1 Minimum Error Thresholding in Bayesian Decision Making

1.1 Description

Minimum error thresholding is a technique used in Bayesian decision making to classify data points (such as pixel values) based on conditional probabilities. The goal is to minimize the probability of misclassification by selecting the class that maximizes the posterior probability.

In the context of image processing, consider a pixel value x. The pixel can belong to one of two classes, C_0 or C_1. To decide the class membership of x, we compare the posterior probabilities $p(x|C_0)p(C_0)$ and $p(x|C_1)p(C_1)$, where $p(x|C_j)$ is the likelihood probability of x given class C_j and $p(C_j)$ is the prior probability of class C_j.

1.2 Mathematical Formulation

Given a pixel value x:

$$\begin{cases} C_0, & \text{if } p(x|C_0)p(C_0) > p(x|C_1)p(C_1) \\ C_1, & \text{otherwise} \end{cases}$$

Assume that the likelihood probabilities follow a normal distribution:

$$p(x|C_j) \in \mathcal{N}(x|\mu_j, \sigma_j^2)$$

This means:

$$p(x|C_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}}$$

The posterior probability can be calculated as:

$$p(x|C_j) \cdot p(C_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}} \cdot p(C_j)$$

Taking the natural logarithm of both sides to simplify the expression:

$$\ln(p(x|C_j) \cdot p(C_j)) = \ln(p(x|C_j)) + \ln(p(C_j))$$

Given that:

$$\ln(p(x|C_j)) = \ln \left[\frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}} \right]$$

We can break it down to:

$$\ln(p(x|C_j)) = \ln \left(\frac{1}{\sqrt{2\pi\sigma_j^2}} \right) + \ln \left(e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}} \right)$$

$$= -\ln \left((2\pi\sigma_j^2)^{1/2} \right) - \frac{(x-\mu_j)^2}{2\sigma_j^2}$$

$$= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_j^2) - \frac{(x-\mu_j)^2}{2\sigma_j^2}$$

Now we can write the sum of natural logarithms for the posterior probability as:

$$\ln p(x|C_j) + \ln p(C_j) = -\frac{1}{2}[\ln(2\pi) + \ln(\sigma_j^2) + \frac{(x-\mu_j)^2}{\sigma_j^2} - 2\ln p(C_j)]$$

To find the optimum solution, given that $\frac{1}{2}\ln(2\pi)$ is constant, we can maximize he following equation:

$$-\left[\frac{(x-\mu_j)^2}{\sigma_j^2}+2(\ln(\sigma_j)-\ln p(C_j))\right]$$

or alternatively find an optimum threshold that minimize this equation:

$$\mathcal{E}_j(x)=\left[\frac{(x-\mu_j)^2}{\sigma_j^2}+2(\ln\sigma_j-\ln p(C_j))\right]$$

Here, $\mathcal{E}_j(x)$ indicates a function that needs to be minimized to decide the class membership of the pixel value x. The components of this function include the squared difference between the pixel value and the class mean, normalized by the class variance, and the logarithmic difference between the class variance and prior probability. In this set up Given a pixel value x:

Assign x to class:

$$\begin{cases} C_0, & \text{if } \mathcal{E}_0(x)\leq\mathcal{E}_1(x) \\ C_1, & \text{otherwise} \end{cases}$$

For threshold q:

All pixels with intensity values:

$$\begin{cases} g\leq q\rightarrow C_0 \\ g>q\rightarrow C_1 \end{cases}$$

The goodness of classification by threshold q over N image pixels $I(u,v)$ can be described by:

$$e(q)=\frac{1}{N}\sum_{u,v}\begin{cases} \mathcal{E}_0(I(u,v)), & \text{for } I(u,v)\leq q \\ \mathcal{E}_1(I(u,v)), & \text{for } I(u,v)>q \end{cases}$$

This can be further simplified to:

$$e(q) = \frac{1}{N}\left[\sum_{g=0}^{q} h(g)\cdot \mathcal{E}_0(g) + \sum_{g=q+1}^{K-1} h(g)\cdot \mathcal{E}_1(g)\right]$$

Where $h(g)$ represents the histogram or frequency of pixel values. This summation can also be expressed as:

$$e(q) = \sum_{g=0}^{q} P(g)\cdot \mathcal{E}_0(g) + \sum_{g=g+1}^{K-1} P(g)\cdot \mathcal{E}_1(g)$$

Where $P(g)$ is the probability of pixel value g, with $P(g) = h(g)/N$.

Recall that error function $\mathcal{E}_j$ can be described as:

$$\mathcal{E}_j = \frac{(x-\mu_j)^2}{\sigma_j^2} + 2\left[\ln(\sigma_j) - \ln(p(C_j))\right]$$

We can place $\mathcal{E}_j$ in the goodness of classification function $e(q)$ which was described as:

$$e(q) = \sum_{g=0}^{q} P(g)\cdot \mathcal{E}_0(g) + \sum_{g=q+1}^{K-1} P(g)\cdot \mathcal{E}_1(g)$$

Breaking down the summation for $g \leq q$:

$$\sum_{g=0}^{q}\left[P(g)\cdot\frac{(x-\mu_j)^2}{\sigma_j^2} + 2P(g)\ln(\sigma_j) - 2P(g)\ln(p(C_j))\right] \tag{1}$$

And for $g > q$:

$$\sum_{g=q+1}^{K-1}\left[P(g)\cdot\frac{(x-\mu_j)^2}{\sigma_j^2}+2P(g)\ln(\sigma_j)-2P(g)\ln(p(C_j))\right]$$

Note that:

1. $\sum_{g=0}^{q} P(g) \approx \mathbb{P}_0(q) = \frac{1}{N}\sum_{g=0}^{q} h(g) = \frac{n_0(g)}{N}$
2. $\sum_{g=q+1}^{K-1} P(g) \approx \mathbb{P}_1(q) = \frac{1}{N}\sum_{g=q+1}^{K-1} h(g) = \frac{n_1(q)}{N}$

Accordingly, we can write equation 1 as:

$$\frac{1}{\sigma_0^2(q)}\underbrace{\sum_{g=0}^{q} P(g)\cdot(g-\mu_0(q))^2}_{=\sigma_0^2(q)\text{ by definition}}+2\ln(\sigma_0)\underbrace{\sum_{g=0}^{q} P(g)}_{\mathbb{P}_0(q)}-2\ln(p(C_0))\underbrace{\sum_{g=0}^{q} P(g)}_{\mathbb{P}_0(q)}$$

This can be expressed as:

$$\begin{aligned}&= \text{constant} + \mathbb{P}_0(q)2\ln(\sigma_0) - 2\mathbb{P}_0(q)\ln(p(C_0))\\ &= \text{constant} + \mathbb{P}_0(q)\ln(\sigma_0^2) - 2\mathbb{P}_0(q)\ln(\mathbb{P}_0(q))\end{aligned}$$

Following the same approach for class 1, we get:

$$e(q) = \text{constant}+\mathbb{P}_0(q)\ln(\sigma_0^2(q))+\mathbb{P}_1(q)\ln(\sigma_1^2(q))-2\mathbb{P}_0(q)\ln(\mathbb{P}_0(q))-2\mathbb{P}_1(q)\ln(\mathbb{P}_1(q))$$

Thus, the optimal threshold q^* is found by minimizing $e(q)$. Note that:

$$\sigma_0^2(q) = \frac{1}{n_0(q)}\left[\sum_{g=0}^{q} h(g)g^2 - \frac{1}{n_0(q)}\left(\sum_{g=0}^{q} h(g)g\right)^2\right]$$

To drive this equation, we start by the definition of $\sigma_0^2(q)$ as follows:

$$\sigma_0^2(q) = \sum_{g=0}^{q} P(g)(g - \mu_0(q))^2 = \sum_{g=0}^{q} [P(g)g^2 + P(g)\mu_0(q)^2 - 2P(g)g\mu_0(q)]$$

Applying the sum over all the terms:

$$= \sum_{g=0}^{q} P(g)g^2 + \mu_0^2(q) \sum_{g=0}^{q} P(g) - 2\mu_0(q) \sum_{g=0}^{q} P(g)g$$

Note that :

1. Accepting some error, we assume $\sum_{g=0}^{q} P(g) \approx 1$
2. $\sum_{g=0}^{q} P(g)g = \mu_0 q$

Then, we further simplify:

$$\approx \sum_{g=0}^{q} P(g)g^2 + \mu_0^2(q) - 2\mu_0^2(q)$$

$$\approx \sum_{g=0}^{q} P(g)g^2 - \mu_0^2(q)$$

$$\approx \sum_{g=0}^{q} P(g)g^2 - \left(\sum_{g=0}^{q} P(g)g \right)^2$$

$$\approx \frac{1}{n_0(q)} \left[B_0(q) - \frac{1}{n_0(q)} A_0^2(q) \right]$$

For $g > q$, the variance $\sigma_1^2(q)$ is:

$$\sigma_1^2(q) \approx \frac{1}{n_1(q)} \left[\sum_{g=q+1}^{K-1} h(g)g^2 - \frac{1}{n_1(q)} \left(\sum_{g=q+1}^{K-1} h(g)g \right)^2 \right]$$

1.3 Definitions

$$A_0(q) = \sum_{g=0}^{q} h(g)g$$

$$B_0(q) = \sum_{g=0}^{q} h(g)g^2$$

$$A_1(q) = \sum_{g=q+1}^{K-1} h(g)g$$

$$B_1(q) = \sum_{g=q+1}^{K-1} h(g)g^2$$

The Recursive Formula for $A_0(q)$

$$A_0(q) = \begin{cases} 0 & \text{for } q = 0 \\ A_0(q-1) + h(q)q & \text{for } 1 \leq q < K-1 \end{cases}$$

The recursive formula for $B_0(q)$ is:

$$B_0(q) = \begin{cases} 0 & \text{if } q = 0 \\ B_0(q-1) + h(q)q^2 & \text{if } 1 \leq q \leq K-1 \end{cases} \tag{2}$$

The recursive formula for $A_1(q)$ is:

$$A_1(q) = \begin{cases} 0 & \text{if } q = K-1 \\ A_1(q+1) + h(q+1) \cdot (q+1) & \text{if } 0 \leq q \leq K-2 \end{cases} \tag{3}$$

The recursive formula for $B_1(q)$ is:

$$B_1(q) = \begin{cases} 0 & \text{if } q = K - 1 \\ B_1(q+1) + h(q+1) \cdot (q+1)^2 & \text{if } 0 \leq q \leq K - 2 \end{cases} \tag{4}$$

1.4 Implementation

1.4.1 Algorithm: Minimum Error Thresholding

Algorithm 1 Minimum Error Thresholding

Require: $h : [0, K-1] \rightarrow \mathbb{N}$ ▷ Grayscale histogram

Output: Optimal threshold th for binary classification or -1 if not found

$K \leftarrow \text{size}(h)$
$(\sigma_0^2, \sigma_1^2, N) \leftarrow \text{MakeSigmaTable}(h, K)$
$n_0 \leftarrow 0$
$q_{\text{min}} \leftarrow -1$
$e_{\text{min}} \leftarrow \infty$
for $q \leftarrow 0$ to $K-2$ **do**
  $n_0 \leftarrow n_0 + h(q)$
  $n_1 \leftarrow N - n_0$
  if $n_0 > 0$ and $n_1 > 0$ **then**
    $P_0 \leftarrow \frac{n_0}{N}$
    $P_1 \leftarrow \frac{n_1}{N}$
    $e \leftarrow P_0 \cdot \ln(\sigma_0^2(q)) + P_1 \cdot \ln(\sigma_1^2(q)) - 2(P_0 \ln(P_0) + P_1 \ln(P_1))$
    if $e < e_{\text{min}}$ **then**
      $e_{\text{min}} \leftarrow e$
      $q_{\text{min}} \leftarrow q$
    end if
  end if
end for
return q_{min}

1.4.2 Algorithm: MakeSigmaTable

Note: The value of $\frac{1}{12}$ is added to avoid 0 when the corresponding class is homogeneous (e.g., 2 classes in a binary image). The variance of a uniform distribution in the unit interval is $\frac{1}{12}$.

Algorithm 2 MakeSigmaTable

Require: h, K

Output: Creates maps $\sigma_0^2, \sigma_1^2 : [0, K-1] \to \mathbb{R}$

$n_0 \leftarrow 0$

$A_0 \leftarrow 0$

$B_0 \leftarrow 0$

for $q \leftarrow 0$ to $K-1$ **do**

  $n_0 \leftarrow n_0 + h(q)$

  $A_0 \leftarrow A_0 + h(q) \cdot q$

  $$\sigma_0^2(q) \leftarrow \begin{cases} \frac{1}{12} + \left(B_0 - \frac{A_0^2}{n_0}\right) / n_0 & \text{if } n_0 > 0 \\ 0 & \text{otherwise} \end{cases}$$

end for

$N \leftarrow n_0$

$n_1 \leftarrow 0$

$A_1 \leftarrow 0$

$B_1 \leftarrow 0$

$\sigma_1^2(K-1) \leftarrow 0$

for $q \leftarrow K-2$ to 0 **do**

  $n_1 \leftarrow n_1 + h(q+1)$

  $A_1 \leftarrow A_1 + h(q+1) \cdot (q+1)$

  $B_1 \leftarrow B_1 + h(q+1) \cdot (q+1)^2$

  $$\sigma_1^2(q) \leftarrow \begin{cases} \frac{1}{12} + \left(B_1 - \frac{A_1^2}{n_1}\right) / n_1 & \text{if } n_1 > 0 \\ 0 & \text{otherwise} \end{cases}$$

end for

return $(\sigma_0^2, \sigma_1^2, N)$

Bibliography

[1] Wilhelm Burger, Mark J. Burge , Principles of Digital Image Processing: Advanced Methods, Chapter 2 - Automatic Thresholding, 2013