Q1)

$$P(x) = \frac{1}{nh_n}\sum_{i=1}^{n}\varphi\left(\frac{x-x_i}{h_n}\right)$$

1- $\boldsymbol{\tilde{p}_n(x) \sim N(\mu_0 h_n^2 + \sigma^2)}$

$$\tilde{p}_n(x) = E\{P_x(x)\} = \sum_{1}^{n} E\left\{\frac{1}{nh_n}\varphi\left(\frac{x-x_i}{h_n}\right)\right\} = \frac{1}{n}\times nE\left\{\frac{1}{h_n}\varphi\left(\frac{x-x_i}{h_n}\right)\right\} = E\left\{\frac{1}{h_n}\varphi\left(\frac{x-x_i}{h_n}\right)\right\}$$

$$= \int_{-\infty}^{\infty}\frac{1}{h_n}\varphi\left(\frac{x-y}{h_n}\right)p(y)dy = \int_{-\infty}^{\infty}\frac{1}{h_n}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-y}{h_n}\right)^2}\times\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}dy$$

$$= \frac{1}{2\pi\sigma h_n}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\left[\left(\frac{x-y}{h_n}\right)^2+\left(\frac{y-\mu}{\sigma}\right)^2\right]}dy = \frac{e^{-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}}}{2\pi\sigma h_n}\int_{-\infty}^{\infty}e^{-\frac{h_n^2+\sigma^2}{2h_n^2\sigma^2}\left[y-\frac{\sigma^2 x+h_n^2\mu}{\sigma^2+h_n^2}\right]^2}dy$$

$$= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{h_n^2+\sigma^2}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}}$$

2- $\boldsymbol{p_n(x) - \tilde{p}_n(x) \cong \frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2\left[1-\left(\frac{x-\mu}{\sigma}\right)^2\right]p(x)}$

$$p_n(x) - \tilde{p}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n}\varphi\left(\frac{x-x_i}{h_n}\right) - \frac{1}{\sqrt{2\pi}\sqrt{h_n^2+\sigma^2}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}}$$

$$= \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}}\left[\sum_{i=1}^{n}\frac{1}{h_n}e^{-\frac{1}{2}\left[\left(\frac{x-x_i}{h_n}\right)^2+\left(\frac{x-\mu}{\sigma}\right)^2\right]}\right] = \frac{1}{1-\frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2}e^{-\frac{1}{2}\left[\left(\frac{x-x_i}{h_n}\right)^2+\left(\frac{x-\mu}{\sigma}\right)^2\right]}$$

$$\approx \frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2\left[1-\left(\frac{x-\mu}{\sigma}\right)^2\right]p(x)$$

3- $\boldsymbol{var(p_n(x)) \cong \frac{1}{2nh_n\sqrt{\pi}}p(x)}$

$$var(p_n(x)) = \sum_{1}^{n} var\left\{\frac{1}{nh_n}\varphi\left(\frac{x-x_i}{h_n}\right)\right\} = n\times var\left\{\frac{1}{nh_n}\varphi\left(\frac{x-x_i}{h_n}\right)\right\}$$

$$= n\left(E\left\{\frac{1}{nh_n}\varphi\left(\frac{x-x_i}{h_n}\right)\right\}^2 - \left(\frac{\tilde{p}_n(x)}{n}\right)^2\right) = n\frac{1}{n^2h_n}E\left\{\frac{1}{h_n}\varphi^2\left(\frac{x-x_i}{h_n}\right)\right\} - \frac{1}{n}\tilde{p}_n^2(x)$$

$$E\left\{\frac{1}{h_n}\varphi^2\left(\frac{x-x_i}{h_n}\right)\right\} = \int_{-\infty}^{\infty}\frac{1}{2nh_n}e^{-\frac{1}{2}\left(\frac{x-y}{h_n}\right)^2}\times\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}dy$$

$$= \frac{1}{\sqrt{(2\pi)^3}h_n\sigma}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\left[\frac{(x-x_i)^2}{\frac{1}{2}h_n^2}+\left(\frac{x-\mu}{\sigma}\right)^2\right]}dy = \frac{1}{2\pi\sqrt{\frac{h_n^2}{2}+\sigma^2}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\frac{h_n^2}{2}+\sigma^2}}$$

$$\rightarrow var(p_n(x)) = \frac{1}{n}\left[\frac{1}{2\pi\sqrt{\frac{h_n^2}{2}+\sigma^2}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\frac{h_n^2}{2}+\sigma^2}} - \tilde{p}_n^2(x)\right] \approx \frac{1}{2nh_n\sqrt{\pi}}p(x)$$

Q2)

$$D(a.b) = \sqrt{\sum_{k=1}^{d} (a_k - b_k)^2} \qquad x'_k = \alpha_k x_k$$

$$D(a',b') = \sqrt{\sum_{k=1}^{d} (\alpha_k a_k - \alpha_k b_k)^2} = \sqrt{\sum_{k=1}^{d} \alpha_k^2 (a_k - b_k)^2}$$

1) Non-Negativity: $D(a',b') = \left(\sum_{k=1}^{d} \alpha_k^2 (a_k - b_k)^2\right)^{\frac{1}{2}} \geq 0$

2) Reflexivity: $a' = b' \rightarrow \alpha a_k = \alpha b_k \rightarrow a_k = b_k \rightarrow D(a',b') = 0$

$$D(a',b') = 0 \rightarrow \alpha_k^2 (a_k - b_k)^2 = 0 \rightarrow a_k = b_k \rightarrow a' = b'$$

3) Symmetry: $D(a',b') = \sqrt{\sum_{k=1}^{d} \alpha_k^2 (a_k - b_k)^2} = \sqrt{\sum_{k=1}^{d} \alpha_k^2 (b_k - a_k)^2} = D(b',a')$

4) Triangle Inequality: $D(a',b')D(c',b') \geq D(a',c')$

$$\sqrt{\sum_{k=1}^{d} \alpha_k^2 (a_k - b_k)^2} \times \sqrt{\sum_{k=1}^{d} \alpha_k^2 (c_k - b_k)^2} \geq^{?} \sqrt{\sum_{k=1}^{d} \alpha_k^2 (a_k - c_k)^2}$$

$$\sum_{k=1}^{d} \alpha_k^2 [(a_k - b_k)^2 + (c_k - b_k)^2] + 2\alpha_k^2 \sum_{k=1}^{d} \sum_{k=1}^{d} (a_k - b_k)^2 (c_k - b_k)^2 \geq^{?} \sum_{k=1}^{d} \alpha_k^2 (a_k - c_k)^2$$

$$\rightarrow 2 \sum_{k=1}^{d} \sum_{k=1}^{d} (a_k - b_k)^2 (c_k - b_k)^2 \geq 0$$

در knn در بعضی موارد بعضی از ابعاد مقادیر بیشتری از ابعاد دیگر دارند و باید اقدامی صورت بگیرد تا ارزش آن‌ها در تصمیم‌گیری لحاظ شود. این اقدام، اسکیل کردن ابعاد است. راهکار این است که هر بعد در α_k ضرب می‌شود و درنهایت با متریک اقلیدوسی فواصل موازنه می‌شوند.

Q3)

Part a)

$$p_Q(e) = \frac{1}{2^Q}\sum_{j=0}^{\frac{k-1}{2}}\binom{Q}{j} \qquad D = \{x^q, \omega_i^q\} \qquad \{x_i\} \quad i = 1, \dots, k$$

k دیتا پوینت موجود است. مساله نیز knn ۲ کلاسه است. به دلیل فرد بودن k یک کلاس حاوی سمپل‌های بیشتری است. اگر j سمبل از Q سمبل اشتباه شود، کلاس آن اشتباه است. تعداد کل حالات نیز 2^Q است. پس داریم:

$$p_Q(e) = \frac{\sum_{j=0}^{\frac{k-1}{2}}\binom{Q}{j}}{\sum_{j=0}^{Q}\binom{Q}{j}} = \frac{1}{2^Q}\sum_{j=0}^{\frac{k-1}{2}}\binom{Q}{j}$$

Part b)

برای k های فرد، عبارت $\sum_{j=0}^{\frac{k-1}{2}}\binom{Q}{j}$ با زیاد شدن k افزایش می‌یابد. تعداد کل حالات نیز همیشه ثابت است. پس با افزایش k، احتمال خطا $p_Q(e)$ افزایش می‌یابد.

این دو کلاس، توزیع یکنواخت روی دایره واحد دارند. بنابراین انتخاب یک k زوج ممکن است حالاتی را به وجود بیارد که تعداد سمبل‌های هر کلاس یکسان شود و تصمیم‌گیری رندوم (50~50) صورت بگیرد که خطای بیشتری دارد.

Part c)

$$\lim_{Q\to\infty} p_Q(e) = \lim_{Q\to\infty}\frac{\sum_{j=0}^{\frac{k-1}{2}}\binom{Q}{j}}{2^Q} = \sum_{j=0}^{\frac{k-1}{2}}\lim_{Q\to\infty}\frac{\frac{Q!}{j!\,(Q-j)!}}{2^Q} = \sum_{j=0}^{\frac{k-1}{2}}\lim_{Q\to\infty}\frac{Q(Q-1)\dots(Q-j+1)(Q-j)!}{j!\,(Q-j)!\,2^Q}$$

$$= \sum_{j=0}^{\frac{k-1}{2}}\frac{1}{j!}\lim_{Q\to\infty}\frac{Q(Q-1)\dots(Q-j+1)}{2^Q} = \sum_{j=0}^{\frac{k-1}{2}}\frac{1}{j!}\lim_{Q\to\infty}\frac{Q^j+\alpha_1 Q^{j-1}+\alpha_2 Q^{j-2}+\cdots}{2^Q} = 0$$