*Problem 1: Parametric Estimation*

1. Since samples $x_1, ..., x_n$ are drawn independently from the Bernoulli distribution,

$$\begin{aligned} p(D|\theta) &= \prod_{i=1}^{n}(p(x_i|\theta)) \\ &= \prod_{i=1}^{n}\theta^{x_i}(1-\theta)^{(1-x_i)} \\ &= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{(n-\sum_{i=1}^{n} x_i)} \end{aligned}$$

   Because $x_i \in \{0, 1\}$, the previous equation can be expressed as $p(D|\theta) = \theta^s(1-\theta)^{(n-s)}$ with $s = \sum_{i=1}^{n} x_i$.

2. Since $\theta$ denotes the probability of getting head or tail and a uniform prior is assumed,

$$\theta \sim U(0,1) = \begin{cases} 1 & 0 \leq \theta \leq 1 \\ 0 & else \end{cases}$$

   According to Bayes parameter estimation:

$$\begin{aligned} p(\theta|D) &= \frac{p(D|\theta)p(\theta)}{p(D)} \\ &= \frac{p(D|\theta)p(\theta)}{\int_0^1 p(D|\theta)p(\theta)d\theta} \end{aligned}$$

   where

$$p(D|\theta)p(\theta) = \begin{cases} \theta^s(1-\theta)^{(n-s)} & 0 \leq \theta \leq 1 \\ 0 & else \end{cases}$$

   and

$$\begin{aligned} &\int_0^1 p(D|\theta)p(\theta)d\theta \\ =& \int_0^1 \theta^s(1-\theta)^{(n-s)}d\theta \\ =& \frac{s!(n-s)!}{(n+1)!} \end{aligned}$$

   Thus we can get

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int_0^1 p(D|\theta)p(\theta)d\theta} = \frac{(n+1)!}{s!(n-s)!}\theta^s(1-\theta)^{n-s}$$

   for $0 \leq \theta \leq 1$

3.

$$\begin{aligned} p(x|D) &= \int_0^1 p(x|\theta)p(\theta|D)d\theta \\ &= \int_0^1 \theta^x(1-\theta)^{(1-x)}\frac{(n+1)!}{s!(n-s)!}\theta^s(1-\theta)^{n-s}d\theta \\ &= \frac{(n+1)!}{s!(n-s)!}\int_0^1 \theta^{(x+s)}(1-\theta)^{(n+1-s-x)}d\theta \\ &= \frac{(n+1)!}{s!(n-s)!}\frac{(x+s)!(n+1-s-x)!}{(n+2)!} \\ &= \frac{(x+s)!(n+1-s-x)!}{s!(n-s)!(n+2)} \end{aligned}$$

Since $x \in \{0,1\}$, $p(x=0|D) = \frac{s!(n-s+1)!}{s!(n-s)!(n+2)} = \frac{n-s+1}{n+2} = 1-\frac{s+1}{n+2}$, $p(x=1|D) = \frac{(1+s)!(n-s)!}{s!(n-s)!(n+2)} = \frac{s+1}{n+2}$, we can get:

$$p(x|D) = (\frac{s+1}{n+2})^x(1-\frac{s+1}{n+2})^{1-x}$$

4. According to maximum likelihood estimation,

$$\hat{\theta} = argmax_\theta p(D|\theta) = argmax_\theta(\theta^s(1-\theta)^{(n-s)})$$

We can get $\hat{\theta}$ through:

$$\frac{dp(D|\theta)}{d\theta} = s\theta^{(s-1)}(1-\theta)^{(n-s)} - (n-s)\theta^s(1-\theta)^{(n-s-1)} = 0$$

Thus

$$\hat{\theta} = \frac{s}{n}$$

and

$$p(x|\hat{\theta}) = \hat{\theta}^x(1-\hat{\theta})^{(1-x)} = (\frac{s}{n})^x(1-\frac{s}{n})^{(1-x)}$$

...

*Problem2: Nonparametric Methods*

1. Because conditional densities are uniform within unit hyperspheres a distance of ten units apart. Which means, the $n$ samples will be distributed inside two different unit hyperspheres based on their labels. If we are given a input $x$ which belongs to $\omega_i$, samples with the same label $\omega_i$ will be nearer to $x$ than samples with the other label, because the two unit hyperspheres for the two classes have distance of ten units.
Thus, if $x$ was classified to a wrong class, number of samples with the same label $\omega_i$ must be less than $(k+1)/2$ such that the kNN method

will include more samples with the other label. Therefore, the probability of error is the probability of having number of samples in $\omega_i$ less than $(k+1)/2$, which can be represented as:

$$P_n(e) = \sum_{j=0}^{(k-1)/2} \binom{n}{j} \frac{1}{2^k} \frac{1}{2^{(n-k)}} = \frac{1}{2^n} \sum_{j=0}^{(k-1)/2} \binom{n}{j}$$

2. For single-nearest neighbor rule, the probability of error is:

$$P_n^1(e) = P_n(e)|_{k=0} = \frac{1}{2^n}$$

For $k > 1$,

$$P_n(e) = \frac{1}{2^n} \sum_{j=0}^{(k-1)/2} \binom{n}{j} = \frac{1}{2^n} + \frac{1}{2^n} \sum_{j=1}^{(k-1)/2} \binom{n}{j} > P_n^1(e)$$

Thus, the single-nearest neighbor rule has a lower error rate than the k-nearest-neighbor error rate for $k > 1$.

3. When $\frac{k-1}{2} < \frac{n}{2}$, $\binom{n}{j} < \binom{n}{(k-1)/2}$ for any $j < \frac{k-1}{2}$, thus we can have:

$$\sum_{j=0}^{(k-1)/2} \binom{n}{j} < \frac{k+1}{2} \binom{n}{(k-1)/2}$$

Furthermore, Since $\binom{n}{j}$ is an increasing function of $j$ with $0 \leq j \leq n/2$, and

$$2^n = \sum_{j=0}^{n} \binom{n}{j} = 2 \sum_{j=0}^{n/2} \binom{n}{j} > 2 \sum_{(k-1)/2}^{n/2} \binom{n}{j}$$

we will have:

$$2^n > 2(\frac{n}{2} - \frac{k-1}{2}) \binom{n}{(k-1)/2}$$

Therefore:

$$P_n(e) < \frac{\frac{k+1}{2} \binom{n}{(k-1)/2}}{2(\frac{n}{2} - \frac{k-1}{2}) \binom{n}{(k-1)/2}} = \frac{\frac{k+1}{2}}{n-(k-1)} < \frac{a\sqrt{n}+1}{2n - 2a\sqrt{n}+2}$$

when $\frac{k-1}{2} < \frac{n}{2}$

Since $k < a\sqrt{n}$, $\frac{k-1}{2} < \frac{n}{2}$ is true when $n > a^2$ Therefore, as $n \to \infty$, $P_n(e) < \frac{a\sqrt{n}+1}{2n-2a\sqrt{n}+2}$, since

$$n \to \infty, \frac{a\sqrt{n}+1}{2n-2a\sqrt{n}+2} \to 0$$

we can prove that $P_n(e) \to 0$, as $n \to \infty$