Robust Adjusted Likelihood Function for Image Analysis

Rong Duan, Wei Jiang, Hong Man

Department of Electrical and Computer Engineering
Stevens Institute of Technology

STEVENS Institute of Technology

Visual Information Environment Laboratory

Outline

- Objective: study parametric classification method when model is misspecified
- Method: *robust adjusted likelihood function (RAL*)
- Contents:
  1. Likelihood function under true model
  2. Model misspecification
  3. Robust adjusted likelihood function
  4. Simulation and application experiment
  5. Conclusion

Likelihood

- Let $x_1, \ldots x_n$ be independent random variables with pdf $f(x_i;\theta)$
  - the likelihood function is defined as the joint density of n independent observations $X=(x_1, \ldots, x_n)'$

$$f(X;\theta) = \prod_{i=1}^{n} f(x_i;\theta) = L(\theta;X)$$

  - the log form is

$$\log(L(\theta;X)) = \sum_{i=1}^{n} \log(f(x_i;\theta))$$

Likelihood

- The Law of Likelihood (Hacking 1965)
  - *If one hypothesis* $\boldsymbol{H_1}$*, implies that a random variable* $\boldsymbol{X}$ *takes the value* $\boldsymbol{x}$ *with probability* $\boldsymbol{f_1(x)}$*, while other hypothesis* $\boldsymbol{H_2}$*, implies that the probability is* $\boldsymbol{f_2(x)}$*, then the observation* $\boldsymbol{X=x}$ *is evidence supporting* $\boldsymbol{H_1}$ *over* $\boldsymbol{H_2}$ *if* $\boldsymbol{f_1(x)>f_2(x)}$*, and the likelihood ratio,* $\boldsymbol{f_1(x)/f_2(x)}$*, measures the strength of that evidence*

Classification

- Binary classification problem: two classes of data $\{X_1\}=\{x_1^{(1)}, \ldots, x_n^{(1)}\}$ and $\{X_2\}=\{x_1^{(2)},\ldots, x_n^{(2)}\}$ from two distributions $g_1(x)$ and $g_2(x)$, where $g_1(x)$ and $g_2(x)$ are true distributions. We denote $l(x, g_2: g_1) = g_2(x)/g_1(x)$ the true likelihood ratio statistic when the data x comes from the true model.
- If the loss function is symmetric and the prior probabilities $q(\theta_k)$ are equal $\{q_{\theta_1} = \ldots = q_{\theta_k}\}$, the Bayes classifier can be expressed as a *maximum likelihood test*

$$i' = \arg\max \log(f_i(x,\theta_i))$$

Classification

- The decision boundary is

$$l(x,\theta_1)= l(x,\theta_2),$$

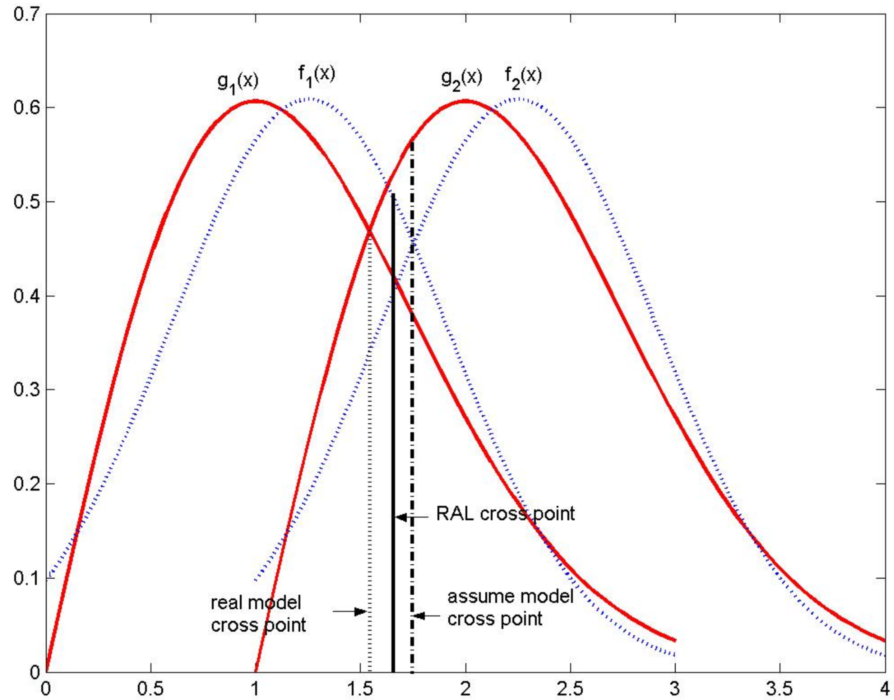
  where $l(x,\theta_i)=log\ f(x,\theta_i)$
- When the model assumption is correct, The Bayes classifier is optimum, it has the minimum error rate.
- The distribution parameters, θ_i, can be learned from training data using maximum likelihood estimation (MLE). However certain estimation error will be introduced, and estimated parameters are denoted as $\hat{\theta}_i$

Model Misspecification

- When the model assumption is incorrect, the maximum likelihood test will yield inferior classification results
  - The estimated model parameters may be erroneous
  - The distribution of the likelihood ratio statistic is no longer chi-square due to the failure of Bartlett's second identity

Model Misspecification

- A model misspecification example:
  – True model: $g_1(x)$, $g_2(x)$; assumed models: $f_1(x)$, $f_2(x)$



Robust Adjustment of Likelihood

- Stafford (1996) proposed a *robust adjustment* of likelihood function in the scalar random variable case,

$$f_\xi(x,\theta)=f(x,\theta)^\xi$$

- The intention is to correct the Bartlett's second identity, which equates the variance of the Fisher score

$$J(\theta)=E_g[u(\theta;X)u^T(\theta;X)]$$

and the expected Fisher information matrix

$$H(\theta)=-E_g\left[\frac{\partial^2\log(L(\theta))}{\partial\theta\partial\theta^T}\right]$$

- Analytical expressions for calculating the parameter, ξ, are only available for a very few distributions.

Robust Adjusted Likelihood Function

- We propose a general *robust adjusted likelihood* (RAL) function

$$f_a(x,\theta)=\eta f(x,\theta)^{\xi}$$

- The RAL classification rule becomes

$$i' = arg\ max\ \{log(\eta)+ \xi\ log(f_i(X,\theta_i))\}$$

- The classification boundary is

$$b + w\ l(x,\theta_1) = l(x,\theta_2),$$

where $b = \{log(\eta_1)- log(\eta_2)\}/\xi_2$ and $w =\xi_1/\xi_2$, this classification boundary is in a form of a linear discriminant function in likelihood space.

Robust Adjusted Likelihood Function

- The RAL introduces a data-driven linear discrimination rule $\boldsymbol{b + w\, l(x,\theta_1) = l(x,\theta_2)}$, where $\boldsymbol{w}$ and $\boldsymbol{b}$ are learned from training data.
  - If $\boldsymbol{w=1}$, the discrimination rule is similar to likelihood ratio tests whose evidence is controlled by the bump function if the parametric family includes $\boldsymbol{g_k(x)}$.
  - If $\boldsymbol{w=1}$ and $\boldsymbol{b=0}$, it reduces to the Bayes classification rule in the data space
- A major advantage of the RAL is that its classification rule includes the Bayes classification rule as a special case. Therefore, similar to likelihood space classification, RAL will not perform worse than Bayes classification.

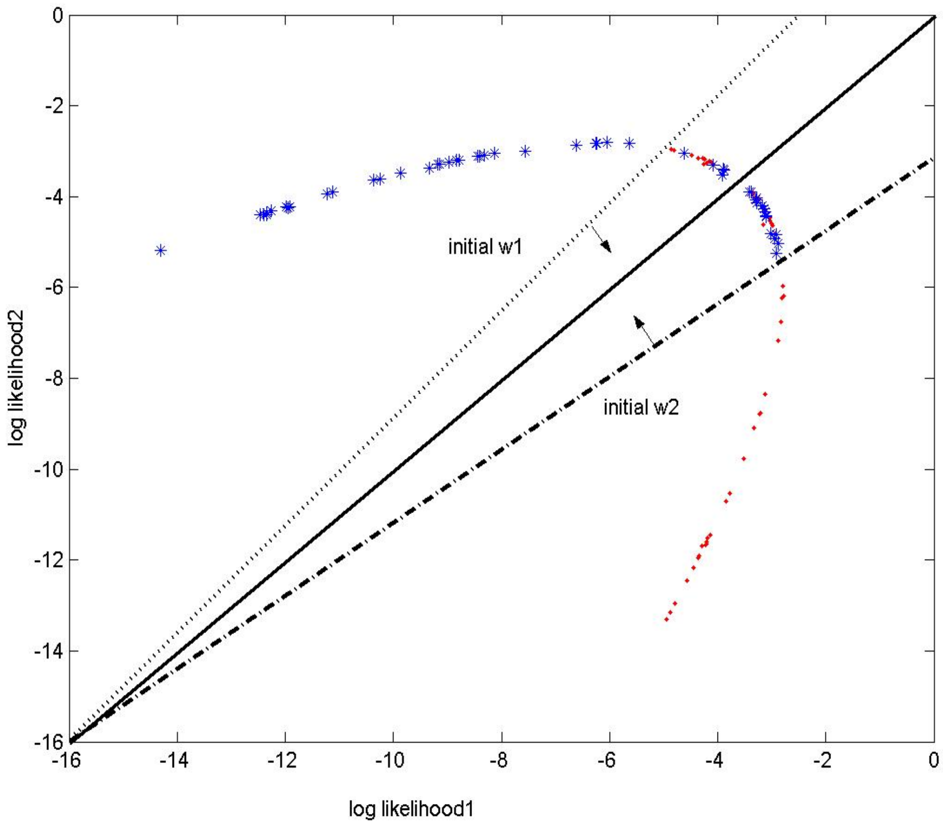
Minimum Error Rate Learning

- Likelihood space *minimum error rate learning* method to estimate (b,w):
  - For two classes of training data, $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$,

$$(b,w)=\arg\min\{P_{g_1}(l(X_1,\theta_2)-wl(X_1,\theta_1)>b)$$
$$+P_{g_2}(l(X_2,\theta_1)-wl(X_2,\theta_2)<b)\}$$

  - Algorithm:
    1. Initialize $\boldsymbol{w_1}$ minimizing error rate for $\boldsymbol{X_1}$, i.e. $\boldsymbol{e_1}$, and $\boldsymbol{w_2}$ minimizing error rate for $\boldsymbol{X_2}$, i.e. $\boldsymbol{e_2}$. Assuming $\boldsymbol{w_1 > w_2}$. Calculate total error rate $\boldsymbol{e=e_1+e_2}$
    2. If $\boldsymbol{w_1 \le w_2}$ or $\boldsymbol{e}$ is minimized, $\boldsymbol{w=(w_1+w_2)/2}$, stop
    3. Else, decrease $\boldsymbol{w_1}$ and increase $\boldsymbol{w_2}$ to calculate new error rate $\boldsymbol{e=e_1+e_2}$, goto step 2

Minimum Error Rate Learning



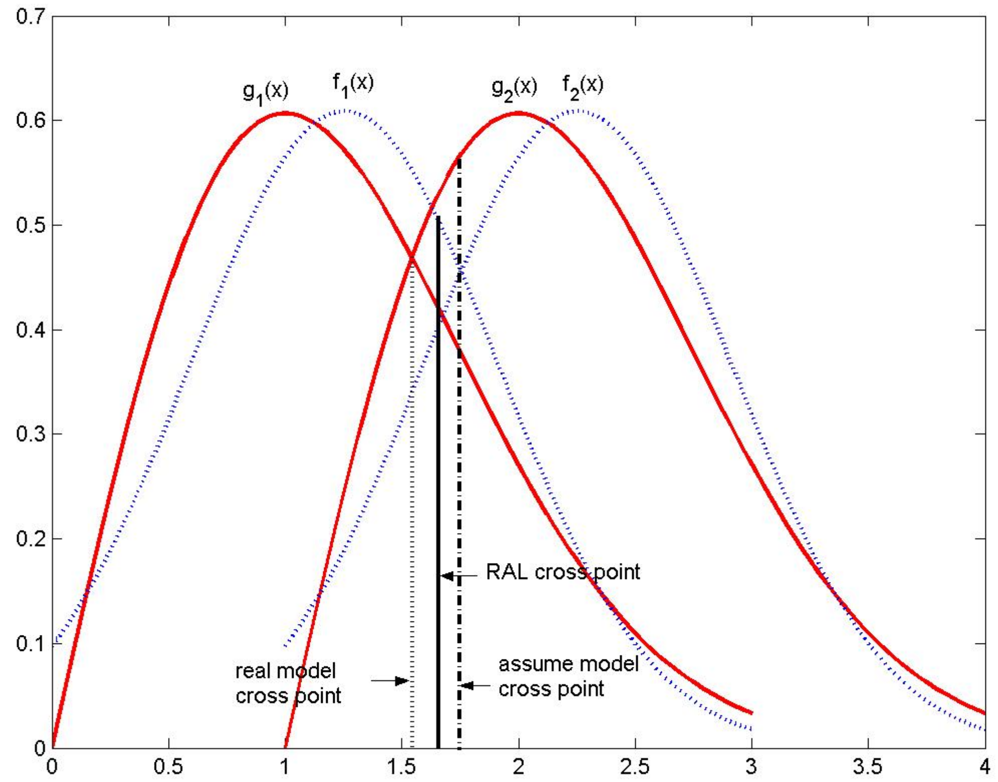
RAL Classification

- RAL classification algorithm
  - Training:
    1. Make model assumption
    2. Estimate model parameters $\boldsymbol{\theta}$ based on maximum likelihood method
    3. Estimate RAL parameter $(\boldsymbol{b},\boldsymbol{w})$ based on minimum error rate method
  - Testing:
    1. Calculate RAL of an input sample $\boldsymbol{y}$,
    2. Classify this sample based on the maximum RAL rule.

Study on Simulated Data

- Experiment:
  1. Two classes data are from two Rayleigh distributions with same scale and different locations. The assumed models are Gaussian distributions with same variance.
  2. The Bayes error rate of the true model, the Bayes error rate of the misspecified model, and the error rate of the robust adjusted likelihood classification are compared
  3. Repeat 100 times to get the average

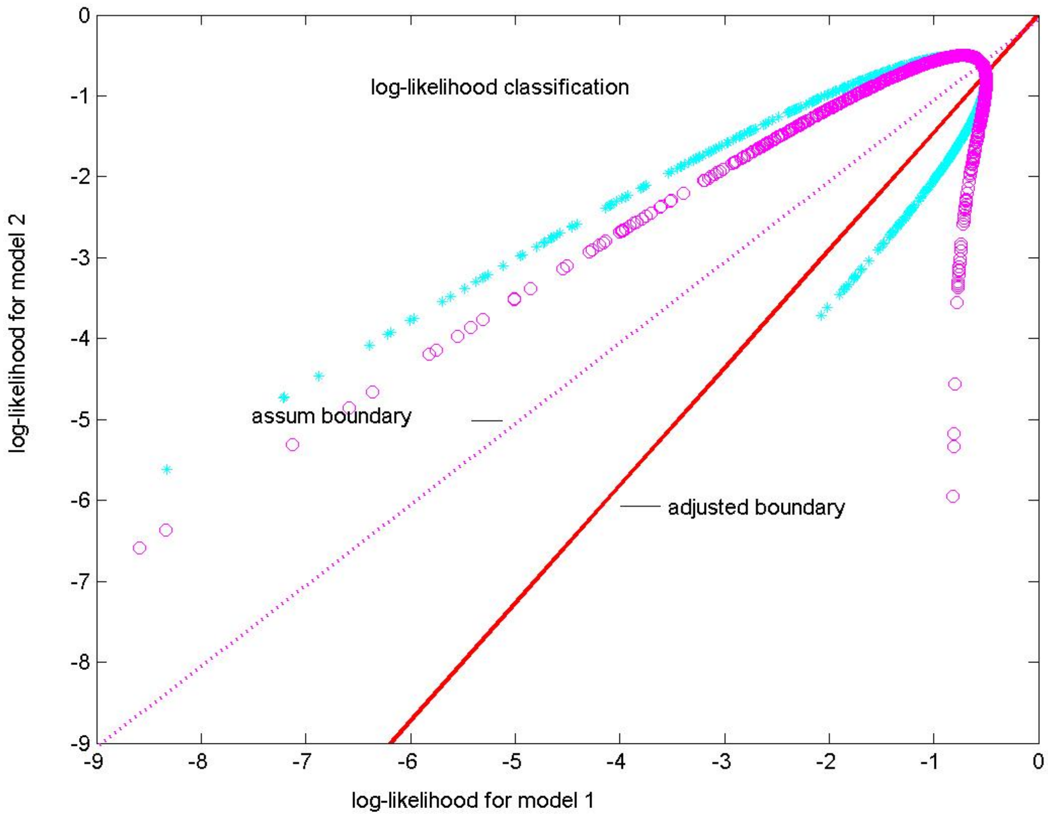
Study on Simulated Data



STEVENS Institute of Technology

Visual Information Environment Laboratory

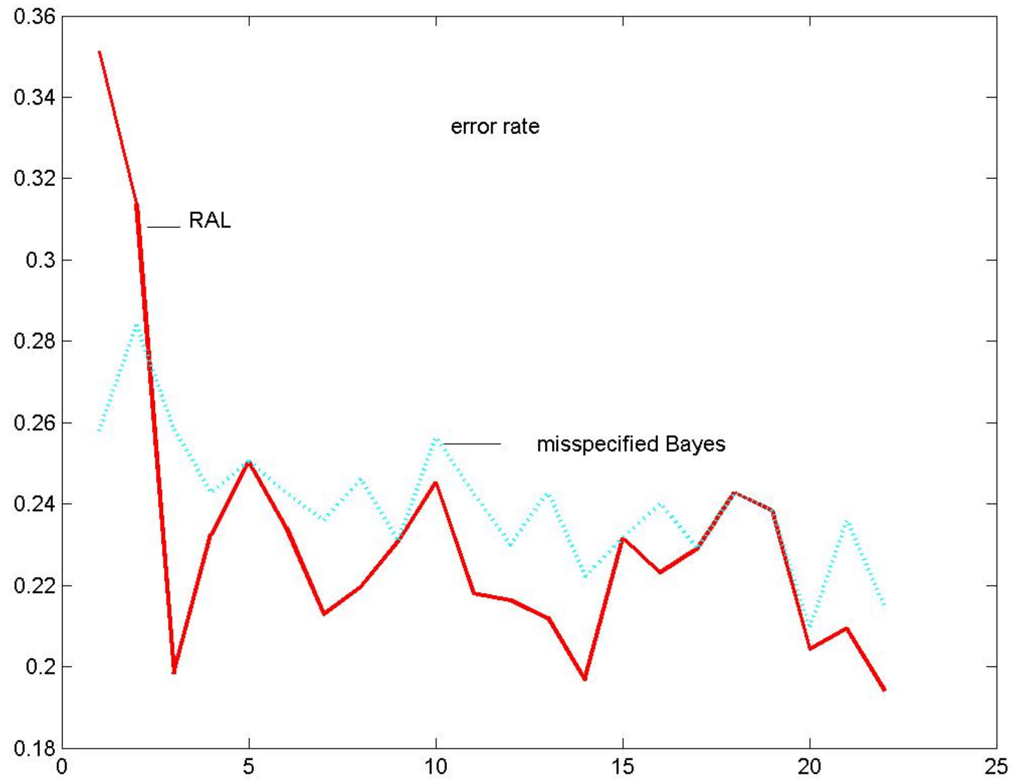
Study on Simulated Data



Application on SAR ATR

- Experiment:
  - MSTAR SAR dataset: T72, BMP2
  - Assumed models: 2 Gaussian Mixture Models (GMM) with 10 mixtures for each class.
  - Classification performance obtained for various training data sizes, with an increase of 10 samples each time.
- Observation:
  - Under a practical situation, accurate model assumption is difficult to obtain, and RAL classification has an advantage to provide certain robustness in parametric classification.

Application on SAR ATR



Conclusion

- The *RAL* classification is robust in classification when model assumption is not correct.
- Minimum error rate method is effective in estimating the raising power and scale parameters from training data
- In theory, *RAL* will not perform worse than the Bayes classifier.
- Further investigation is needed to obtain theoretical performance bound for *RAL* under various practical situations