(a) Theoretical Error Rate for Classification(a) 理论分类错误率

The theoretical error rate in a classification problem can be expressed as the probability of misclassifying a sample. The optimal classification decision is made by selecting the class ω_i that maximizes the posterior probability $p(\omega_i|\mathbf{x})$, according to **Bayes' Theorem**:

分类问题中的理论错误率可以表示为错误分类样本的概率。通过选择类别来做出最佳分类决策 ω_i 最大化后验概率 $p(\omega_i|\mathbf{x})$ ，根据**贝叶斯定理**：

$$p(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)p(\omega_i)}{p(\mathbf{x})}$$

Here:

- $p(\omega_i)$ is the prior probability of class ω_i,$p(\omega_i)$ 是类别的先验概率 ω_i ，
- $p(\mathbf{x}|\omega_i)$ is the class-conditional probability density function for class ω_i,

  $p(\mathbf{x}|\omega_i)$ 是类的类条件概率密度函数 ω_i ，
- $p(\mathbf{x})$ is the total probability of observing $\mathbf{x}$, and$p(\mathbf{x})$ 是观察到的总概率 $\mathbf{x}$ ， 和
- $p(\omega_i|\mathbf{x})$ is the posterior probability of class ω_i given the input sample $\mathbf{x}$.

  $p(\omega_i|\mathbf{x})$ 是类别的后验概率 ω_i 给定输入样本 $\mathbf{x}$ 。

To calculate the **theoretical error rate**, we first define the decision rule: we classify the input sample $\mathbf{x}$ into class ω_i if $p(\omega_i|\mathbf{x})$ is the highest among all classes. The error occurs when $\mathbf{x}$ is classified into the wrong class, i.e., when the true class of $\mathbf{x}$ is not the one with the highest posterior probability.

为了计算**理论错误率**，我们首先定义决策规则：我们对输入样本进行分类 $\mathbf{x}$ 进入课堂 ω_i 如果 $p(\omega_i|\mathbf{x})$ 是所有班级中最高的。错误发生时 $\mathbf{x}$ 被分类到错误的类别，即，当真正的类别 $\mathbf{x}$ 不是后验概率最高的那个。

Let's define the **theoretical error rate** P_e as the probability of misclassification. The error occurs when the input sample $\mathbf{x}$ is assigned to a class ω_j where $j \neq i$, and the true class is ω_i. The error rate is thus given by:

让我们定义**理论错误率** P_e 作为错误分类的概率。当输入样本时出现错误 $\mathbf{x}$ 被分配到一个班级 ω_j 在哪里 $j \neq i$ ，真实的类别是 ω_i 。因此，错误率由下式给出：

$$P_e = 1 - \int_{\mathbf{x}} \max_i \left(p(\omega_i|\mathbf{x})\right) p(\mathbf{x}) d\mathbf{x}$$

This expression captures the idea that the theoretical error rate is the complement of the probability of making the correct classification, which is determined by the maximum posterior probability across all classes.

该表达式体现了这样的思想：理论错误率是正确分类概率的补充，而正确分类概率由所有类别的最大后验概率决定。

Answer the question (b)回答问题（b）

(b) Decision Rule that Minimizes the Theoretical Error Rate

(b) 最小化理论错误率的决策规则

To minimize the theoretical error rate, we need to derive the decision rule based on **Bayes' decision theory**. This is done by classifying the input sample $\mathbf{x}$ to the class ω_i that maximizes the posterior probability $p(\omega_i|\mathbf{x})$.

为了最小化理论错误率，我们需要基于**贝叶斯决策理论**推导决策规则。这是通过对输入样本进行分类来完成的 $\mathbf{x}$ 到班级 ω_i 最大化后验概率 $p(\omega_i|\mathbf{x})$ 。

Using **Bayes' Theorem**, the posterior probability $p(\omega_i|\mathbf{x})$ can be written as:

使用**贝叶斯定理**，后验概率 $p(\omega_i|\mathbf{x})$ 可以写成：

$$p(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)p(\omega_i)}{p(\mathbf{x})}$$

where:

- $p(\omega_i)$ is the prior probability of class ω_i,$p(\omega_i)$ 是类别的先验概率 ω_i ，
- $p(\mathbf{x}|\omega_i)$ is the class-conditional probability density function for class ω_i,

  $p(\mathbf{x}|\omega_i)$ 是类的类条件概率密度函数 ω_i ，
- $p(\mathbf{x}) = \sum_{i=1}^{c} p(\mathbf{x}|\omega_i)p(\omega_i)$ is the total probability of observing $\mathbf{x}$.

  $p(\mathbf{x}) = \sum_{i=1}^{c} p(\mathbf{x}|\omega_i)p(\omega_i)$ 是观察到的总概率 $\mathbf{x}$ 。

The denominator $p(\mathbf{x})$ is the same for all classes, so it does not affect the decision rule. Thus, the decision rule that minimizes the theoretical error rate is based on the numerator:

分母 $p(\mathbf{x})$ 对于所有类别都是相同的，因此它不会影响决策规则。因此，最小化理论错误率的决策规则基于分子：

$$\text{Assign } \mathbf{x} \text{ to class } \omega_i \text{ if } p(\mathbf{x}|\omega_i)p(\omega_i) > p(\mathbf{x}|\omega_j)p(\omega_j) \text{ for all } j \neq i$$

or equivalently,或同等地，

$$\text{Assign } \mathbf{x} \text{ to class } \omega_i \text{ if } p(\mathbf{x}|\omega_i)p(\omega_i) \text{ is maximum among all classes.}$$

This is known as the **Bayes decision rule** or **Maximum a Posteriori (MAP) rule**.

这称为**贝叶斯决策规则**或**最大后验 (MAP) 规则。**

In summary, the decision rule is:综上，决策规则为：

$$\hat{\omega}(\mathbf{x}) = \arg\max_i \left(p(\mathbf{x}|\omega_i)p(\omega_i)\right)$$

This rule assigns the sample $\mathbf{x}$ to the class ω_i that maximizes the product of the prior probability $p(\omega_i)$ and the class-conditional probability density $p(\mathbf{x}|\omega_i)$. This minimizes the theoretical error rate.

该规则分配样本 $\mathbf{x}$ 到班级 ω_i 最大化先验概率的乘积 $p(\omega_i)$ 和类条件概率密度 $p(\mathbf{x}|\omega_i)$ 。这最大限度地减少了理论错误率。

Answer the question (c)回答问题(c)

(c) Using Training Samples to Estimate $p(\omega_i)$, $p(\mathbf{x}|\omega_i)$, and $p(\mathbf{x})$ Leading to the k-Nearest Neighbor Classifier

(c) 使用训练样本进行估计 $p(\omega_i)$, $p(\mathbf{x}|\omega_i)$ ， 和 $p(\mathbf{x})$ 导致 k -最近邻分类器

In practice, the probabilities $p(\omega_i)$, $p(\mathbf{x}|\omega_i)$, and $p(\mathbf{x})$ are typically not known beforehand and need to be estimated from training data. Here's how we can estimate these probabilities using training samples:

在实践中，概率 $p(\omega_i)$, $p(\mathbf{x}|\omega_i)$ ， 和 $p(\mathbf{x})$ 通常事先不知道，需要根据训练数据进行估计。以下是我们如何使用训练样本来估计这些概率：

1. **Estimating $p(\omega_i)$ (Class Prior Probability):估计 $p(\omega_i)$ （类别先验概率）：**

   The prior probability $p(\omega_i)$ represents the likelihood that any random sample belongs to class ω_i. It can be estimated using the **relative frequency** of the training samples for each class:

   先验概率 $p(\omega_i)$ 表示任何随机样本属于该类的可能性 ω_i 。可以使用每个类别的训练样本的**相对频率**来估计：

   $$p(\omega_i) = \frac{N_i}{N}$$

   where:

   - N_i is the number of training samples in class ω_i,N_i 是班级训练样本的数量 ω_i ，
   - N is the total number of training samples.N 是训练样本的总数。

   This gives a simple estimate of the prior probability based on the proportion of samples in each class.这根据每个类别中的样本比例给出了先验概率的简单估计。

2. **Estimating $p(\mathbf{x}|\omega_i)$ (Class-Conditional Probability Density):**

   估计 $p(\mathbf{x}|\omega_i)$ （类条件概率密度）：

   The class-conditional probability density function $p(\mathbf{x}|\omega_i)$ represents the likelihood of observing a sample $\mathbf{x}$ given that it belongs to class ω_i. Estimating $p(\mathbf{x}|\omega_i)$ directly can be difficult, especially in high-dimensional feature spaces. In the absence of explicit parametric models for the data, one can use **non-parametric methods** such as the k-**Nearest Neighbor (k-NN)** approach.

   类条件概率密度函数 $p(\mathbf{x}|\omega_i)$ 表示观察样本的可能性 $\mathbf{x}$ 鉴于它属于类 ω_i 。估计 $p(\mathbf{x}|\omega_i)$ 直接地可能很困难，特别是在高维特征空间中。在缺乏明确的数据参数模型的情况下，可以使用**非参数方法**，例如 k - **最近邻（k-NN）**方法。

   In the k-NN classifier, the probability $p(\mathbf{x}|\omega_i)$ is approximated based on the density of training samples near $\mathbf{x}$. For a given sample $\mathbf{x}$, the k-NN method finds the k nearest neighbors in the training data (using a distance metric like Euclidean distance). The fraction of those neighbors that belong to class ω_i is used as an estimate of the class-conditional probability density:

   在 k -NN分类器，概率 $p(\mathbf{x}|\omega_i)$ 根据附近训练样本的密度进行近似 $\mathbf{x}$ 。对于给定的样本 $\mathbf{x}$ ，这 k -NN 方法找到 k 训练数据中的最近邻居（使用欧几里得距离等距离度量）。属于该类的邻居的比例 ω_i 用作类条件概率密度的估计：

   $$p(\mathbf{x}|\omega_i) \propto \frac{\text{Number of neighbors in class } \omega_i}{k}$$

   Essentially, the k-NN method approximates the probability distribution by considering how many of the k nearest samples belong to the class ω_i.

   本质上， k -NN 方法通过考虑有多少个来近似概率分布 k 最近的样本属于该类 ω_i 。

3. **Estimating $p(\mathbf{x})$ (Total Probability of $\mathbf{x}$):估计 $p(\mathbf{x})$ （总概率 $\mathbf{x}$ ）：**

   The total probability of observing the sample $\mathbf{x}$, $p(\mathbf{x})$, can be expressed as a sum over the class-conditional probabilities weighted by the prior probabilities:

   观察样本的总概率 $\mathbf{x}$, $p(\mathbf{x})$ ，可以表示为先验概率加权的类条件概率的总和：

   $$p(\mathbf{x}) = \sum_{i=1}^{c} p(\mathbf{x}|\omega_i)p(\omega_i)$$

   However, in the context of classification, we typically don't need to estimate $p(\mathbf{x})$ explicitly because it is the same for all classes and thus cancels out in the decision rule.

   然而，在分类的背景下，我们通常不需要估计 $p(\mathbf{x})$ 明确地，因为它对于所有类都是相同的，因此在决策规则中被取消。

k-Nearest Neighbor Classifier and Decision Rule

k -最近邻分类器和决策规则

In the k-NN classifier, we classify a new input sample $\mathbf{x}$ by finding the k closest training samples (neighbors) and assigning $\mathbf{x}$ to the class that appears most frequently among those neighbors. This directly corresponds to a non-parametric estimate of the class posterior probability $p(\omega_i|\mathbf{x})$, which can be written as:

在 k -NN分类器，我们对新的输入样本进行分类 $\mathbf{x}$ 通过找到 k 最接近的训练样本（邻居）并分配 $\mathbf{x}$ 到这些邻居中出现最频繁的类别。这直接对应于类后验概率的非参数估计 $p(\omega_i|\mathbf{x})$ ，可以写成：

$$p(\omega_i|\mathbf{x}) \approx \frac{\text{Number of neighbors in class } \omega_i}{k}$$

The **decision rule** for the k-NN classifier is then:**决策规则**为 k -NN分类器则为：

$$\hat{\omega}(\mathbf{x}) = \arg\max_i \left(\frac{\text{Number of neighbors in class } \omega_i}{k}\right)$$

This means we classify $\mathbf{x}$ into the class that has the highest number of neighbors among the k nearest samples. The k-NN method does not require explicit estimation of the probability density functions $p(\mathbf{x}|\omega_i)$ or $p(\mathbf{x})$; instead, it makes decisions based on local neighborhood information in the feature space.

这意味着我们分类 $\mathbf{x}$ 进入邻居数量最多的班级 k 最近的样本。这 k -NN方法不需要显式估计概率密度函数 $p(\mathbf{x}|\omega_i)$ 或者 $p(\mathbf{x})$;相反，它根据特征空间中的局部邻域信息做出决策。

Summary of Estimation and Decision Rule:估计和决策规则总结：

- **Estimate $p(\omega_i)$:** Use the relative frequency of the class in the training set.

  估计 $p(\omega_i)$ ： 使用训练集中类别的相对频率。
- **Estimate $p(\mathbf{x}|\omega_i)$:** Use the proportion of the k-nearest neighbors that belong to class ω_i

  .**估计 $p(\mathbf{x}|\omega_i)$ ：** 使用比例 k -属于类的最近邻居 ω_i 。
- **Decision Rule:** Assign $\mathbf{x}$ to the class that appears most frequently among its k-nearest neighbors, which minimizes the theoretical error rate without the need for explicit probability density estimates.

  决策规则： 分配 $\mathbf{x}$ 到其中出现最频繁的类 k - 最近邻，它可以最大限度地减少理论错误率，而不需要明确的概率密度估计。

The k-NN classifier is a simple yet effective non-parametric approach that indirectly estimates these probabilities and minimizes classification error by relying on the local structure of the training data.

这 k -NN分类器是一种简单而有效的非参数方法，它通过依赖训练数据的局部结构来间接估计这些概率并最小化分类误差。