Upper Bounds for Bayes Error
slecture for ECE662

Jeehyun Choe

Purdue University

April 14, 2014
Chapter 1

Introduction

This lecture describes the theoretical upper bounds for Bayes Error. First, in chapter 2, the error bound is expressed in terms of the Bayes classifiers. This error bound expression includes a \( \min \) function that by using a lemma, the \( \min \) function can be replaced with the expression of a theoretical error bound, the Chernoff bound. In chapter 3, we specifically derive the Chernoff bound for the Normally distributed data. We also derive the Chernoff distance in the case of Normally distributed data. In section 3.2, some examples for the Chernoff bound are provided.

The materials given in this lecture are based on the lecture notes and the discussions that were shared in Prof. Boutin’s ECE662 Spring 2014 course at Purdue University. Examples were designed to reflect the theories taught in the class.
Chapter 2

Upper Bounds for Bayes Error

2.1 Classifier using Bayes rule

To classify the dataset of feature vectors with $C$ labels, we choose class $w_i$ where

$$\arg \max_{\omega_i \in \{\omega_1, \ldots, \omega_C\}} \text{Prob}(\omega_i | x). \quad (2.1)$$

However, it is difficult to directly estimate $\text{Prob}(\omega_i | x)$. So instead of solving eq. 2.1, we use the Bayes rule to change the problem to

$$\arg \max_{\omega_i \in \{\omega_1, \ldots, \omega_C\}} \rho(x | \omega_i) \text{Prob}(\omega_i). \quad (2.2)$$

As can be seen from eq. 2.2, we need to know all the distributions and priors of the classes in the dataset to apply the Bayesian classifier. If the distributions and priors in the dataset is all known, the Bayesian classifier is an optimal classifier since the decision taken following Bayes rule minimizes the probability of error.

2.2 Upper Bounds for Bayes Error

The expected value of the error when deciding following the Bayes rule can be computed as below.

$$E(error) = \int_{\mathbb{R}} \text{Prob}(error | x) \rho(x) dx$$

$$= \int_{\mathbb{R}} \min(\text{Prob}(\omega_1 | x), \text{Prob}(\omega_2 | x)) \rho(x) dx \quad (2.3)$$

$$= \int_{\mathbb{R}} \min(\rho(x | \omega_1) \text{Prob}(\omega_1), \rho(x | \omega_2) \text{Prob}(\omega_2)) dx \quad (2.4)$$

And we can obtain the upper bound for eq. 2.5 using the following lemma.

$$\min\{a, b\} \leq a^\beta b^{1-\beta} \quad (2.6)$$

where $\forall a, b \geq 0$ and $0 \leq \beta \leq 1$. Therefore, from eq. 2.5 and 2.6, we can derive the following error bound.

$$E(error) \leq \int_{\mathbb{R}} (\rho(x | \omega_1) \text{Prob}(\omega_1))^\beta (\rho(x | \omega_2) \text{Prob}(\omega_2))^{1-\beta} dx = \varepsilon_\beta \quad (2.7)$$
Eq. 2.7 is called Chernoff Bound. This is the error bound that can be computed given the full knowledge of probability of each class.

Since priors are independent of $x$, we can take priors out of the integral.

$$
\varepsilon_\beta = \text{Prob}(\omega_1)^\beta \text{Prob}(\omega_2)^{1-\beta} \int_{\mathbb{R}} \rho(x \mid \omega_1)^\beta (\rho(x \mid \omega_2))^{1-\beta} dx
$$

For the specific case where $\beta = \frac{1}{2}$, we call $\varepsilon_{\frac{1}{2}}$ as Bhattacharyya bound.

The smallest of such bound in eq. 2.8

$$
S := \min_{\beta \in [0,1]} (\varepsilon_\beta)
$$

is also sometimes called the Chernoff Bound. This can be used as a error bound since the Bayes error lies somewhere in between 0 and $\min (\varepsilon_\beta)$.

The Chernoff bound $\varepsilon_\beta$ is the probability of error as a function of $\beta$. It is usually used for two-category cases and if to be used for multi-category cases, the error bound can be split into several two-category cases and add up all to form the final error bound.
Chapter 3

Chernoff Bound for Normally Distributed Data

In this section, we will look at the upper bounds for Bayes error in 1-dimensional Normally distributed data.

Assume $\rho(x | \omega_1)$ and $\rho(x | \omega_2)$ are Gaussians with the distribution

$$
\rho(x | \omega_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[ -\frac{1}{2} \left( \frac{x-\mu_i}{\sigma_i} \right)^2 \right].
$$

(3.1)

Since Gaussian distribution is one of the exponential families, eq. 3.1 can be expressed as a following form [1] [3].

$$
\rho(x | \omega_i) = \exp \left( \langle t(x), \theta_i \rangle - F(\theta_i) + k(x) \right)
$$

(3.2)

where $t(x)$ is sufficient statistic and $\theta$ is parameters.

It is easier to derive the Chernoff Bound for Normally distributed data if we use the form in eq. 3.2. Therefore, we will first find the exponential form of the Gaussian distribution in eq. 3.1.

$$
\rho(x | \omega_i) = \exp \left( \frac{1}{2} \ln \left( \frac{1}{\sqrt{2\pi} \sigma_i} \right) - \frac{1}{2} \left( \frac{x^2}{\sigma_i^2} + \frac{x\mu_i}{\sigma_i^2} - \frac{\mu_i^2}{2\sigma_i^2} \right) - F(\theta_i) + k(x) \right)
$$

(3.3)

From eq. 3.4, we can set the sufficient statistic $t(x)$ and parameters $\theta_i$ as

$$
t(x) = (x, -x^2)
$$

(3.5)

$$
\theta_i = \left( \frac{\mu_i}{\sigma_i^2}, \frac{1}{2\sigma_i^2} \right).
$$

(3.6)

We can also express $F(\theta_i)$ and $k(x)$ as

$$
F(\theta_i) = \frac{\theta_i^2}{4\theta_i^2} - \frac{1}{2} \ln \theta_i^2 + \frac{1}{2} \ln \pi
$$

(3.7)
where \( \theta_{1,1} \) and \( \theta_{1,2} \) refer to the first and second components in eq. 3.6.

Now by substituting this exponential form of Gaussian densities into eq. 2.8, we can derive the simpler formula of Chernoff bound.

### 3.1 Derivation of Chernoff Distance

Let the integral part in eq. 2.8 be the Chernoff \( \beta \)-coefficient [3].

\[
\int_{\mathbb{R}} \rho(x \mid \omega_p)^{\beta} \rho(x \mid \omega_q)^{1-\beta} dx = \int_{\mathbb{R}} \left[ \exp \left( (t(x), \theta_p) - F(\theta_p) \right) \right]^{\beta} \left[ \exp \left( (t(x), \theta_q) - F(\theta_q) \right) \right]^{1-\beta} dx
\]

\[
= \int_{\mathbb{R}} \exp \left[ (t(x), \beta \theta_p) - \beta F(\theta_p) \right] \exp \left[ (t(x), (1-\beta) \theta_q) - (1-\beta) F(\theta_q) \right] dx
\]

\[
= \int_{\mathbb{R}} \exp \left( (t(x), \beta \theta_p) + (t(x), (1-\beta) \theta_q) - \beta F(\theta_p) - (1-\beta) F(\theta_q) \right) dx
\]

\[
= \exp \left( -\beta F(\theta_p) - (1-\beta) F(\theta_q) \right) \int_{\mathbb{R}} \exp \left( (t(x), \beta \theta_p + (1-\beta) \theta_q) \right) dx
\]

\[
= \exp \left( -\beta F(\theta_p) - (1-\beta) F(\theta_q) \right) \int_{\mathbb{R}} \exp \left( -F(\beta \theta_p + (1-\beta) \theta_q) \right) dx
\]

\[
= \exp \left( -\beta F(\theta_p) - (1-\beta) F(\theta_q) + F(\beta \theta_p + (1-\beta) \theta_q) \right)
\]

We can see that eq. 3.11 is in the form \( e^{-k(\beta)} \). \( k(\beta) \) is sometimes called Chernoff distance. To obtain the formula for \( -k(\beta) \), we can substitute \( F() \) terms in eq. 3.11 with the expression in terms of \( \mu_i \) or \( \sigma_i \) using eq. 3.6 and eq. 3.7.

\[
\begin{align*}
\beta F(\theta_p) &= \beta \frac{\mu_p^2}{2\sigma_p^2} - \beta \frac{1}{2} \ln \left( \frac{1}{2\sigma_p^2} \right) + \beta \frac{1}{2} \ln \pi \\
(1 - \beta) F(\theta_q) &= (1 - \beta) \frac{\mu_q^2}{2\sigma_q^2} - (1 - \beta) \frac{1}{2} \ln \left( \frac{1}{2\sigma_q^2} \right) + (1 - \beta) \frac{1}{2} \ln \pi \\
-F(\beta \theta_p + (1-\beta) \theta_q) &= -F(\theta_{tot})
\end{align*}
\]

The last line in eq. 3.12 can be computed by first finding the \( \theta_{tot} \) as

\[
\theta_{tot} = \beta \left( \frac{\mu_p}{\sigma_p}, \frac{1}{2\sigma_p^2} \right) + (1 - \beta) \left( \frac{\mu_q}{\sigma_q}, \frac{1}{2\sigma_q^2} \right) = \left( \frac{\beta \mu_p \sigma_q^2 + (1 - \beta) \mu_q \sigma_p^2}{\sigma_p^2 \sigma_q^2}, \frac{\beta \sigma_q^2 + (1 - \beta) \sigma_p^2}{2\sigma_p^2 \sigma_q^2} \right)
\]

and substituting this result into the eq. 3.7 as follows:

\[
-F(\theta_{tot}) = -\frac{[\beta \mu_p \sigma_q^2 + (1 - \beta) \mu_q \sigma_p^2]^2}{2\sigma_p^2 \sigma_q^2 (\beta \sigma_q^2 + (1 - \beta) \sigma_p^2)} + \frac{1}{2} \ln \frac{\beta \sigma_q^2 + (1 - \beta) \sigma_p^2}{(\beta \sigma_q^2 + (1 - \beta) \sigma_p^2)} - \frac{1}{2} \ln \pi.
\]

Therefore, \( k(\beta) \) can be derived as follows:

\[
k(\beta) = \beta F(\theta_p) + (1 - \beta) F(\theta_q) - F(\theta_{tot})
\]

\[
= \beta \frac{\mu_p^2}{2\sigma_p^2} + (1 - \beta) \frac{\mu_q^2}{2\sigma_q^2} - \frac{[\beta \mu_p \sigma_q^2 + (1 - \beta) \mu_q \sigma_p^2]^2}{2\sigma_p^2 \sigma_q^2 (\beta \sigma_q^2 + (1 - \beta) \sigma_p^2)} + \frac{1}{2} \ln \left( \frac{\beta \sigma_q^2 + (1 - \beta) \sigma_p^2}{(\beta \sigma_q^2 + (1 - \beta) \sigma_p^2)} \right)
\]
\[ k(\beta) = \beta (1 - \beta) (\mu_p - \mu_q)^2 + \left( \frac{1}{2} \ln \left[ \frac{\beta \sigma_q^2 + (1 - \beta) \sigma_p^2}{\sigma_p^{2(1-\beta)} \sigma_q^{2\beta}} \right] \right). \] (3.14)

Therefore, the error bound would be

\[ \varepsilon_\beta = \text{Prob}(\omega_1)^\beta \text{Prob}(\omega_2)^{1-\beta} \exp \left[ -k(\beta) \right]. \] (3.15)

### 3.2 Example

By using the closed form equation of the Chernoff bound shown in eq. 3.15 and eq. 3.14, we can visualize several examples of Chernoff bounds for Normally distributed data.

![The Chernoff bound with various priors](image)

**Figure 3.1:** Chernoff bound for \(\text{N}(0,1)\) and \(\text{N}(5,1)\)

In Figure 3.1, we can see that the Chernoff bound is generally low, since the gap between the two means is high compared to the variances. The Chernoff bound doesn’t change much as the prior changes from 0.1 to 0.9 in this case.

Figure 3.2 shows the case where the gap between two means is low compared to the variances. In this case, as the prior for one category gets larger than the other prior, the Chernoff bound goes down. However, when both priors have similar values, the Chernoff bound gets higher.

We can observe that both in Figure 3.1 and Figure 3.2, the Chernoff bound is close to the Bhattacharyya bound when the Priors have the equal values.

Figure 3.3 and Figure 3.4 show the cases when the variances of two categories differ. In previous examples where the variance for both categories are the same, the Chernoff bound was close to the Bhattacharyya bound when the Priors have the equal values. However, Figure 3.3 and Figure 3.4 don’t show such patterns.
The Chernoff bound with various priors
Normal distributions of $N(0, 1)$ with $P_1$ and $N(1, 1)$ with $(1-P_1)$

- $P_1 = 0.10$ ($E_{min} = 0.1000$, $\beta_{min} = 1.00$)
- $P_1 = 0.20$ ($E_{min} = 0.2000$, $\beta_{min} = 1.00$)
- $P_1 = 0.30$ ($E_{min} = 0.3000$, $\beta_{min} = 1.00$)
- $P_1 = 0.40$ ($E_{min} = 0.3982$, $\beta_{min} = 0.91$)
- $P_1 = 0.50$ ($E_{min} = 0.4413$, $\beta_{min} = 0.49$)
- $P_1 = 0.60$ ($E_{min} = 0.3982$, $\beta_{min} = 0.09$)
- $P_1 = 0.70$ ($E_{min} = 0.3000$, $\beta_{min} = 0.00$)
- $P_1 = 0.80$ ($E_{min} = 0.2000$, $\beta_{min} = 0.00$)
- $P_1 = 0.90$ ($E_{min} = 0.1000$, $\beta_{min} = 0.00$)

Figure 3.2: Chernoff bound for $N(0,1)$ and $N(1,1)$

The Chernoff bound with various priors
Normal distributions of $N(0, 1)$ with $P_1$ and $N(5, 2)$ with $(1-P_1)$

- $P_1 = 0.10$ ($E_{min} = 0.0747$, $\beta_{min} = 0.60$)
- $P_1 = 0.20$ ($E_{min} = 0.1022$, $\beta_{min} = 0.47$)
- $P_1 = 0.30$ ($E_{min} = 0.1134$, $\beta_{min} = 0.41$)
- $P_1 = 0.40$ ($E_{min} = 0.1155$, $\beta_{min} = 0.37$)
- $P_1 = 0.50$ ($E_{min} = 0.1111$, $\beta_{min} = 0.33$)
- $P_1 = 0.60$ ($E_{min} = 0.1013$, $\beta_{min} = 0.31$)
- $P_1 = 0.70$ ($E_{min} = 0.0865$, $\beta_{min} = 0.28$)
- $P_1 = 0.80$ ($E_{min} = 0.0666$, $\beta_{min} = 0.25$)
- $P_1 = 0.90$ ($E_{min} = 0.0403$, $\beta_{min} = 0.21$)

Figure 3.3: Chernoff bound for $N(0,1)$ and $N(5,2)$
The Chernoff bound with various priors
Normal distributions of N(0, 1) with $P_1$ and N(1, 2) with $(1-P_1)$

<table>
<thead>
<tr>
<th>$P_1$</th>
<th>$E_{\text{min}}$</th>
<th>$\beta_{\text{min}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.1000</td>
<td>1.00</td>
</tr>
<tr>
<td>0.20</td>
<td>0.2000</td>
<td>1.00</td>
</tr>
<tr>
<td>0.30</td>
<td>0.3000</td>
<td>1.00</td>
</tr>
<tr>
<td>0.40</td>
<td>0.3992</td>
<td>0.90</td>
</tr>
<tr>
<td>0.50</td>
<td>0.4209</td>
<td>0.37</td>
</tr>
<tr>
<td>0.60</td>
<td>0.3744</td>
<td>0.17</td>
</tr>
<tr>
<td>0.70</td>
<td>0.2957</td>
<td>0.07</td>
</tr>
<tr>
<td>0.80</td>
<td>0.2000</td>
<td>0.00</td>
</tr>
<tr>
<td>0.90</td>
<td>0.1000</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Figure 3.4: Chernoff bound for N(0,1) and N(1,2)
Chapter 4

Summary

In this lecture, we have shown how the Chernoff bound can be derived. We have derived the Chernoff bound in a closed form for the 1D Gaussian distributions. We also saw some examples of the Chernoff bounds on different sets of Normally distributed 1-dimensional data.
Bibliography

