**LOGISTIC DISCRIMINANT ANALYSIS**

(Paper From IEEE)

----------------------------------------------------------------------------------------------------

Prerequisite

* Bayesian Decision Theory
* Multi-Variate Linear Algebra
* Numerical Computations

----------------------------------------------------------------------------------------------------

(Methods Used Before Logistic Discriminant Analysis)

Linear Discriminant Analysis

It is well known linear-based method to extract the features for multi-class discrimination/classification.

The main objective of the linear discriminant analysis is to maximize the discriminant criterion which is used to evaluate the performance of new feature vector y, which is generated due to

Discriminant Criterion

Where ∑T and ∑B are total covarience and between class covarience.

dimension reduction linear mapping in original input feature vector. We maximize the discriminant criterion to obtain the **optimal coefficient matrix** A by solving an eigen equation.



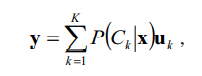
*The jth column of A is the eigenvector corresponding to the jth largest eigenvalue*. Therefore, the importance of each element of the new feature vector y is evaluated by the corresponding eigenvalues. ***The dimension of the new feature vector y is bounded by min (K -1, N).***

----------------------------------------------------------------------------------------------------

(Foundation for Logistic Discriminant Analysis)

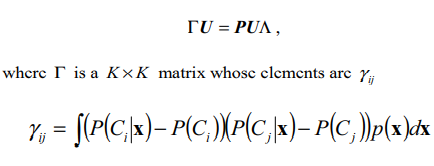
Optimal Non-Linear Discriminant Analysis

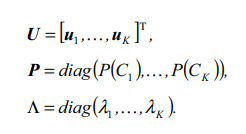
In this analysis, a dimension reducing non-linear mapping is constructed to maximize the discriminant criterion J.

Optimal Non-Linear Discriminant Mapping

Where

The vectors uk (k = 1, …, K) are class representative vectors which are determined by following eigen-equation:

And,



It is important to remember that the optimal non-linear mapping is closely related to Bayesian decision theory, namely the *posterior probabilities* P(Ck|x).

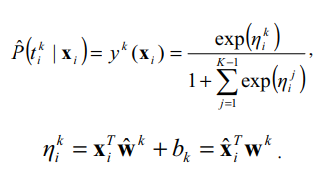
**Thus, we construct optimal nonlinear discriminant features by ONDA from a given input features if we can know or estimate all the Bayesian posteriori probabilities correctly.**

----------------------------------------------------------------------------------------------------

Multi-Nominal Logistic Regression

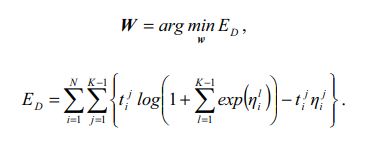
Logistic regression (LR) is one of the simplest models for binary classification and can directly estimate the posterior probabilities.

Multi-nominal logistic regression (MLR) is a natural extension of LR to multi-class classification problems. By modifying the outputs of the linear predictor by the link function, MLR can naturally estimate the posterior probabilities.

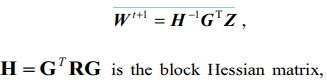
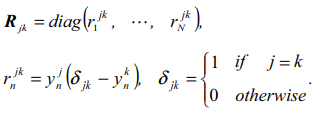
For K-class classification problem, The outputs of MLR estimate the posterior probabilities P(tki | xi) . They are defined as follows:

Outputs Of MLR / Posterior Probabilities

The optimal parameters of MLR are obtained by minimizing the negative log-likelihood which is a convex optimization problem and it has only a single, global minimum.

Minimized negative log-likelihood

Again, the optimal parameter W can be efficiently found using Newton-Raphson method or an iterative re-weighted least squares (IRLS) procedure. In each iteration step, W is updated as

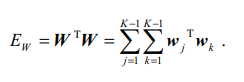
 

Parameter W is updated continuously until it converges.

*Regularization of MLR*

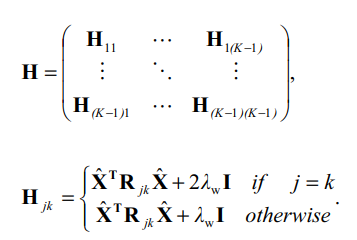
In general, the regularization term is introduced to control the over-fitting. The regularization methods of MLR were proposed such as the shrinkage method (regularized MLR) and locality preserving multi-nominal logistic regression (LPMLR).

In shrinkage method, unnecessary growth of the parameters is penalized by introducing the regularization term EW defined as follows:



the optimal parameters of the regularized MLR is determined by minimizing the negative log-likelihood as given above, which is a convex optimization problem, and lambda w is the pre-specified regularization parameter of Ew .

The multiplicative update rule for the regularized MLR is the same as (19). However, the elements of the block Hessian matrix H are different from MLR [10]. the block Hessian matrix H of the regularized MLR is defined as follows:



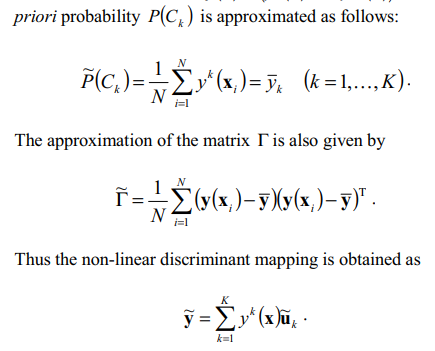
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Logistic Discriminant Analysis

the outputs of the ordinal MLR or the regularized MLR can be interpreted as estimates of the Bayesian posterior probabilities (P(C1|X),….., P(Ck|X))T.

Logistic Discriminant Analysis is obtained by substituting the Bayesian probabilities with the outputs of regularized MLR in ONDA.

It is expected that the discriminant space constructed by LgDA is better than the one constructed by LDA, because MLR is more natural as the estimates of the posterior probabilities than the linear approximation of them used in LDA.



The representative vectors of each class u~k are determined by the following eigen equation:

Class Representative Vector



*The nonlinear discriminant mapping constructs the good approximation of the ultimate nonlinear discriminant mapping ONDA in terms of the discriminant criterion.*

---------------------------------------------------------------------------------------------------

Examples of Discriminations

