2.2.3.2 Naive Bayes Classification

The problem with Kernel density estimation is that it performs very poorly in high dimensions. While one can calculate 2.2.3.1 Bayes Classification > 2.2.12 for $x \in \mathbb{R}^d$ by using a non-negative function $K_\lambda : \mathbb{R}^d \to \mathbb{R}$ with $\int_{\mathbb{R}^d} K_\lambda(x)dy = 1$, an accurate estimation requires an extremely large number of data points, meaning N must be very large.
Another problem that we cannot address with the previous methods is the approximation of the conditional probabilities $\mathbb{P}(x|K_x)$ **for mixed variables** $x \in \mathbb{R}^d$, e.g. with qualitative (color) and quantitative (weight) component.
An alternative approach that circumvents both difficulties is naive Bayes classification, which operates under the **strong assumption** that the individual features x_{ij} for $j = 1, \dots, d$ for all data points x_i are independent within a class. In this case, we can use Bayes' theorem to get:

$$\mathbb{P}(K_c|x) = \frac{\mathbb{P}(K_c)\mathbb{P}(x|K_c)}{\sum_{i=1}^{C} \mathbb{P}(K_i)\mathbb{P}(x|K_i)} = \frac{\pi_c \prod_{j=1}^{d} \mathbb{P}(x_j|K_c)}{\sum_{i=1}^{C} \pi_i \prod_{j=1}^{d} \mathbb{P}(x_j|K_i)}$$

remember that π_c was considered as the probability $\mathbb{P}(K_c)$ and the estimator was $\hat{\pi}_c := \frac{\#\{i=1,\dots,N|y_i=l_c\}}{N}$

or for densities

$$\mathbb{P}(K_c|x) = \frac{\pi_c \prod_{j=1}^{d} p(x_j|K_c)}{\sum_{i=1}^{C} \pi_i \prod_{j=1}^{d} p(x_j|K_i)}, \quad x \in \mathbb{R}^d$$

Thus, in this case although $x \in \mathbb{R}^d$, we only need to approximate one-dimensional probabilities or densities, namely $\mathbb{P}(x_j|K_c)$ or $p(x_j|K_c)$ for $j = 1, \dots, d$. This can be achieved using one of the three approaches discussed above. The approximated naive Bayes classifier then becomes

$$x \mapsto \arg\max_{c=1}^{C} \hat{\pi}_c \prod_{j=1}^{d} \hat{P}(x_j|K_c), \quad \text{or} \quad x \mapsto \arg\max_{c=1}^{C} \hat{\pi}_c \prod_{j=1}^{d} \hat{p}(x_j|K_c)$$

Exercise 2.2.1

Calculate the naive Bayes classifier for the dataset below for x = (195,g, yellow). Use a Gaussian Mixture Model for the weight variable and a discrete model for the color

Class:	Weight:	Color:
Apple	162g	red
Apple	186g	yellow
Banana	112g	yellow
Banana	142g	green
Banana	128g	yellow

Table 2.1: Fruit Dataset