XCS299i Problem Set #2

JUAN RICARDO PEDRAZA ESCOBAR

Machine Learning

Stanford Center for Professional Development

August,2021

1.a

Since and , it follows *∂h*(*x*)*/∂θk* = .

Letting = , we have

Replacing *∂h(x)/∂θk ,*

Substituting into our equation for , we have

Consequ**e**ntly, the (k, l) entry of the Hessian is given by,

By deriving whit respect to *,*

Using the fact that Xij = xixj if and only if X = xxT , we have

To prove that H is positive semi-definite, show ≥ 0 for all z ∈ Rd.

* Note that the function follows 0 ≤ ≥ 1 and the square term is always positive so ,

1.c

For shorthand, we let denote the parameters for the problem. Since the given formulae are conditioned on y, use Bayes rule to get:

We replace the equations that we use to model the distribution of ,

We take the quadratic term located in the previous expression denominator and apply properties of distribution,

Simplifying and reorganizing the term,

Therefore,

Where,

Finally gives,

1.d

First, derive the expression for the log-likelihood of the training data:

Using the equations for this model,

Now, the likelihood is maximized by setting the derivative (or gradient) with respect to each of the parameters to zero.

# For :

Setting this equal to zero and solving for gives the maximum likelihood estimate.

# For :

Hint: Remember that (and thus ) is symmetric.

Setting this gradient to zero gives the maximum likelihood estimate for µ0.

# For µ1:

Hint: Remember that Σ (and thus Σ−1) is symmetric.

Setting this gradient to zero gives the maximum likelihood estimate for µ1.

* For Σ, note that Σ =,

Setting this gradient to zero gives the maximum likelihood estimate for

1.f

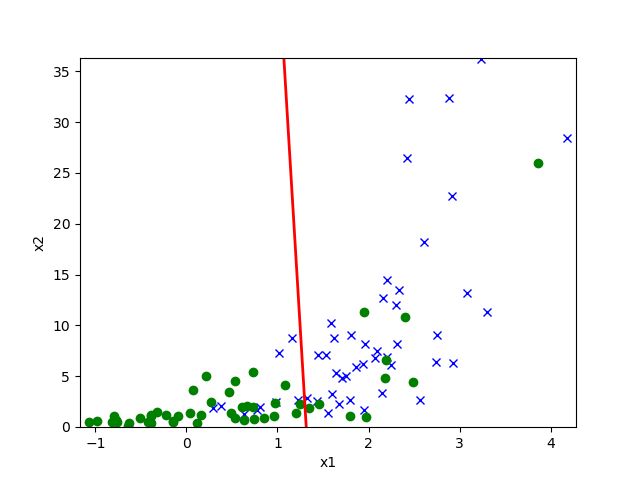


Figure 1 . GDA decision boundary on data set 1

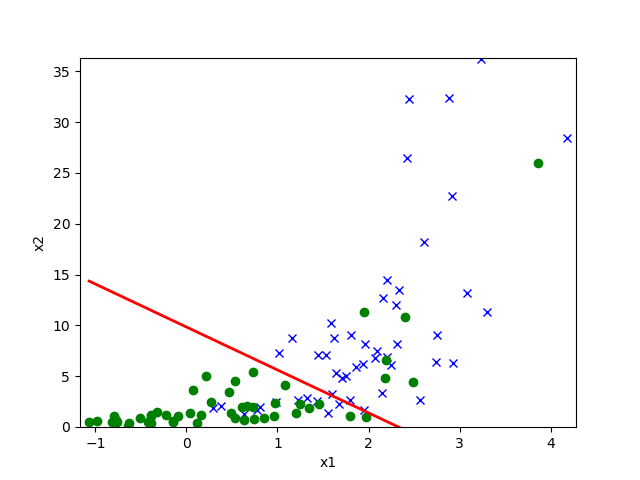


Figure 2. Logistic regression decision boundary dataset 1.

* Comment

The decision boundary from logistic regression seems more reasonable on Dataset1. The logistic decision boundary puts weight on both x1 and x2. However, with the GDA decision boundary the accuracy is lower on Dataset 1.

1.g

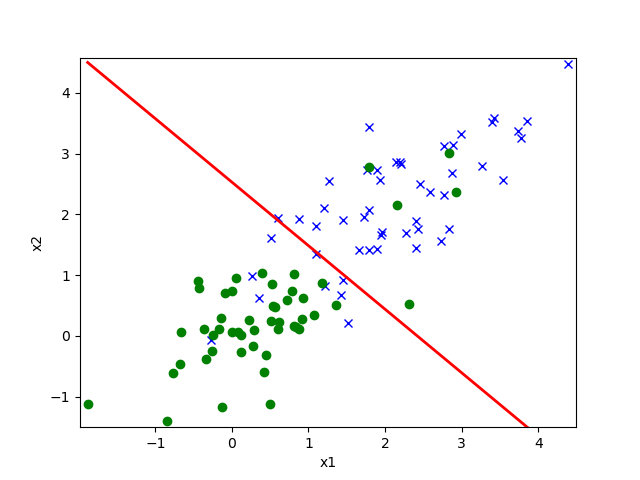


Figure 3. GDA Decision Boundary on Dataset 2.

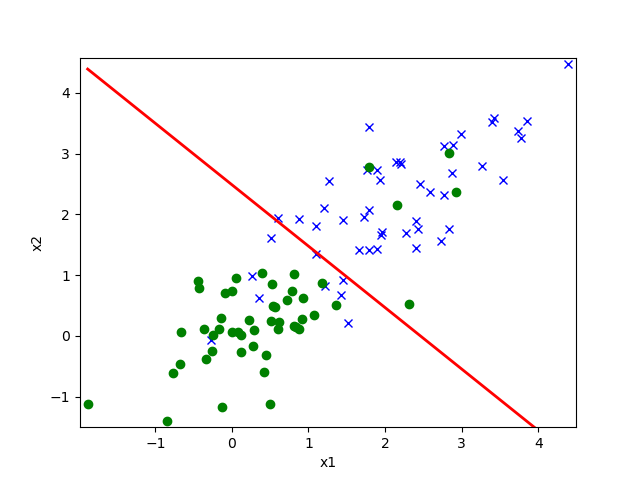


Figure 4. Logistic regression Decision Boundary on Dataset 2.

* Comment

Decision boundaries from logistic regression and GDA are nearly identical for Dataset 2. However, in the first dataset (Dataset 1) GDA seem to perform worse than logistic regression. This probably relates to the fact that the training data(xi) are not Gaussian and make the GDA model performs worse. Also, as we already know logistic regression is more powerful when the data set is not drawn from a multivariate Gaussian.

1.h

* Comment.

By setting,

all in the Dataset 2 become Gaussian, therefore the GDA model performs significantly better.

2.a

Using the standard form for the exponential family to compare,

2.b

For GML model the canonical response function will be,

2.c

The log-likelihood of an example is defined as. To derive the stochastic gradient ascent rule, use the results in part (a) and the standard GLM assumption that

|

Thus, the stochastic gradient ascent update rule should be:

which reduces here to: