3. Let $p_{\omega_i}(\omega_i)$ be the prior probability of class ω_i and $p_{\omega_i}(\omega_i|\boldsymbol{x})$ be the posterior probability of class ω_i. Let $p_x(\boldsymbol{x}|\omega_i)$ be the class-conditional probability density function and $p_x(\boldsymbol{x})$ be the probability density function over all classes. For symbolic simplicity, the subscripts of p_{ω_i} and p_x are omitted in the following, where it is clear from the context, which probability density function is referred to.

(a) If you know the value of the data $\boldsymbol{x}$ and make a decision that $\boldsymbol{x}$ belongs to a class ω_i, find the probability of the wrong decision $p(e_i|\boldsymbol{x})$ and hence derive the decision rule that minimizes $p(e_i|\boldsymbol{x})$, using one of the above quantities.

(5 Marks)

Note: Question No. 3 continues on page 3.

(b) Suppose none of the above quantities is available but we can use the available training data to estimate $p(\omega_i)$, $p(\boldsymbol{x}|\omega_i)$ and $p(\boldsymbol{x})$, derive the decision rule that minimizes the probability of the wrong decision so that you can make a decision based on the estimated $p(\omega_i)$ and $p(\boldsymbol{x}|\omega_i)$.

(5 Marks)

(c) Estimate $p(\omega_i)$, $p(\boldsymbol{x}|\omega_i)$ and $p(\boldsymbol{x})$ so that the decision rule that minimizes the probability of the wrong decision leads to the k-nearest neighbor classifier.

22S1 Q3

Solution

(a) $p(e_i|x) = 1 - \max_i P_{w_i}(w_i|x)$

max

(b) $\boxed{p(w_i|x) = \frac{p(x|w_i)\,p(w_i)}{p(x)}}$ X

(c) find k nearest training sample to x

the count the frequency of each class among the k samples. Then choose the class with the most neighbors as the predicted class for x

6222 2022-2023

3 (a) know: $p(x)$ $p(w_i|x)$ target: $p(e_i|x)$

~~p(e~~ $p(e_i|x) = 1 - p(w_i|x)$ $p(w_i|x) = \frac{p(x|w_i)p(w_i)}{p(x)}$

(b) minimize ~~the~~ $p(e_i|x) \Rightarrow$ maximize $p(w_i|x)$

$p(x)$ is constant $\Longrightarrow$ maximize $p(x|w_i) \cdot p(w_i)$

(c) suppose that we place a cell of volumn V around x and capture k samples, k_i of which turn out to be labeled w_i.

n labeled samples

$\hat{p}(x) = \frac{k}{nV(k)}$ 4分

感觉就是p（x）=k/nV

$\hat{p}(w_i) = \frac{n_i}{n}$

p(wi)=ni/n

$\hat{p}(x|w_i) = \frac{k_i}{n_i V(k)}$

p(x|wi)=ki/niV

$\hat{p}(e_i|x) = 1 - \frac{p(x|w_i)p(w_i)}{p(x)}$

几个式子一列最后再写个
p(wi|x)=ki/k

$= 1 - \frac{k_i}{\cancel{n_i}\cancel{V(k)}} \cdot \frac{\cancel{n_i}}{\cancel{n}} \cdot \frac{\cancel{n}\cancel{V(k)}}{k}$

$= 1 - \frac{k_i}{k}$

(a)

问题要求：已知数据 x 的值，做出决策认为 x 属于某个类别 ω_i，计算错误决策的概率 $p(e_i|x)$，并推导出最小化 $p(e_i|x)$ 的决策规则。

解答步骤：

1. **后验概率定义**：
   - $p(\omega_i|x) = \frac{p(x|\omega_i)p(\omega_i)}{p(x)}$
   - 这里，$p(x) = \sum_{i=1}^{C} p(x|\omega_i)p(\omega_i)$，其中 C 是类别的总数。
2. **错误概率的定义**：
   - 假设我们选择 ω_i 为类别，那么错误概率为：

     $$p(e_i|x) = 1 - p(\omega_i|x)$$

   - 即错误概率等于不属于 ω_i 的概率总和。
3. **最小化错误概率的决策规则**： $\underset{w_i}{\operatorname{argmax}}$
   - 为了最小化错误概率，我们应该选择具有最大后验概率的类别：

     $$\omega^* = \arg\max_{\omega_i} p(\omega_i|x)$$

   - 这被称为贝叶斯决策规则，即最大后验概率决策 (Maximum A Posteriori, MAP)。 $=\max_{i}$

(b)

问题要求：假设没有已知量，但可以通过训练数据估计 $p(\omega_i)$、$p(x|\omega_i)$ 和 $p(x)$。推导出最小化错误决策概率的决策规则。

解答步骤：

1. **估计先验概率** $p(\omega_i)$：
   - 可以通过训练数据中类别 ω_i 的样本数量 N_i 来估计：

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

   其中 N 为总样本数。
2. **估计条件概率** $p(x|\omega_i)$：
   - 通过类别 ω_i 的训练样本估计 x 在类别 ω_i 中的分布。
3. **估计** $p(x)$：
   - $p(x) = \sum_{i=1}^{C} p(x|\omega_i)p(\omega_i)$。
4. **决策规则**：
   - 依然使用贝叶斯决策规则，即选择具有最大后验概率的类别：

     $$\omega^* = \arg\max_{\omega_i} p(\omega_i|x)$$

   - 这意味着我们需要使用估计的 $p(\omega_i|x)$ 进行决策。

(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 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。