$$\begin{aligned}
&\leq \frac{1}{nV_n}\int \frac{1}{V_n}\varphi^2\left(\frac{\mathbf{x}-\mathbf{V}}{h_n}\right)p(\mathbf{V})d\mathbf{V} \\
&\leq \frac{\mathrm{Sup}(\varphi)}{nV_n}\int \frac{1}{V_n}\varphi\left(\frac{\mathbf{x}-\mathbf{V}}{h_n}\right)p(\mathbf{V})d\mathbf{V} \\
&\leq \frac{\mathrm{Sup}(\varphi)\bar{p}_n(\mathbf{x})}{nV_n}.
\end{aligned}$$

We note that $\mathrm{Sup}(\varphi) < \infty$ and that in the limit $n \to \infty$ we have $\bar{p}_n(\mathbf{x}) \to p(\mathbf{x})$ and $nV_n \to \infty$. We put these results together to conclude that

$$\lim_{n\to\infty} \sigma_n^2(\mathbf{x}) = 0,$$

for all $\mathbf{x}$.

2. Our normal distribution is $p(x) \sim N(\mu, \sigma^2)$ and our Parzen window is $\varphi(x) \sim N(0,1)$, or more explicitly,

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2},$$

and our estimate is

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

(a) The expected value of the probability at x, based on a window width parameter h_n, is

$$\begin{aligned}
\bar{p}_n(x) &= \mathcal{E}[p_n(x)] = \frac{1}{nh_n}\sum_{i=1}^{n}\mathcal{E}\left[\varphi\left(\frac{x-x_i}{h_n}\right)\right] \\
&= \frac{1}{h_n}\int_{-\infty}^{\infty}\varphi\left(\frac{x-v}{h_n}\right)p(v)dv \\
&= \frac{1}{h_n}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-v}{h_n}\right)^2\right]\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2\right]dv \\
&= \frac{1}{2\pi h_n\sigma}\exp\left[-\frac{1}{2}\left(\frac{x^2}{h_n^2}+\frac{\mu^2}{\sigma^2}\right)\right]\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}v^2\left(\frac{1}{h_n^2}+\frac{1}{\sigma^2}\right)-2v\left(\frac{x}{h_n^2}+\frac{\mu}{\sigma^2}\right)\right]dv \\
&= \frac{1}{2\pi h_n\sigma}\exp\left[-\frac{1}{2}\left(\frac{x^2}{h_n^2}+\frac{\mu^2}{\sigma^2}\right)+\frac{1}{2}\frac{\alpha^2}{\theta^2}\right]\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\left(\frac{v-\alpha}{\theta}\right)^2\right]dv,
\end{aligned}$$

where we have defined

$$\theta^2 = \frac{1}{1/h_n^2 + 1/\sigma^2} = \frac{h_n^2\sigma^2}{h_n^2+\sigma^2}$$

and

$$\alpha = \theta^2\left(\frac{x}{h_n^2}+\frac{\mu}{\sigma^2}\right).$$

We perform the integration and find

$$\begin{aligned}
\bar{p}_n(x) &= \frac{\sqrt{2\pi}\theta}{2\pi h_n \sigma} \exp\left[-\frac{1}{2}\left(\frac{x^2}{h_n^2} + \frac{\mu^2}{\sigma^2} \right) + \frac{1}{2}\frac{\alpha^2}{\theta^2} \right] \\
&= \frac{1}{\sqrt{2\pi} h_n \sigma} \frac{h_n \sigma}{\sqrt{h_n^2 + \sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x^2}{h_n^2} + \frac{\mu^2}{\sigma^2} - \frac{\alpha^2}{\theta^2} \right) \right].
\end{aligned}$$

The argument of the exponentiation can be expressed as follows

$$\begin{aligned}
\frac{x^2}{h_n^2} + \frac{\mu^2}{\sigma^2} - \frac{\alpha^2}{\theta^2} &= \frac{x^2}{h_n^2} + \frac{\mu^2}{\sigma^2} - \frac{\theta^4}{\theta^2}\left(\frac{x}{h_n^2} + \frac{\mu}{\sigma^2} \right)^2 \\
&= \frac{x^2 h_n^2}{(h_n^2 + \sigma^2) h_n^2} + \frac{\mu^2 \sigma^2}{(h_n^2 + \sigma^2)\sigma^2} - \frac{2x\mu}{h_n^2 + \sigma^2} \\
&= \frac{(x-\mu)^2}{h_n^2 + \sigma^2}.
\end{aligned}$$

We substitute this back to find

$$\bar{p}_n(x) = \frac{1}{\sqrt{2\pi}\sqrt{h_n^2 + \sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2 + \sigma^2} \right],$$

which is the form of a Gaussian, denoted

$$\bar{p}_n(x) \sim N(\mu, h_n^2 + \sigma^2).$$

(b) We calculate the variance as follows:

$$\begin{aligned}
\mathrm{Var}[p_n(x)] &= \mathrm{Var}\left[\frac{1}{nh_n} \sum_{i=1}^n \varphi\left(\frac{x - x_i}{h_n} \right) \right] \\
&= \frac{1}{n^2 h_n^2} \sum_{i=1}^n \mathrm{Var}\left[\varphi\left(\frac{x - x_i}{h_n} \right) \right] \\
&= \frac{1}{nh_n^2} \mathrm{Var}\left[\varphi\left(\frac{x - v}{h_n} \right) \right] \\
&= \frac{1}{nh_n^2} \left\{ \mathcal{E}\left[\varphi^2\left(\frac{x - v}{h_n} \right) \right] - \left(\mathcal{E}\left[\varphi\left(\frac{x - v}{h_n} \right) \right] \right)^2 \right\},
\end{aligned}$$

where in the first step we used the fact that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent samples drawn according to $p(\mathbf{x})$. We thus now need to calculate the expected value of the square of the kernel function

$$\begin{aligned}
\mathcal{E}\left[\varphi^2\left(\frac{x - v}{h_n} \right) \right] &= \int \varphi^2\left(\frac{x - v}{h_n} \right) p(v) dv \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left[-\left(\frac{x - v}{h_n} \right)^2 \right] \exp\left[-\frac{1}{2}\left(\frac{v - \mu}{\sigma} \right)^2 \right] \frac{1}{\sqrt{2\pi}\sigma} dv \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - v}{h_n/\sqrt{2}} \right)^2 \right] \exp\left[-\frac{1}{2}\left(\frac{v - \mu}{\sigma} \right)^2 \right] \frac{1}{\sqrt{2\pi}\sigma} dv.
\end{aligned}$$

From part (a) by a similar argument with $h_n/\sqrt{2}$ replacing h_n, we have

$$\frac{1}{h_n/\sqrt{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-v}{h_n/\sqrt{2}}\right)^2\right] \exp\left[-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2\right] \frac{1}{\sqrt{2\pi}\sigma} dv$$
$$= \frac{1}{\sqrt{2\pi}\sqrt{h_n^2/2+\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2/2+\sigma^2}\right].$$

We make the substitution and find

$$\mathcal{E}\left[\varphi^2\left(\frac{x-v}{h_n}\right)\right] = \frac{h_n/\sqrt{2}}{2\pi\sqrt{h_n^2/2+\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2/2+\sigma^2}\right],$$

and thus conclude

$$\frac{1}{nh_n^2}\mathcal{E}\left[\varphi^2\left(\frac{x-v}{h_n}\right)\right] = \frac{1}{nh_n}\frac{1}{2\sqrt{\pi}}\frac{1}{\sqrt{2\pi}\sqrt{h_n^2/2+\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2/2+\sigma^2}\right].$$

For small h_n, $\sqrt{h_n^2/2+\sigma^2} \simeq \sigma$, and thus the above equation can be approximated as

$$\begin{aligned}
\frac{1}{nh_n^2}\mathcal{E}\left[\varphi^2\left(\frac{x-v}{h_n}\right)\right] &\simeq \frac{1}{nh_n 2\sqrt{\pi}}\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \\
&= \frac{1}{2nh_n\sqrt{\pi}} p(x). \qquad (*)
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\frac{1}{nh_n^2}\left\{\mathcal{E}\left[\varphi\left(\frac{x-v}{h_n}\right)\right]\right\}^2 &= \frac{1}{nh_n^2} h_n^2 \frac{1}{\sqrt{2\pi}\sqrt{h_n^2+\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right] \\
&= \frac{h_n}{nh_n}\frac{1}{\sqrt{2\pi}\sqrt{h_n^2+\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right] \\
&\simeq \frac{h_n}{nh_n}\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] \simeq 0, (**)
\end{aligned}$$

valid for small h_n. From $(*)$ and $(**)$ we have, (still for small h_n)

$$\mathrm{Var}[P_n(x)] \simeq \frac{p(x)}{2nh_n\sqrt{\pi}}.$$

(c) We write the bias as

$$\begin{aligned}
p(x) - \bar{p}_n(x) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] - \frac{1}{\sqrt{2\pi}\sqrt{h_n^2+\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right] \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right]\left\{1 - \frac{\sigma}{\sqrt{h_n^2+\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2} + \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]\right. \\
&= p(x)\left\{1 - \frac{1}{\sqrt{1+(h_n/\sigma)^2}} \exp\left[-\frac{(x-\mu)^2}{2}\left\{\frac{1}{h_n^2+\sigma^2} - \frac{1}{\sigma^2}\right\}\right]\right\} \\
&= p(x)\left\{1 - \frac{1}{\sqrt{1+(h_n/\sigma)^2}} \exp\left[\frac{1}{2}\frac{h_n^2(x-\mu)^2}{h_n^2+\sigma^2}\right]\right\}.
\end{aligned}$$

For small h_n we expand to second order:

$$\frac{1}{\sqrt{1+(\frac{h_n}{\sigma})^2}} \simeq 1-\frac{1}{2}\left(h_n/\sigma\right)^2$$

and

$$\exp\left[\frac{h_n^2}{2\sigma^2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right] \simeq 1+\frac{h_n^2}{2\sigma^2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}.$$

We ignore terms of order higher than h_n^2 and find

$$\begin{aligned}
p(x)-\bar{p}_n(x) &\simeq p(x)\left\{1-\left(1-\frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2\right)\left(1+\frac{h_n^2}{2\sigma^2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right)\right\} \\
&\simeq p(x)\left\{1-1+\frac{1}{2}\frac{h_n^2}{\sigma^2}-\frac{h_n^2}{2\sigma^2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right\} \\
&\simeq \frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2\left[1-\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right]p(x) \\
&\simeq \frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2\left[1-\left(\frac{x-\mu}{\sigma}\right)^2\right]p(x).
\end{aligned}$$

3. Our (normalized) distribution is

$$p(x)=\begin{cases} 1/a & 0\le x\le a \\ 0 & \text{otherwise,}\end{cases}$$

and our Parzen window is

$$\varphi(x)=\begin{cases} e^{-x} & x\ge 0 \\ 0 & x<0.\end{cases}$$

(a) The expected value of the Parzen window estimate is

$$\begin{aligned}
\bar{p}_n(x) &= \mathcal{E}\left[\frac{1}{nh_n}\sum_{i=1}^{n}\varphi\left(\frac{x-x_i}{h_n}\right)\right]=\frac{1}{h_n}\int\varphi\left(\frac{x-v}{h_n}\right)p(v)dv \\
&= \frac{1}{h_n}\int_{x\ge v} e^{-\frac{(x-v)}{h_n}}p(v)dv \\
&= \frac{\exp[-x/h_n]}{h_n}\int_{\substack{x\ge v,\\ 0<v<a}}\frac{1}{a}\exp[v/h_n]dv \\
&= \begin{cases} 0 & \text{if } x<0 \\ \frac{e^{-x/h_n}}{ah_n}\int_0^x e^{v/h_n}dv & \text{if } 0\le x\le a \\ \frac{e^{-x/h_n}}{ah_n}\int_0^a e^{v/h_n}dv & \text{if } x\ge a \end{cases}
\end{aligned}$$

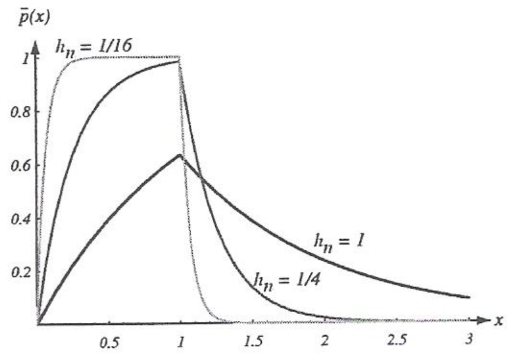
$$= \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{a}(1 - e^{-x/h_n}) & \text{if } 0 \le x \le a \\ \frac{1}{a}(e^{a/h_n} - 1)e^{-x/h_n} & \text{if } x \ge a, \end{cases}$$

where in the first step we used the fact that $x_1, \ldots, x_n$ are independent samples drawn according to $p(v)$.

(b) For the case $a = 1$, we have

$$\bar{p}_n(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x/h_n} & 0 \le x \le 1 \\ (e^{1/h_n} - 1)e^{-x/h_n} & x > 1, \end{cases}$$

as shown in the figure.



(c) The bias is

$$\begin{aligned} p(x) - \bar{p}_n(x) &= \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{a} - \frac{1}{a}(1 - e^{-x/h_n}) & \text{if } 0 \le x \le a \\ 0 - \frac{1}{a}(e^{a/h_n} - 1)e^{-x/h_n} & \text{if } x \ge a \end{cases} \\ &= \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{a}e^{-x/h_n} & \text{if } 0 \le x \le a \\ -\frac{1}{a}(e^{a/h_n} - 1)e^{-x/h_n} & \text{if } x \ge a. \end{cases} \end{aligned}$$

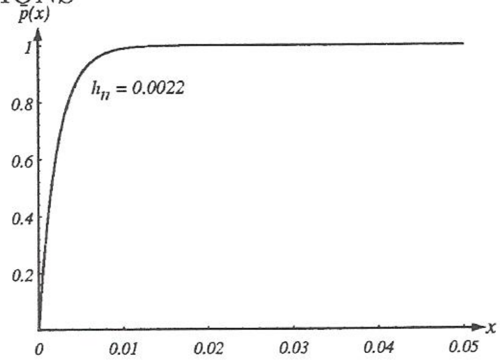
Formally, a bias lower than 1% over 99% of the range $0 < x < a$, means that

$$\frac{p(x) - \bar{p}(x)}{p(x)} \le 0.01$$

over 99% of the range $0 < x < a$. This, in turn, implies

$$\begin{aligned} \frac{1/ae^{-x/h_n}}{1/a} &\le 0.01 \quad \text{over 99\% of } 0 < x < a \text{ or} \\ h_n &\le \frac{0.01a}{\ln{(100)}}. \end{aligned}$$

For the case $a = 1$, we have that $h_n \le 0.01/(\ln 100) = 0.0022$, as shown in the figure. Notice that the estimate is within 1% of $p(x) = 1/a = 1.0$ above $x \sim 0.01$, fulfilling the conditions of the problem.



4. We have from Algorithm 2 in the text that the discriminant functions of the PNN classifier are given by

$$g_i(\mathbf{x}) = \sum_{k=1}^{n_i} \exp\left[\frac{\mathbf{w}_k^t\mathbf{x} - 1}{\sigma_2}\right] a_{k_i} \qquad i = 1, \ldots, c$$

where $\|\mathbf{w}_k\| = \|\mathbf{x}\| = 1$, n_i is the number of training patterns belonging to ω_i and

$$a_{k_i} = \begin{cases} 1 & \text{if } \mathbf{w}_k \in \omega_i \\ 0 & \text{otherwise.} \end{cases}$$

(a) Since $\|\mathbf{w}_k\| = \|\mathbf{x}\| = 1$, $g_i(\mathbf{x})$ can be written as

$$g_i(\mathbf{x}) = \sum_{i=1}^{n_i} \exp\left[-\frac{\|\mathbf{x} - \mathbf{w}_k\|^2}{2\sigma^2}\right] a_{k_i}.$$

Note that $g_i(\mathbf{x})/n_i$ is a radial Gaussian based kernel estimate, $p_n(\mathbf{x}|\omega_i)$, of $p(\mathbf{x}|\omega_i)$. If we use n_i/n as the estimate of the prior class probability $P(\omega_i)$, then $g_i(\mathbf{x})$ can be rewritten as

$$g_i(\mathbf{x}) = nP_n(\omega_i)p(\mathbf{x}|\omega_i).$$

Thus $g_i(\mathbf{x})$ properly accounts for the class priors.

(b) The optimal classification rule for unequal costs is given by

$$\text{Choose } \omega_k \text{ if } g_k^* = \min_{i \le c} g_i^*,$$

where the λ_{ij} represent the costs and

$$g_i^*(\mathbf{x}) = \sum_{j=1}^{c} \lambda_{ij} P(\omega_j|\mathbf{x}) = \sum_{j=1}^{c} \frac{P(\omega_j)p(\mathbf{x}|\omega_j}{p(\mathbf{x})}.$$

This discriminant function can be written simply as

$$g_i^*(\mathbf{x}) = \sum_{j=1}^{c} \lambda_{ij} P(\omega_j)p(\mathbf{x}|\omega_j).$$

Consequently, the PNN classifier must estimate $g_i^*(\mathbf{x})$. From part (a) we have that $g_i = nP_n(\omega_j)p(\mathbf{x}|\omega_j)$. Thus the new discriminant functions are simply

$$\hat{g}_i(\mathbf{x}) = \frac{1}{n} \sum_{j=1}^{c} \lambda_{ij} g_j(\mathbf{x}),$$

$$= \int_0^{cr/(c-1)} \left(1 - \frac{1/c}{p^2(x)}\right) p(x)dx$$

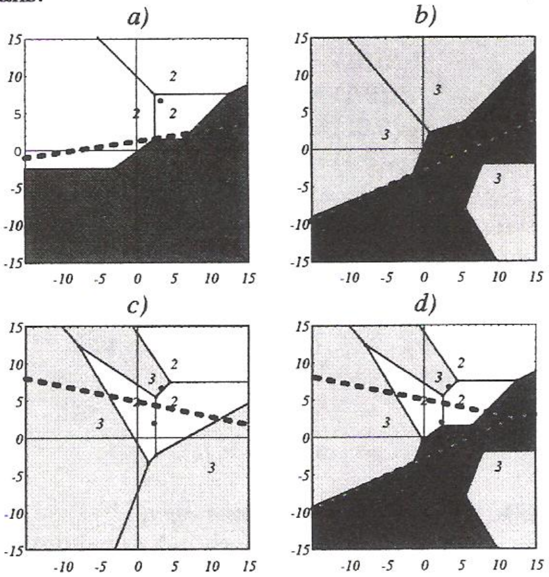
$$= \int_0^{cr/(c-1)} \left(1 - \frac{1}{c}\right) dx = \left(1 - \frac{1}{c}\right)\frac{cr}{c-1} = r.$$

Thus we have demonstrated that $P^* = P = r$ in this nontrivial case.

9. Our data are given in the table.

ω_1	ω_2	ω_3
(10,0)	(5,10)	(2,8)
(0,-10)	(0,5)	(-5,2)
(5,-2)	(5,5)	(10,-4)

Throughout the figures below, we let dark gray represent category ω_1, white represent category ω_2 and light gray represent category ω_3. The data points are labeled by their category numbers, the means are shown as small black dots, and the dashed straight lines separate the means.



10. The Voronoi diagram of n points in d-dimensional space is the same as the convex hull of those same points projected orthogonally to a hyperboloid in $(d+1)$-dimensional space. So the editing algorithm can be solved either with a Voronoi diagram algorithm in d-space or a convex hull algorithm in $(d+1)$-dimensional space. Now there are scores of algorithms available for both problems all with different complexities.

A theorem in the book by Preparata and Shamos refers to the complexity of the Voronoi diagram itself, which is of course a lower bound on the complexity of computing it. This complexity was solved by Victor Klee, "On the complexity of d-dimensional Voronoi diagrams," *Archiv. de Mathematik.*, vol. 34, 1980, pp. 75-80. The complexity formula given in this problem is the complexity of the convex hull algorithm of Raimund Seidel, "Constructing higher dimensional convex hulls at logarithmic cost per face," *Proc. 18th ACM Conf. on the Theory of Computing*, 1986, pp. 404-413.

So here d is one bigger than for Voronoi diagrams. If we substitute d in the formula in this problem with $(d-1)$ we get the complexity of Seidel's algorithm for Voronoi diagrams, as discussed in A. Okabe, B. Boots and K. Sugihara, **Spatial Tessellations: Concepts and Applications of Voronoi Diagrams**, John Wiley, 1992.

11. Consider the "curse of dimensionality" and its relation to the separation of points randomly selected in a high-dimensional space.

(a) The sampling density goes as $n^{1/d}$, and thus if we need n_1 samples in $d=1$ dimensions, an "equivalent" number samples in d dimensions is n_1^d. Thus if we needed 100 points in a line (i.e., $n_1 = 100$), then for $d = 20$, we would need $n_{20} = (100)^{20} = 10^{40}$ points — roughly the number of atoms in the universe.

(b) We assume roughly uniform distribution, and thus the typical inter-point Euclidean (i.e., L_2) distance δ goes as $\delta^d \sim$ volume, or $\delta \sim (\text{volume})^{1/d}$.

(c) Consider points uniformly distributed in the unit interval $0 \leq x \leq 1$. The length containing fraction p of all the points is of course p. In d dimensions, the width of a hypercube containing fraction p of points is $l_d(p) = p^{1/d}$. Thus we have

$$\begin{aligned} l_5(0.01) &= (0.01)^{1/5} &= 0.3910 \\ l_5(0.1) &= (0.1)^{1/5} &= 0.7248 \\ l_{20}(0.01) &= (0.01)^{1/20} &= 0.8609 \\ l_{20}(0.1) &= (0.1)^{1/20} &= 0.8609. \end{aligned}$$

(d) The L_∞ distance between two points in d-dimensional space is given by Eq. 57 in the text, with $k \to \infty$:

$$\begin{aligned} L_\infty(\mathbf{x}, \mathbf{y}) &= \lim_{k\to\infty} \sqrt[1/k]{\sum_{i=1}^{d} |x_i - y_i|^k} \\ &= \max[|x_1 - y_1|, |x_2 - y_2|, \ldots, |x_d - y_d|] \\ &= \max_i |x_i - y_i|. \end{aligned}$$

In other words, consider each axis separately, $i = 1, \ldots, d$. There is a separation between two points $\mathbf{x}$ and $\mathbf{y}$ along each individual direction i, that is, $|x_i - y_i|$. One of these distances is the greatest. The L_∞ distance between two points is merely this maximum distance.

Informally we can see that for two points randomly selected in the unit d-dimensional hypercube $[0,1]^d$, this L_∞ distance approaches 1.0 as we can nearly always find an axis i for which the separation is large. In contrast, the L_∞ distance to *the closest* of the faces of the hypercube approaches 0.0, because we can nearly always find an axis for which the distance to a face is small. Thus, nearly every point is closer to a face than to another randomly selected point. In short, nearly every point is on the "outside" (that is, on the "convex hull") of the set of points in a high-dimensional space — nearly every point is an "outlier."

We now demonstrate this result formally. Of course, $\mathbf{x}$ is always closer to a wall than 0.5 — even for $d = 1$ — and thus we consider distances l^* in the

(b) We follow the logic of part (a). Now our target function is $f(\mathbf{x}) = \sum_{i=1}^{d} a_i B_i(\mathbf{x})$ where each member in the basis set of M basis functions $B_i(\mathbf{x})$ is a function of the d-component vector $\mathbf{x}$. The approximation function is

$$\hat{f}(\mathbf{x}) = \sum_{m=1}^{M} \hat{a}_m B_m(\mathbf{x}),$$

and, as before, the coefficients are least-squares estimates

$$\hat{a}_i = \arg\min_{a_i} \sum_{i=1}^{n} \left[y_i - \sum_{m=1}^{q} a_m B_m(\mathbf{x}) \right]^2$$

and $y_i = f(\mathbf{x}_i) + N(0, \sigma^2)$. Now $\mathbf{y}$ will be approximated by $\mathbf{B}\hat{\mathbf{a}}$, the projection of $\mathbf{y}$ onto the column space of $\mathbf{B}_j$, that is, the subspace spanned by the M vectors

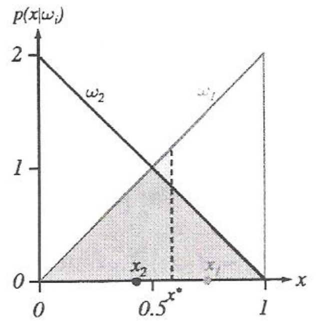
$$\begin{bmatrix} B_i(\mathbf{x}_1) \\ B_i(\mathbf{x}_2) \\ \vdots \\ B_i(\mathbf{x}_n) \end{bmatrix}.$$

As in part (a), we have

$$\mathcal{E}[(f(\mathbf{x}) - \hat{f}(\mathbf{x}))^2] = \frac{M\sigma^2}{n},$$

which is independent of d, the dimensionality of the original space.

13. We assume $P(\omega_1) = P(\omega_2) = 0.5$ and the distributions are as given in the figure.



(a) Clearly, by the symmetry of the problem, the Bayes decision boundary is $x^* = 0.5$. The error is then the area of the dark shading in the figure, divided by the total possible area, that is

$$P^* = \int_0^1 \min[P(\omega_1)p(x|\omega_1), P(\omega_2)p(x|\omega_2)]\, dx$$

$$
\begin{aligned}
&= P(\omega_1)\int_0^{0.5} 2x\ dx + P(\omega_2)\int_{0.5}^{1}(2-2x)\ dx \\
&= 0.5\frac{1}{4} + 0.5\frac{1}{4} = 0.25.
\end{aligned}
$$

(b) Suppose a point is randomly selected from ω_1 according to $p(x|\omega_1)$ and another point from ω_2 according to $p(x|\omega_2)$. We have that the error is

$$
\int_0^1 dx_1\ p(x_1|\omega_1)\int_0^1 dx_2\ p(x_2|\omega_2)\int_{(x_1-x_2)/2}^{1} dx\ p(x|\omega_2)\int_0^{(x_1-x_2)/2} dx\ p(x|\omega_1).
$$

(c) From part (d), below, we have for the special case $n = 2$,

$$
P_2(e) = \frac{1}{3} + \frac{1}{(2+1)(2+3)} + \frac{1}{2(2+2)(2+3)} = \frac{51}{120} = 0.425.
$$

(d) By symmetry, we may assume that the test point belongs to category ω_2. Then the chance of error is the chance that the nearest sample point ot the test point is in ω_1. Thus the probability of error is

$$
\begin{aligned}
P_n(e) &= \int_0^1 P(x|\omega_2)\text{Pr}[\text{nearest } y_i \text{ to } x \text{ is in } \omega_1]dx \\
&= \int_0^1 P(x|\omega_2)\sum_{i=1}^{n}\text{Pr}[y_i \in \omega_1 \text{ and } y_i \text{ is closer to } x \text{ than } y_j, \forall j \neq i]dx.
\end{aligned}
$$

By symmetry the summands are the same for all i, and thus we can write

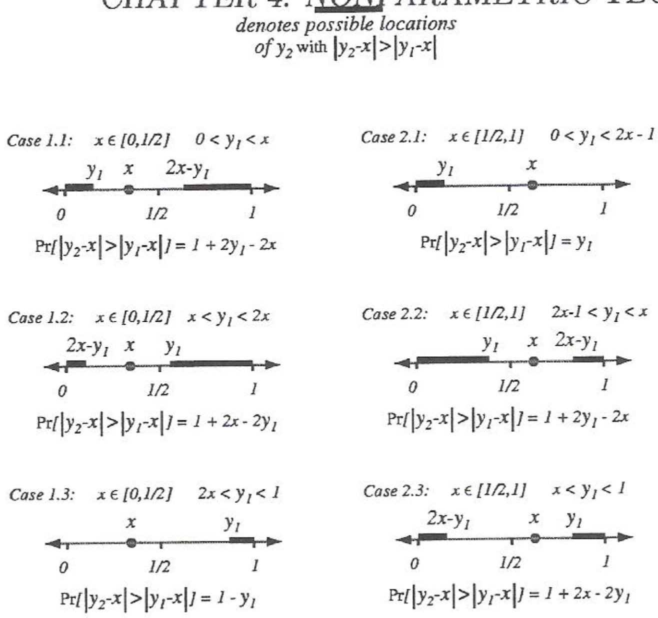
$$
\begin{aligned}
P_n(e) &= \int_0^1 P(x|\omega_2)n\text{Pr}[y_1 \in \omega_1 \text{ and } |y_1 - x| > |y_i - x|, \forall i > 1]dx \\
&= \int_0^1 P(x|\omega_2)n\int_0^1 P(\omega_1|y_1)\text{Pr}[|y_i - x| > |y_1 - x|, \forall i > 1]dy_1\ dx \\
&= \int_0^1 P(x|\omega_2)n\int_0^1 P(\omega_1|y_1)\text{Pr}[|y_2 - x| > |y_1 - x|]^{n-1}dy\ dx,
\end{aligned}
$$

where the last step again relies on the symmetry of the problem.

To evalute $\text{Pr}[|y_2 - x| > |y_1 - x|]$, we divide the integral into six cases, as shown in the figure.

We substitute these values into the above integral, and break the integral into the six cases as

$$
P_n(e) = \int_0^{1/2} P(x|\omega_2)n\left[\int_0^x P(\omega_1|y_1)(1 + 2y_1 - 2x)^{n-1}dy_1\right.
$$



$$
\begin{aligned}
& + \int_x^{2x} P(\omega_1|y_1)(1+2x-2y_1)^{n-1}dy_1 \\
& \left. + \int_{2x}^{1} P(\omega_1|y_1)(1-y_1)^{n-1}dy_1 \right] dx \\
+ \int_{1/2}^{1} P(x|\omega_2)n & \left[\int_0^{2x-1} P(\omega_1|y_1)y_1^{n-1}dy_1 \right. \\
& + \int_{2x-1}^{x} P(\omega_1|y_1)(1+2y_1-2x)^{n-1}dy_1 \\
& \left. + \int_x^{1} P(\omega_1|y_1)(1+2x-2y_1)^{n-1}dy_1 \right] dx.
\end{aligned}
$$

Our density and posterior are given as $p(x|\omega_2) = 2(1-x)$ and $P(\omega_1|y) = y$ for $x \in [0,1]$ and $y \in [0,1]$. We substitute these forms into the above large integral and find

$$
\begin{aligned}
P_n(e) &= \int_0^{1/2} 2n(1-x) \left[\int_0^x y_1(1+2y_1-2x)^{n-1}dy_1 \right. \\
& \qquad \int_x^{2x} y_1(1+2x-2y_1)^{n-1}dy_1 \\
& \qquad \left. \int_{2x}^{1} y_1(1-y_1)^{n-1}dy_1 \right] dx
\end{aligned}
$$

$$+\int_{1/2}^{1} 2n(1-x)\left[\int_{0}^{2x-1} y_1^n dy_1 \right.$$
$$\int_{2x-1}^{x} y_1(1+2y_1-2x)^{n-1}dy_1$$
$$\left.\int_{x}^{1} y_1(1+2x-2y_1)^{n-1}dy_1\right] dx.$$

There are two integrals we must do twice with different bounds. The first is:

$$\int_a^b y_1(1+2y_1-2x)^{n-1}dy_1.$$

We define the function $u(y_1)=1+2y_1-2x$, and thus $y_1=(u+2x-1)/2$ and $dy_1=du/2$. Then the integral is

$$\begin{aligned}\int_a^b y(1+2y_1-2x)^{n-1}dy_1 &= \frac{1}{4}\int_{u(a)}^{u(b)} (u+2x-1)u^{n-1}du \\ &= \frac{1}{4}\left[\frac{2x-1}{n}u^n+\frac{1}{n+1}u^{n+1}\right]_{u(a)}^{u(b)}.\end{aligned}$$

The second general integral is:

$$\int_a^b y_1(1+2x-2y_1)^{n-1}dy_1.$$

We define the function $u(y_1)=1+2x-2y_1$, and thus $y_1=(1+2x-u)/2$ and $dy_1=-du/2$. Then the integral is

$$\begin{aligned}\int_a^b y_1(1+2x-2y_1)^{n-1}dy_1 &= -\frac{1}{4}\int_{u(a)}^{u(b)} (1+2x+u)u^{n-1}du \\ &= -\frac{1}{4}\left[\frac{2x+1}{n}u^n-\frac{1}{n+1}u^{n+1}\right]_{u(a)}^{u(b)}.\end{aligned}$$

We use these general forms to evaluate three of the six components of our full integral for $P_n(e)$:

$$\begin{aligned}\int_0^x y_1(1+2y_1-2x)^{n-1}dy_1 &= \frac{1}{4}\left[\frac{2x-1}{n}u^n-\frac{1}{n+1}u^{n+1}\right]_{1=u(x)}^{1-2x=u(0)} \\ &= \frac{1}{4}\left(\frac{2x+1}{n}+\frac{1}{n+1}\right)+\frac{1}{4}(1-2x)^{n+1}\left(\frac{1}{n}-\frac{1}{n+1}\right)\end{aligned}$$

$$\begin{aligned}
\int_x^{2x} y_1(1+2x-2y_1)^{n-1}dy_1 &= -\frac{1}{4}\left[\frac{2x+1}{n}u^n - \frac{1}{n+1}u^{n+1}\right]_{1=u(x)}^{1-2x=u(2x)} \\
&= \frac{1}{4}\left(\frac{2x+1}{n} - \frac{1}{n+1}\right) - \frac{1}{2n}(1-2x)^n + \frac{1}{4}(1-2x)^{n+1}\left(\frac{1}{n} + \frac{}{r}\right. \\
\int_{2x}^{1} y_1(1-y_1)^{n-1}dy_1 &= \int_0^{1-2x}(1-u)u^{n-1}du = \left[\frac{1}{n}u^n - \frac{1}{n+1}u^{n+1}\right]_0^{1-2x} \\
&= \frac{1}{n}(1-2x)^n - \frac{1}{n+1}(1-2x)^{n+1}.
\end{aligned}$$

We add these three integrals together and find, after a straightforward calculation that the sum is

$$\frac{x}{n} + \frac{1}{2n}(1-2x)^n + \left(\frac{1}{2n} - \frac{1}{n+1}\right)(1-2x)^{n+1}. \qquad (*)$$

We next turn to the three remaining three of the six components of our full integral for $P_n(e)$:

$$\begin{aligned}
\int_0^{2x-1} y_1^n dy_1 &= \frac{1}{n+1}(2x-1)^{n+1} \\
\int_{2x-1}^{x} y_1(1+2y_1-2x)^{n-1}dy_1 &= \frac{1}{4}\left[\frac{2x-1}{n}u^n + \frac{1}{n+1}u^{n+1}\right]_{2x-1=u(2x-1)}^{1=u(x)} \\
&= \frac{1}{4}\left(\frac{2x-1}{n} + \frac{1}{n+1}\right) - \frac{1}{4}(2x-1)^{n+1}\left(\frac{1}{n} + \frac{1}{n+1}\right) \\
\int_x^1 y_1(1+2x-2y_1)^{n-1}dy_1 &= -\frac{1}{4}\left[\frac{2x+1}{n}u^n - \frac{1}{n+1}u^{n+1}\right]_{1=u(x)}^{2x-1=u(1)} \\
&= \frac{1}{4}\left(\frac{2x+1}{n} - \frac{1}{n+1}\right) - \frac{1}{2n}(2x-1)^n - \frac{1}{4}(2x-1)^{n+1}\left(\frac{1}{n} - \right.
\end{aligned}$$

The sum of these latter three terms is

$$\frac{x}{n} - \frac{1}{2n}(2x-1)^n - \left(\frac{1}{2n} - \frac{1}{n+1}\right)(2x-1)^{n+1}. \qquad (**)$$

Thus the sum of the three Case 1 terms is

$$\begin{aligned}
\int_0^{1/2} 2n(1-x)\Bigg[&\int_0^x y_1(1+2y_1-2x)^{n-1}dy_1 \\
&+ \int_x^{2x} y_1(1+2x-2y_1)^{n-1}dy_1 \\
&+ \int_{2x}^{1} y_1(1-y_1)^{n-1}dy_1\Bigg]dx
\end{aligned}$$

$$= \int_0^{1/2} 2n(1-x)\left[\frac{x}{n} + \frac{1}{2n}(1-2x)^n + (1-2x)^{n+1}\left(\frac{1}{2n} - \frac{1}{n+1}\right)\right] dx$$

$$= \int_0^{1} n\left(\frac{1+u}{2}\right)\left[\frac{1-u}{2n} + \frac{1}{2n}u^n + \left(\frac{1}{2n} - \frac{1}{n+1}\right)u^{n+1}\right] du,$$

where we used the substitution $u = 1-2x$, and thus $x = (1-u)/2$, and $dx = -du/2$ and $1-x = (1+u)/2$.

Likewise, we have for the three Case 2 terms

$$\int_{1/2}^{1} 2n(1-x)\left[\int_0^{2x-1} y_1^n dy_1 + \int_{2x-1}^{x} y_1(1+2y_1-2x)^{n-1}dy_1 + \int_x^1 y_1(1+2x-2y_1)^{n-1}dy_1\right] dx$$

$$= \int_{1/2}^{1} 2n(1-x)\left[\frac{x}{n} - \frac{1}{2n}(2x-1)^n - (2x-1)^{n+1}\left(\frac{1}{2n} - \frac{1}{n+1}\right)\right] dx$$

$$= \int_0^{1} n\left(\frac{1-u}{2}\right)\left[\frac{1+u}{2n} - \frac{1}{2n}u^n - \left(\frac{1}{2n} - \frac{1}{n+1}\right)u^{n+1}\right] du,$$

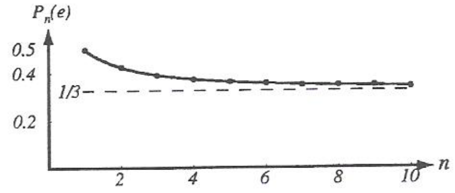
where we used the substitution $u = 2x-1$, and thus $x = (1+u)/2$ and $dx = du/2$ and $1-x = (1-u)/2$.

Now we are ready to put all these results together:

$$\begin{aligned}
P_n(e) &= \int_0^{1} n\left(\frac{1+u}{2}\right)\left[\frac{1-u}{2n} + \frac{1}{2n}u^n + \left(\frac{1}{2n} - \frac{1}{n+1}\right)u^{n+1}\right] du \\
&\quad + \int_0^{1} n\left(\frac{1-u}{2}\right)\left[\frac{1+u}{2n} - \frac{1}{2n}u^n - \left(\frac{1}{2n} - \frac{1}{n+1}\right)u^{n+1}\right] du \\
&= n\int_0^1 \left[\frac{1}{2n}(1-u^2) + \frac{1}{2n}u^{n+1} + \left(\frac{1}{2n} - \frac{1}{n+1}\right)u^{n+2}\right] du \\
&= n\left[\frac{1}{2n}\left(u - \frac{u^3}{3}\right) \frac{1}{2n(n+2}u^{n+2} + \frac{1}{n+3}\left(\frac{1}{2n} - \frac{1}{n+1}\right)u^{n+3}\right]_0^1 \\
&= n\left[\frac{1}{2n}\frac{2}{3} + \frac{1}{2n(n+2)} + \frac{1}{n+3}\left(\frac{1-n}{2n(n+1)}\right)\right] \\
&= \frac{1}{3} + \frac{(n+1)(n+3)-(n-1)(n+2)}{2(n+1)(n+2)(n+3)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{3} + \frac{3n+5}{2(n+1)(n+2)(n+3)} \\
&= \frac{1}{3} + \frac{1}{(n+1)(n+3)} + \frac{1}{2(n+2)(n+3)},
\end{aligned}
$$

which decays as $1/n^2$, as shown in the figure.



We may check this result for the case $n = 1$ where there is only one sample. Of course, the error in that case is $P_1(e) = 1/2$, since the true label on the test point may either match or mismatch that of the single sample point with equal probability. The above formula above confirms this

$$
\begin{aligned}
P_1(e) &= \frac{1}{3} + \frac{1}{(1+1)(1+3)} + \frac{1}{2(1+2)(1+3)} \\
&= \frac{1}{3} + \frac{1}{8} + \frac{1}{24} = \frac{1}{2}.
\end{aligned}
$$

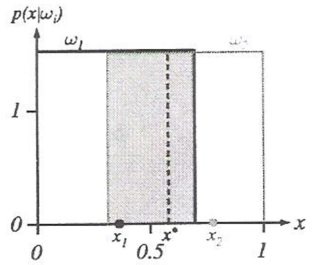
(e) The limit for infinite data is simply

$$\lim_{n\to\infty} P_n(e) = \frac{1}{3},$$

which is larger than the Bayes error, as indeed it must be. In fact, this solution also illustrates the bounds of Eq. 52 in the text:

$$
\begin{aligned}
P^* &\leq P \leq P^*(2 - 2P^*) \\
\frac{1}{4} &\leq \frac{1}{3} \leq \frac{3}{8}.
\end{aligned}
$$

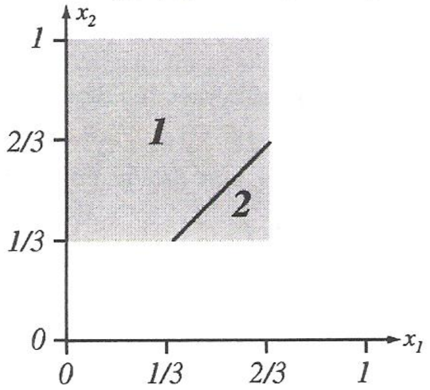
14. We assume $P(\omega_1) = P(\omega_2) = 0.5$ and the distributions are as given in the figure.



(a) This is a somewhat unusual problem in that the Bayes decision can be any point $1/3 \leq x^* \leq 2/3$. For simplicity, we can take $x^* = 1/3$. Then the Bayes error is then simply

$$\begin{aligned} P^* &= \int_0^1 \min[P(\omega_1)p(x|\omega_1), P(\omega_2)p(x|\omega_2)]dx \\ &= \int_{1/3}^{2/3} P(\omega_1)p(x|\omega_1)dx \\ &= 0.5(1/3)(3/2) = 0.25. \end{aligned}$$

(b) The shaded area in the figure shows the possible (and equally likely) values of a point x_1 chosen from $p(x|\omega_1)$ and a point x_2 chosen from $p(x|\omega_2)$.



There are two functionally separate cases, as numbered, corresponding to the position of the decision boundary $x^* = (x_1 + x_2)/2$. (Note that we will also have to consider which is larger, x_1 or x_2. We now turn to the decision rule and probability of error in the single nearest-neighbor classifier in these two cases:

case 1 : $x_2 \geq x_1$ and $1/3 \leq (x_1 + x_2)/2 \leq 2/3$: Here the decision point x^* is between 1/3 and 2/3, with $\mathcal{R}_2$ at large values of x. This is just the Bayes case described in part (a) and the error rate is thus 0.25, as we saw. The relative probability of **case 2** occuring is the relative area of the gray region, that is, 7/8.

case 2 : $x_1 \geq x_2$ and $1/3 \leq (x_1 + x_2)/2 \leq 2/3$: Here the decision boundary is between 1/3 and 2/3 (in the Bayes region) but note especially that $\mathcal{R}_1$ is for large values of x, that is, the decision is the opposite of the Bayes decision. Thus the error is 1.0 minus the Bayes error, or 0.75. The relative probability of **case 2** occuring is the relative area of the gray region, that is, 1/8.

We calculate the average error rate in the case of one point from each category by merely adding the probability of occurrence of each of the three cases (proportional to the area in the figure), times the expected error given that case, that is,

$$P_1 = \frac{7}{8}0.25 + \frac{1}{8}0.75 = \frac{5}{16} = 0.3125,$$

which is of course greater than the Bayes error.

(c) PROBLEM NOT YET SOLVED

(d) PROBLEM NOT YET SOLVED

(e) In the limit $n \to \infty$, every test point x in the range $0 \leq x \leq 1/3$ will be properly classified as ω_1 and every point in the range $2/3 \leq x \leq 1$ will be properly classified as ω_2. Test points in the range $1/3 \leq x \leq 2/3$ will be misclassified half of the time, of course. Thus the expected error in the $n \to \infty$ case is

$$\begin{aligned} P_\infty &= P(\omega_1)\mathrm{Pr}[0 \leq x \leq 1/3|\omega_1] \cdot 0 + P(\omega_1)\mathrm{Pr}[1/3 \leq x \leq 2/3|\omega_1] \cdot 0.5 \\ &\quad + P(\omega_2)\mathrm{Pr}[1/3 \leq x \leq 2/3|\omega_2] \cdot 0.5 + P(\omega_2)\mathrm{Pr}[2/3 \leq x \leq 1|\omega_2] \cdot 0 \\ &= 0.5\ 0.5\ 0.5 + 0.5\ 0.5\ 0.5 = 0.25. \end{aligned}$$

Note that this is the same as the Bayes rate. This problem is closely related to the "zero information" case, where the posterior probabilities of the two categories are equal over a range of x. If the problem specified that the distributions were equal throughout the full range of x, then the Bayes error and the P_∞ errors would equal 0.5.

15. An faster version of Algorithm 3 in the text deletes prototypes as follows:

Algorithm 0 (Faster nearest-neighbor)

begin initialize $j \leftarrow 0, \mathcal{D}, n =$ number of prototypes
  Construct the full Voronoi diagram of $\mathcal{D}$
do $j \leftarrow j + 1$ (for each prototype $\mathbf{x}'_j$)
if $\mathbf{x}'_j$ is not marked **then** find the Voronoi neighbors of $\mathbf{x}'_j$
  if any neighbor is not from the$\mathbf{x}'_j$class **then** mark $\mathbf{x}'_j$ and its neighbors in other classes
  until $j = n$
Discard all unmarked prototypes
returnVoronoi diagram of the remaining (marked) prototypes
end

If we have k Voronoi neighbors on average of any point $\mathbf{x}'_j$, then the probability that i out of these k neighbors are not from the same class as $\mathbf{x}'_j$ is given by the binomial law:

$$P(i) = \binom{k}{i}(1 - 1/c)^k (1/c)^{k-1},$$

where we have assumed that all the classes have the same prior probability. Then the expected number of neighbors of any point $\mathbf{x}'_j$ belonging to different class is

$$E(i) = k(1 - 1/c).$$

Since each time we find a prototype to delete we will remove $k(1-1/c)$ more prototypes on average, we will be able to speed up the search by a factor $k(1 - 1/c)$.

16. Consider Algorithm 3 in the text.

(a) In the figure, the training points (black for ω_1, white for ω_2) are constrained to the intersections of a two-dimensional grid. Note that prototype f does not contribute to the class boundary due to the existence of points e and d. Hence f should be removed from the set of prototypes by the editing algorithm (Algorithm 3 in the text). However, this algorithm detects that f has a prototype