18794 Pattern Recognition Theory
Prof. Marios Savvides

HMW-Alexander

November 9, 2016

Contents

---

1 Introduction

---

2 Decision Theory

2.1 Terms

- Feature Space:
  - **Feature**: a distinctive characteristic or quality of the object
  - **Feature vector**: combine more than one feature as a vector
  - **Feature space**: The space defined by the feature vectors
- Classifiers:
  - **Decision regions**: a classifier partitions the feature space into class-corresponding decision regions.
  - **Decision boundaries**: the borders between the decision regions.

2.2 Bayes Rule

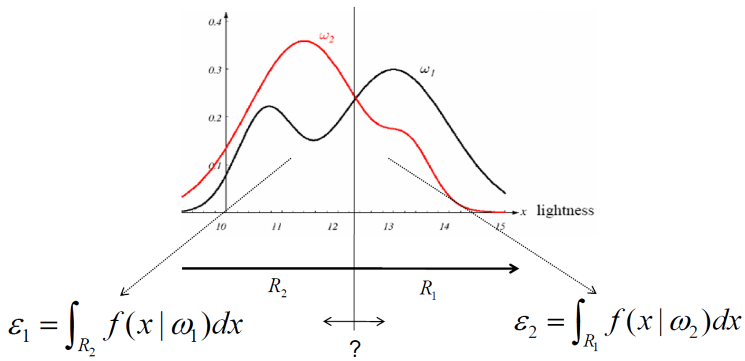
$$P(w_i|x) = \frac{P(x, w_i)}{P(x)} = \frac{P(x|w_i)P(w_i)}{\sum_{k=1}^{C} P(x|w_k)P(w_k)} \tag{1}$$

- **Posterior Probability** $P(w_i|x)$: the conditional probability of correct class being w_i given that feature value x has been observed.
- **Evidence** $P(x)$: the total probability of observing the feature value of x.
- **Likelihood** $P(x|w_i)$: the conditional probability of observing a feature value of x given that the correct class is w_i.

- **Prior Probability** $P(w_i)$: the probability of class w_i, $\sum_{k=1}^{C} P(w_k) = 1$.
- **Bayes Classifiers** decide on the class that has the **largest posterior probability** ($\max_{w_i} P(w_i|x)$). They are statistically the best classifiers i.e. they are minimum error classifiers (optimal).

2.3 Minimum Probability of Error

- $\epsilon = P(error|class)$: probability of assigning x to the wrong class w.
- $P_e = \sum_{k=1}^{C} P(w_k)\epsilon_k$: total probability of error.



For the two class case shown above, we want to minimize P_e as below:

$$\begin{aligned} P_e &= P(w_1)\epsilon_1 + P(w_2)\epsilon_2 \\ &= P(w_1)\int_{R_2} f(x|w_1)dx + P(w_2)\int_{R_1} f(x|w_2)dx \\ &= P(w_1)(1 - \int_{R_1} f(x|w_1)dx) + P(w_2)\int_{R_1} f(x|w_2)dx \\ &= P(w_1) + \int_{R_1} (P(w_2)f(x|w_2) - P(w_1)f(x|w_1))dx \end{aligned} \tag{2}$$

To minimize P_e, we want $P(w_2)f(x|w_2) - P(w_1)f(x|w_1)$ to be always negative(< 0) in the region R_1:

$$\begin{aligned} P(w_1)f(x|w_1) - P(w_2)f(x|w_2) > 0 \quad &\Rightarrow \quad w_1 \\ P(w_1)f(x|w_1) - P(w_2)f(x|w_2) < 0 \quad &\Rightarrow \quad w_2 \end{aligned} \tag{3}$$

2.4 Likelihood Ratio

- **Likelihood ratio**: $l(x) = \frac{f(x|w_1)}{f(x|w_2)}$
- **Log likelihood ratio**: $ln(l(x)) = ln(\frac{f(x|w_1)}{f(x|w_2)}) = ln(f(x|w_1)) - ln(f(x|w_2))$
- **Ratio of a priori probabilities**: $T = \frac{P(w_2)}{P(w_1)}$
- **Log ratio of a priori probabilities**: $ln(T) = ln(\frac{P(w_2)}{P(w_1)}) = ln(P(w_2)) - ln(P(w_1))$

$$\begin{aligned} ln(l(x)) = ln(\tfrac{f(x|w_1)}{f(x|w_2)}) > ln(\tfrac{P(w_2)}{P(w_1)}) = ln(T) \quad &\Rightarrow \quad w_1 \\ ln(l(x)) = ln(\tfrac{f(x|w_1)}{f(x|w_2)}) < ln(\tfrac{P(w_2)}{P(w_1)}) = ln(T) \quad &\Rightarrow \quad w_2 \end{aligned} \tag{4}$$

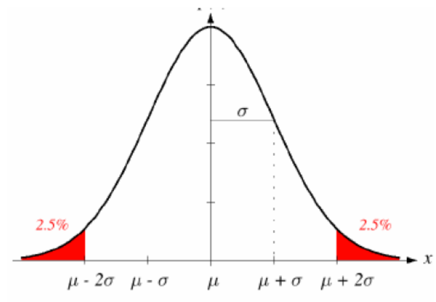
2.4.1 Likelihood as Gaussian distribution

Assume likelihood $f(x|w_i)$ are Gaussian distributions with mean μ_i and variance σ_i^2.

$$f(x|w_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right) \tag{5}$$

The log likelihood ratio:

$$f(x \mid \omega_i)$$



$$\begin{aligned} ln(l(x)) &= ln\left(\frac{\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right)}{\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right)}\right) \\ &= ln(\tfrac{\sigma_2}{\sigma_1}) + \tfrac{(x-\mu_2)^2}{2\sigma_2^2} - \tfrac{(x-\mu_1)^2}{2\sigma_1^2} \end{aligned} \tag{6}$$

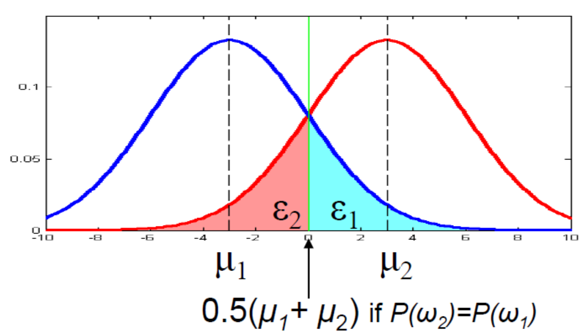
Case: $\sigma_1 = \sigma_2 = \sigma$

$$ln(l(x)) = \tfrac{2x(\mu_1-\mu_2)-(\mu_1^2-\mu_2^2)}{2\sigma^2} \tag{7}$$

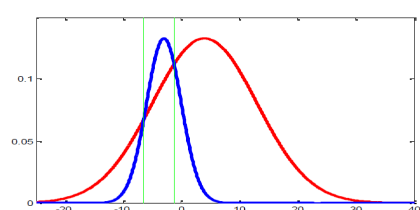
$$\begin{aligned} x(\mu_1-\mu_2) - \tfrac{\mu_1^2-\mu_2^2}{2} > \sigma^2 ln(\tfrac{P(w_2)}{P(w_1)}) &\quad \Rightarrow \quad w_1 \\ x(\mu_1-\mu_2) - \tfrac{\mu_1^2-\mu_2^2}{2} < \sigma^2 ln(\tfrac{P(w_2)}{P(w_1)}) &\quad \Rightarrow \quad w_2 \end{aligned} \tag{8}$$

If $P(w_1) = P(w_2)$

$$x = \frac{\mu_1 + \mu_2}{2} \tag{9}$$



Case: $\sigma_1 \neq \sigma_2$



$$\ln\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2 - \frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$

$$ax^2 + bx + c \gtrless 0$$

$$(x - x_1)(x - x_2) \gtrless 0$$