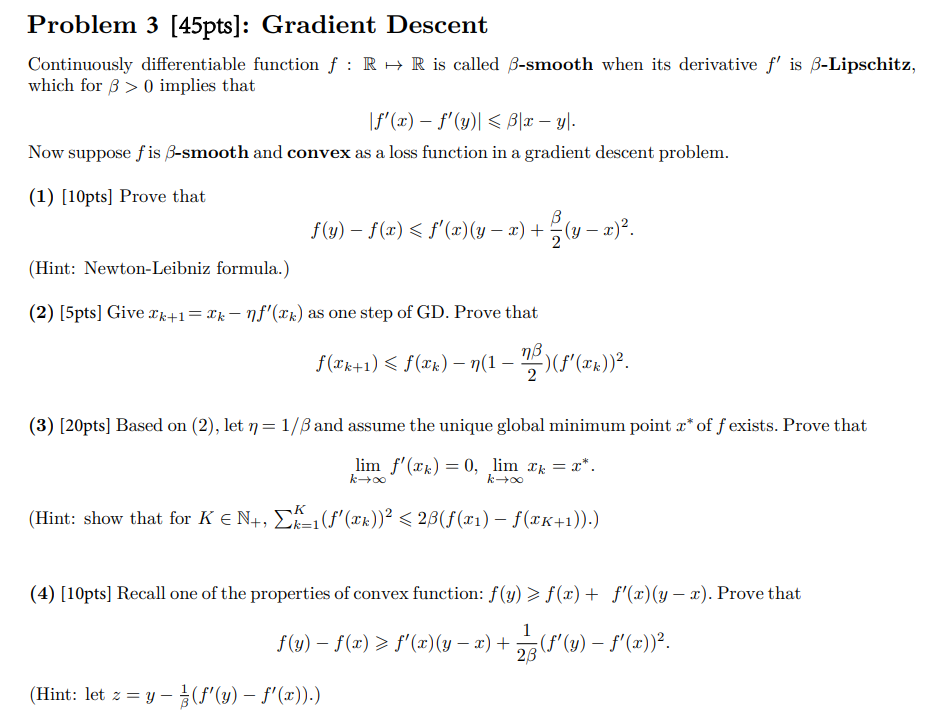
Next, only need to substitute into , then we can calculate all the probabilities .

Finally, the log-likelihood function of multiclass logistic regression model is

Let , then can be written as in short. And let , then likelihood term in can be rewritten as

Finally, we can obtain

1. The gradient of log-likelihood function is



1. From the Newton-Leibniz formula, it is apparent that

Then by the property of convex function and *β*-smoothness

1. Based on (1)
2. Based on (2), let , we have

is converge, hence .

Next, we can prove by contradiction. Assume that , which means

From , we know

Note that is the unique global minimum point and is monotony, hence take and , we get . Let , . However, for the same ,there should be , which means . This is contradictory. In conclusion, .

1. Let , based on (1), we have

Therefore,

From the property of convex function

Add the above two formulas together