# CSE 569 Homework #3 Solutions

## Problem 1 Solution:

(a) For any $x$, the probability of $x$ falling into bin $B_j$ is estimated by the normalized histogram value for that bin, i.e., $P(x \in B_j) = \frac{Y_j}{n}$. Accordingly, we estimate the density for that bin as $\hat{p}(x) = \frac{Y_j}{n} m$. (since $\int_{B_j} \hat{p}(x)dx = P(x \in B_j)$)

To write this for any $x$ that may fall into any bin, we may have

$$\hat{p}(x) = \sum_{j=1}^{m} \frac{mY_j}{n} I(x \in B_j), \text{ where } I(x \in B_j) = \begin{cases} 1, & \text{if } x \in B_j \\ 0, & \text{o.w.} \end{cases},$$

(b) $E(\hat{p}(x)) = E\left(\sum_{j=1}^{m} \frac{mY_j}{n} I(x \in B_j)\right)$. [Note: we take the expectation because $\hat{p}(x)$ depends on $D$; The expectation is NOT with respect to $x$; that's the reason the problem states "for a given $x$".]

For a given $x$, it should belong to some $B_j$,

thus $E(\hat{p}(x)) = E\left[\frac{mY_j}{n}\right] = \frac{m}{n} E\{Y_j\}$, for some $j$.

What is $E\{Y_j\}$? See slides 4-5 of Notes05, $Y_j$ is binomial with a parameter $p = \int_{B_j} p(x)dx$, and thus its mean is $np = n\int_{B_j} p(x)dx$

$\therefore E(\hat{p}(x)) = \frac{m}{n} E\{Y_j\} = \frac{m}{n} \times n \int_{B_j} p(x)\,dx$

$$= m \int_{B_j} p(x)dx$$

(You can stop here.)

But, what's good about this $E[\hat{p}(x)]$?

If $m$ is big, or $\frac{1}{m}$ (the bin size) is small, $\int_{B_j} p(x)dx \approx \frac{1}{m} p(x')$ for some $x' \in B_j$

$\therefore E\{\hat{p}(x)\} \approx m \times \frac{1}{m} p(x') = p(x')$ for $x' \in B_j$

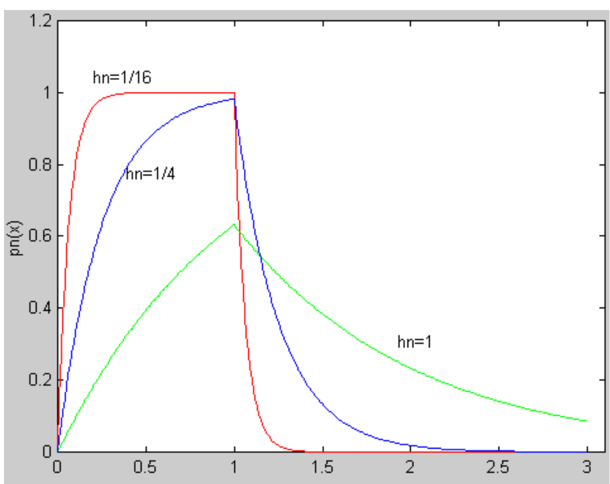
This shows $\hat{p}(x)$ converges to $p(x)$ when $m \to \infty$.

**Problem 2 Solution**:

(a) The mean (expected value) of the Parzen window estimate is computed as (assuming that the samples are i.i.d random variables):

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

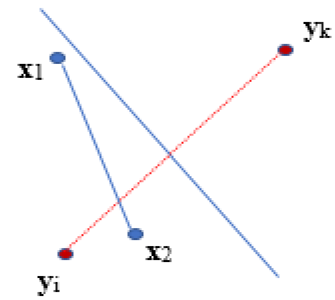
(b) When a=1, we have the following plot




(c) From the figure in (b), it is clear that, in 0<*x*<*a*, when $h_n$ gets very small, the bias really happens only towards the 0-end of the interval. So we really want there is only <=0.01 error in the interval of [0.01, 1) (since "over 99%" of the range). Using this, plus the result of (a), you can figure out $h_n$.

(d) This is straightforward after (c).

**Problem 4 Solution:**




Consider two adjacent training sample $\mathbf{y}_i$. and $\mathbf{y}_k$, which define a segment of the border of the Voronoi cell of $\mathbf{y}_i$. This segment of the border should be part of the hyperplane that is orthogonal to the line $\mathbf{y}_i$ - $\mathbf{y}_k$ and passes through the middle point between $\mathbf{y}_i$. and $\mathbf{y}_k$, as illustrated to the right. That is, $\mathbf{y}_i$'s Voronoi cell lies on one side of the above hyperplane. The half spaces separated by a hyperplane are convex (see below "Note"). The argument is true for $\mathbf{y}_i$ and all its other immediate neighbors. So $\mathbf{y}_i$'s Voronoi cell should be convex on all sides.

Note: To explicitly show that for any two points $\mathbf{x}_1$ and $\mathbf{x}_2$ lying on the $\mathbf{y}_i$ side of the above hyperplane, any point $\mathbf{x}^* = \alpha\, \mathbf{x}_1 + (1-\alpha)\, \mathbf{x}_2$, for $0< \alpha < 1$, also lies on the same side, we can write the hyperplane equation as $g(\mathbf{x})=\mathbf{w}^t\mathbf{x} + w_0 = 0$. Then $\mathbf{x}_1$ and $\mathbf{x}_2$ lying on the same side of the hyperplane means both $g(\mathbf{x}_1)<0$ and $g(\mathbf{x}_2) <0$ (or both $>0$, as long as they have the same sign). Then we can show that $g(\mathbf{x}^*) = g(\alpha\, \mathbf{x}_1 + (1-\alpha)\, \mathbf{x}_2) = \mathbf{w}^t[\alpha\, \mathbf{x}_1 + (1-\alpha)\, \mathbf{x}_2] + w_0 = \alpha\, (\mathbf{w}^t\mathbf{x}_1 + w_0) + (1-\alpha)(\mathbf{w}^t\mathbf{x}_2 + w_0) = \alpha\, g(\mathbf{x}_1) + (1-\alpha)g(\mathbf{x}_2)$, which will have the same sign as $g(\mathbf{x}_1)$ (and $g(\mathbf{x}_2)$), for $0< \alpha$. So $\mathbf{x}^*$ is on the same side of the hyperplane.

**Problem 5. Computer Exercise: Outline of the steps to develop a solution**

Let X be a random variable representing the samples of $D_1$, and Y of $D_2$. So X ~ $N(\mu_X, 1)$, Y~$N(\mu_Y, 1)$. Let Z=Y-X, then Z ~ $N(\mu_Y - \mu_X, 2)$.

Note: (1) the variance of Z is doubled; (2) for simulating the data $D_1$ and $D_2$, we assume we know $\mu_X = 1$, and $\mu_Y = 1.5$.

For given $n$ samples, the sample mean Z_bar = Y_bar – X_bar, where Y_bar and X_bar are respectively the sample mean of $D_2$ and $D_1$.

Under the $H_0$ hypothesis "the means for X and Y are the same", we compute q as

$q$ = ( Z_bar- ($\mu_Y$- $\mu_X$) ) / (*sqrt*(2)/*sqrt*($n$))

= ( Z_bar - 0 ) / (*sqrt*(2)/*sqrt*($n$)); // since we assume $\mu_Y = \mu_X$

and this $q$ should follow the $N(0, 1)$ distribution (see lecture slides), under $H_0$.

From a standard $N(0, 1)$ table, we know for a significant level 0.05, the acceptance region is [-1.967, 1.967] (see lecture slides), and we would have 0.95 probability mass for $N(0, 1)$ on this interval.

So, in the simulation, if the computed q falls in the above region, we accept $H_0$ with significance level 0.05 (or, roughly, we believe $D_1$ and $D_2$ are from a normal density with a common mean and a known variance 1, with 95% confidence). Otherwise, we reject $H_0$.

For the given parameter settings (with $\mu_X = 1$, $\mu_Y = 1.5$, and three different n values), the answer/results may vary for different datasets (since your random number generator will give you different $D_1$ and $D_2$).

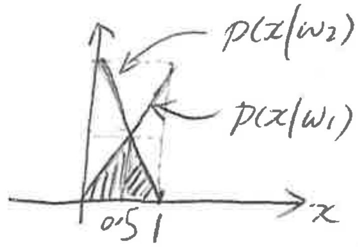
That is, sometimes you accept $H_0$, and sometimes you reject $H_0$, even if the parameters are fixed (but with different datasets). But if you repeat the experiment many times (under the same setting but with different datasets), you may find:

(1) with n=20, you might accept $H_0$ very often.
(2) with n=100, you may accept $H_0$ much less often.
(3) with n=600, you rarely or even never accept $H_0$.

This should be intuitive: with more data, we become more confident that $D_1$ and $D_2$ do not have the same mean.

**Problem 3 Solution**: See the following scanned pages.

(a) The densities are illustrated below.




Since the priors are equal, we decide $w_1$ when

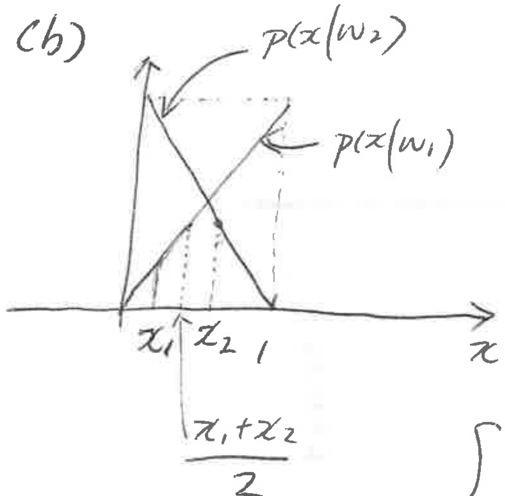
$$p(x|w_1) > p(x|w_2)$$

This gives us the decision boundary $x = 0.5$, i.e.,

Decide $w_1$ if $x > 0.5$; o.w. decide $w_2$.

The Bayes error is the area of the shaded region (weighted by the priors), and thus the Bayes error is $\frac{1}{4}$.

(b)




If we are given $x_1$ from $w_1$, $x_2$ from $w_2$, and $x_1 < x_2$, as illustrated below, then clearly the decision rule is:

Decide $w_1$ if $x < \frac{x_1+x_2}{2}$ ☆

O.W. Decide $w_2$.

[Note: $\frac{x_1+x_2}{2}$ is the middle point between $x_1$ and $x_2$, and any $x$ less than $\frac{x_1+x_2}{2}$ will be closer to $x_1$ than to $x_2$. Hence the rule ☆ above.]

In this case, the error probability is:

$$P(w_1)\,P[\text{error for } w_1] + P(w_2)\,P[\text{error for } w_2]$$

$$= \frac{1}{2}\int_{\frac{x_1+x_2}{2}}^{1} p(x|w_1)\,dx + \frac{1}{2}\int_{0}^{\frac{x_1+x_2}{2}} p(x|w_2)\,dx$$

$$= \frac{1}{2}\int_{\frac{x_1+x_2}{2}}^{1} 2x\,dx + \frac{1}{2}\int_{0}^{\frac{x_1+x_2}{2}} 2(1-x)\,dx$$

$$= \frac{1}{2}\,x^2\Big|_{\frac{x_1+x_2}{2}}^{1} + \frac{1}{2}\left[2x - x^2\right]\Big|_{0}^{\frac{x_1+x_2}{2}} = \frac{1}{2} + \frac{x_1+x_2}{2} - \frac{(x_1+x_2)^2}{4} \overset{\text{fixed}}{\triangleq} P_{NN}(e)$$

This error is in general larger than $\frac{1}{2}$, since

$$0 \le \frac{x_1+x_2}{2} \le 1 \Rightarrow \frac{x_1+x_2}{2} \ge \left(\frac{x_1+x_2}{2}\right)^2 \Rightarrow \frac{x_1+x_2}{2} - \frac{(x_1+x_2)^2}{4} \ge 0.$$

Thus $P_{NN}^{fixed}(e) = \frac{1}{2} + \frac{x_1+x_2}{2} - \frac{(x_1+x_2)^2}{4} \ge \frac{1}{2}$.

Note, the given $x_1$ & $x_2$ with $x_1 < x_2$ are really "poor" training samples, since frome the PDFs, $p(x|\omega_1)$ should have higher probability giving larger values. So, chances are, if we are given $x_1$ from $\omega_1$ and $x_2$ from $\omega_2$, in most cases we should have $x_1 > x_2$. In that case, the error should be less than the case of $x_1 < x_2$.

(c) Continuing the discussion above, when given $x_1$ from $\omega_1$ and $x_2$ from $\omega_2$, we really have two cases: $x_1 < x_2$ or $x_1 \ge x_2$ (We arbitarily assign the "=" case to "≥", which is not important for continuous PDFs.)

Therefore, the probability of error should be the weighted averaged errors of both cases, with the weights being the probabilities of the cases, i.e.,

$$P_{NN}(e) = P(x_1 < x_2)\,P[\text{error} \mid x_1 < x_2] + P[x_1 \ge x_2]\,P[\text{error} \mid x_1 \ge x_2]$$

To actually figure out each term, we'll have to do much more. Let's go step by step.

Step 1. Find $P[x_1 < x_2]$. (Accordingly, $P[x_1 \geq x_2] = 1 - P[x_1 < x_2]$)

As we discussed above, for the given PDFs, intuitively we know that in most case $x_1 > x_2$. So we expect $P[x_1 < x_2]$ should be far less than 0.5. But what is its exact value? To find this, we can define a new random variable $Z \triangleq X_2 - X_1$. Then we can find the PDF for $Z$. For doing this, you need to review your probability book on "PDFs of functions of random variables". After we figure out the PDF for $Z$, $\int_{z>0} P_Z(z)\,dz$ will give us the probability of $X_1 < X_2$.

For this problem, we will find $P[x_1 < x_2] = \frac{1}{6}$. (so $P[x_1 \geq x_2] = \frac{5}{6}$)

Step 2. Find $P[\text{error} \mid x_1 < x_2]$

In Part (b), we already see that, given fixed $x_1$ & $x_2$, with $x_1 < x_2$, the error is $P_{NN}^{\text{fixed}}(e)$. In general, even if given $x_1 < x_2$, $x_1$ & $x_2$ can still take various values (as long as $x_1 < x_2$), therefore $P_{NN}^{\text{fixed}}(e)$ will take different values depending on the specific $x_1$ & $x_2$ values. So $P[\text{error} \mid x_1 < x_2] = E\left[P_{NN}^{\text{fixed}}(e)\right]$ and the expectation is over all possible $x_1 < x_2$ cases.

Note that $P_{NN}^{\text{fixed}}(e)$ is actually only dependent of $\frac{x_1 + x_2}{2} \triangleq W$.

Then $W$ is a new random variable, whose PDF can be determined.

then $E\left[P_{NN}^{\text{fixed}}(e)\right] = \int P_{NN}^{\text{fixed}}(e)\, f_W(w)\, dw$.

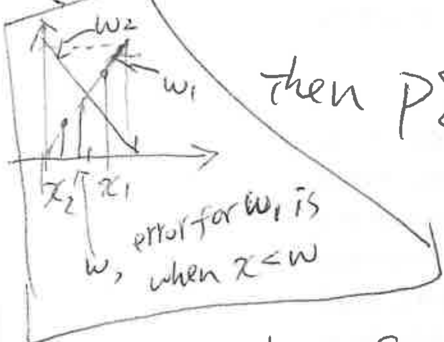
You can find out $f_W(w) = \begin{cases} 16w^2 - \frac{32}{3}w^3, & 0 \leq w \leq \frac{1}{2} \\ 16(1-w)^2 - \frac{32}{3}(1-w)^3, & \frac{1}{2} \leq w \leq 1. \end{cases}$

See earlier note on "PDFs of functions of random variables"

**Step 3**. Find $P[\text{error} \mid x_1 \geq x_2]$.

This is similar to Step 2, except that $P_{NN}^{fixed}(e)$ will be different than what used in step 2. Basically, the new

$$P_{NN}^{fixed2}(e) = \frac{1}{2}\int_0^{\frac{x_1+x_2}{2}} 2x\,dx + \frac{1}{2}\int_{\frac{x_1+x_2}{2}} 2(1-x)\,dx = \frac{1}{2} - w + w^2, \quad w = \frac{x_1+x_2}{2}$$




then $$P[\text{error} \mid x_1 \geq x_2] = E\left[P_{NN}^{fixed2}(e)\right] = \int P_{NN}^{fixed2}(e) f_W(w)\,dw.$$

where $f_W(w)$ is the PDF for $W \triangleq \frac{x_1+x_2}{2}$, as in Step 2.

BTW, $f_W(w)$ on $[0, \frac{1}{2}]$ is symmetric to $f_W(w)$ on $[\frac{1}{2}, 1]$

In Step 2, by pluggin in $P_{NN}^{fixed}(e)$ & $f_W(w)$, you will find

$$E\left[P_{NN}^{fixed}(e)\right] = \frac{13}{18}.$$

In Step 3, similarly, you will find $E\left[P_{NN}^{fixed2}(e)\right] = \frac{15}{18}$

**Step 4**. Finally, we have

$$P_{NN}(e) = \frac{1}{6} \times \frac{13}{18} + \frac{5}{6} \times \frac{15}{18} = \frac{38}{108} \simeq 0.35.$$

This is the NN-rule error rate (for the case of having 1 sample from each class)