Problem Set 2

Patrick Altmeyer

10 February, 2021

1 Piece-wise CDFs

We are given the conditional CDFs

CDFs:

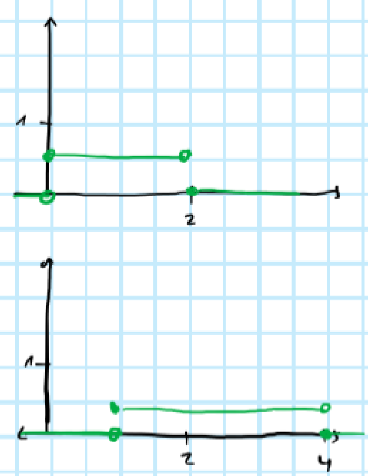
$$\mathbb{P}(X < x \mid Y=0) = \begin{cases} 0 & \text{if } x \le 0 \\ \frac{x}{2} & \text{if } 0 < x \le 2 \\ 1 & \text{if } x > 2 \end{cases}$$

$$\mathbb{P}(X < x \mid Y=1) = \begin{cases} 0 & \text{if } x \le 1 \\ \frac{x-1}{3} & \text{if } 1 < x \le 4 \\ 1 & \text{if } x > 4 \end{cases}$$

from which the conditional PDFs can be derived:

$$f(x \mid Y=0) = \begin{cases} \frac{1}{2} & \text{if } 0 < x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

$$f(x \mid Y=1) = \begin{cases} \frac{1}{3} & \text{if } 1 < x \le 4 \\ 0 & \text{otherwise} \end{cases}$$



Then we have for the joint PDF:

Taking it together:

$$f(x) = f(x \mid Y=0)\,\mathbb{P}(\overset{\frac{1}{2}}{\cancel{Y}=0}) + f(x \mid Y=1)\,\mathbb{P}(\overset{\frac{1}{2}}{\cancel{Y}=1})$$

$$f(x) = \begin{cases} 0.5\,\cancel{\tfrac{1}{2} + 0.5}\cdot 0 \quad \tfrac{6}{24} & \text{if} \quad 0 < x \leq 1 \\ 0.5\,\cancel{\tfrac{1}{2} + 0.5}\,\tfrac{1}{3} \quad \tfrac{10}{24} & \text{if} \quad 1 < x \leq 2 \\ 0.5 \cdot 0 + 0.5 \cdot \tfrac{1}{3} \quad \tfrac{4}{24} & \text{if} \quad 2 < x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

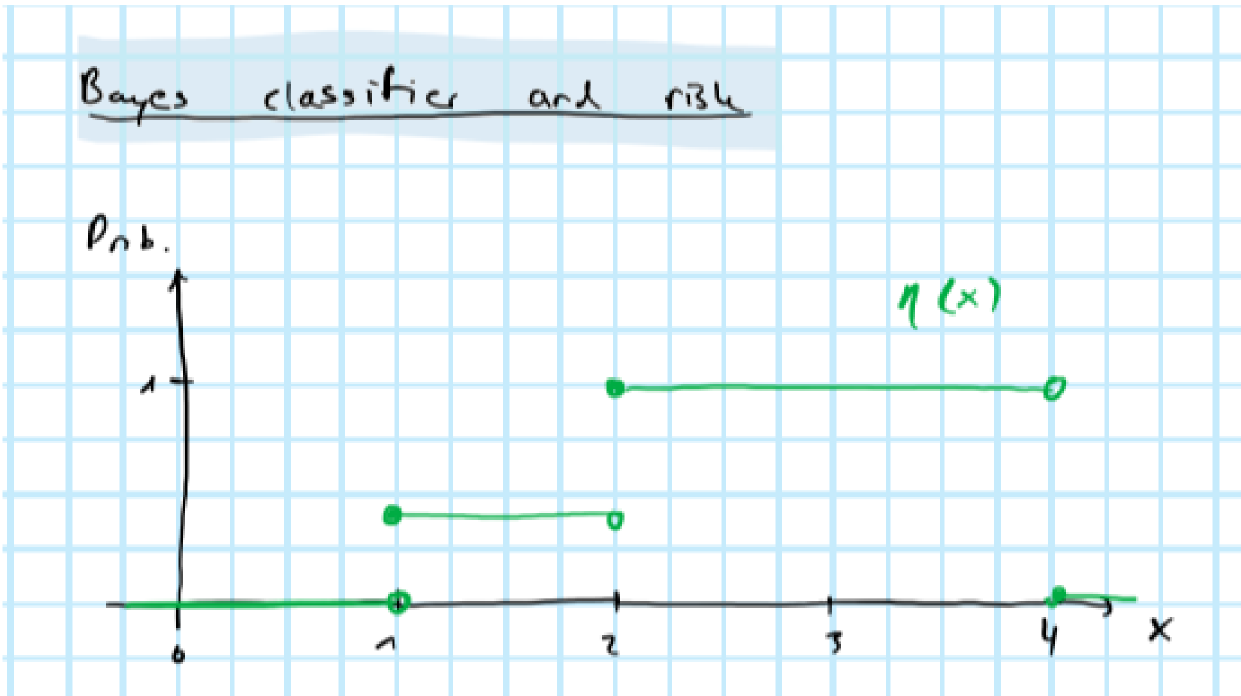
Using Bayes rule we can then determine a functional form for $\eta(\mathbf{X})$

Bayes rule

$$\eta(x) = \frac{\mathbb{P}(X=x \mid Y=1)\,\mathbb{P}(Y=1)}{\mathbb{P}(X=x)}$$

$$\eta(x) = \begin{cases} \overbrace{0}^{P(x \mid Y=1)} \cdot \overbrace{\tfrac{1}{2}}^{P(Y=1)} \cdot \overbrace{\tfrac{24}{6}}^{P(x)^{-1}} = 0 & \text{if} \quad 0 < x \leq 1 \\ \tfrac{1}{3} \cdot \tfrac{1}{2} \cdot \tfrac{24}{10} = 0.4 & \text{if} \quad 1 < x \leq 2 \\ \tfrac{1}{3} \cdot \tfrac{1}{2} \cdot \tfrac{24}{4} = 1 & \text{if} \quad 2 < x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

which can be illustrated as follows:



1.1 Bayes classifier and risk

Then we have for the Bayes classifier and corresponding risk:

Classifier :

$$g^*(x) = \begin{cases} 1 & \text{if} \quad 2 < x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

Bayes risk :

$$R^* = \mathbb{E}\left[\mathbb{1}_{\eta(x) \leq \frac{1}{2}} \, \eta(x) + \mathbb{1}_{\eta(x) > \frac{1}{2}} \, (1 - \eta(x)) \right]$$

$$= \mathbb{E} \min \left[\eta(x), (1 - \eta(x)) \right]$$

$$R^* = \mathbb{E} \begin{cases} 0.4 & 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

$$= \underbrace{\frac{5}{12}}_{\mathbb{P}(1 < x \leq 2)} \cdot \frac{4}{10} = \frac{20}{120} = \frac{1}{6}$$

1.2 1-NN

For the 1-NN we can not that the following holds asymptotically:

Since $\eta(x)$ is (piece-wise) continuous we have that

$$d\left(x, x_{(1)}(x)\right) \to 0$$

$$\Rightarrow \qquad x \approx x_{(1)}(x)$$

$$\Rightarrow \qquad \eta(x) \approx \eta\left(x_{(1)}(x)\right)$$

for sufficiently large n (i.e. asymptotically).

So, the probability of error is

$$R(g_n) = \mathbb{P}\left(\underbrace{Y_{(1)}(x)}_{\text{label of nearest neighbour}} \neq Y \mid D_n\right) = \mathbb{E}\left[\mathbb{P}\left(Y_{(1)}(x) \neq Y \mid X, D_n\right) \mid D_n\right]$$

$$\approx \mathbb{P}\left(Y' \neq Y \mid X\right)$$

where Y', Y have the same distribution given X and Y' is conditionally independent of Y

We have $\left.\begin{array}{l} Y \sim \mathrm{Bern}(\eta) \\ Y' \sim \mathrm{Bern}(\eta) \end{array}\right\}$

$$\begin{aligned} \mathbb{P}(Y \neq Y') &= \mathbb{P}(Y = 1, Y' = 0) + \mathbb{P}(Y = 0, Y' = 1) \\ &= \eta(1-\eta) + \eta(1-\eta) \\ &= 2\eta(1-\eta) \end{aligned}$$

And consequently applying this here we get:

And hence

$$R(g_n^{1NN}) \approx \mathbb{E}\left(2\eta(x)\left(1-\eta(x)\right)\right)$$

Plugging in for η

$$R^{1NN}(g_n) = \mathbb{E}\begin{cases} \overset{0.48}{\overline{2 \cdot 0.4 \cdot 0.6}} & \text{if } 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases} = \frac{5}{12} \cdot 0.48 = \frac{\overset{1}{\cancel{50}}}{\underset{\cancel{10}\,5}{\cancel{120}}} \cdot \frac{\overset{\cancel{4}\,2}{\cancel{48}}}{\underset{2}{\cancel{100}}} = \frac{2}{10}$$

1.3 3-NN

Similarly, we can show for the 3-NN rule:

3NN - classifier

By the same logic as for 1NN:

$$\mathbb{P}(\text{majority}(Y_1', Y_2', Y_3') \neq Y) =$$
$$= \mathbb{P}(Y=1, \text{maj}(Y_1', Y_2', Y_3') = 0) + \mathbb{P}(Y=0, \text{maj}(Y_1', Y_2', Y_3') = 1)$$
$$= \eta \, \mathbb{P}(\text{maj}(Y_1' \dots) = 0) + (1-\eta)\, \mathbb{P}(\text{maj}(Y_1' \dots) = 1)$$
$$= \eta\left[(1-\eta)^3 + 3(1-\eta)^2 \eta\right] + (1-\eta)\left[\eta^3 + 3(1-\eta)\eta^2\right]$$
$$= \eta(1-\eta)\left[(1-\eta)^2 + 6\eta(1-\eta) + \eta^2\right]$$

So, the asymptotic prob. of error of the 3NN classifier is

$$R^{3NN} = \lim_{n\to\infty} \mathbb{E}(R(g_n)) = \mathbb{E}\left[\eta(x)(1-\eta(x))\left[(1-\eta(x))^2 + \eta(x)^2 + 6\eta(x)(1-\eta(x))\right]\right]$$
$$= \mathbb{E}\left[\eta(x)(1-\eta(x))\right] + 4\,\mathbb{E}\left[\eta(x)^2(1-\eta(x))^2\right]$$

$$R^{3NN}(g_n) = \mathbb{E}\begin{cases} 0.24 + 4\left[0.4^2\, 0.6^2\right] = 0.4704 & \text{if } 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

$$= \frac{5}{12} \cdot 0.4704 = 0.196$$

2 Nearest neighbor regression

2.1 Maths

We can derive the optimal predictor as follows:

$$f^* = \arg\min_f \; R(f)$$

$$= \arg\min_f \; \mathbb{E}\,[f(x) - Y]^2$$

$$= \arg\min_f \; \mathbb{E}_x\left[\mathbb{E}_{Y|x}\left[[f(x) - Y]^2 \mid x\right]\right]$$

$$f^* = \arg\min_f \; \mathbb{E}_{Y|x}[(f - Y)^2 \mid x]$$

FOC: $$\frac{d}{df}\,\mathbb{E}_{Y|x}[(f - Y)^2 \mid x] = \mathbb{E}_{Y|x}[2f - 2Y \mid x] = 0$$

$$\Leftrightarrow \quad 2f - \mathbb{E}_{Y|x}[2Y \mid x] = 0$$

$$\Rightarrow \quad f^* = \mathbb{E}_{Y|x}[Y \mid x]$$

In other words, the optimal predictor f^* of $\mathbf{Y}$ given $\mathbf{X}$ is just the conditional mean of $\mathbf{Y}$ given $\mathbf{X}$.

Then the Bayes risk just corresponds to the irreducible error $R^* = \sigma_\varepsilon^2$:

The Bayes risk is then

$$R^* = R(f^*) = \mathbb{E}\left[(f^*(x) - Y)^2 \mid x\right]$$

$$= \mathbb{E}\left[(Y - \mathbb{E}(Y))^2 \mid x\right]$$

$$= \mathbb{E}\left[(Y - [Y + \varepsilon])^2 \mid x\right]$$

$$= \mathbb{E}\left[\varepsilon^2 \mid x\right]$$

$$= \sigma_\varepsilon^2$$

Now note that for any KNN regressor we can decompose its risk as follows:

$$f_n^{kNN}(x) = \frac{1}{k}\sum_{\ell=1}^{k} Y_{(\ell)}(x)$$

Then

$$R(f_n^{kNN}(X)) = \mathbb{E}\left[(Y - f_n^{kNN}(X))^2 \mid X = x\right] =$$

$$= \mathbb{E}\left[([f(X)+\varepsilon] - f_n^{kNN}(X))^2 \mid X = x\right]$$

$\mathbb{E}(\varepsilon \mid X=x)=0$

$$= \mathbb{E}\left[(f(X)+\varepsilon)^2 - 2(f(X)+\varepsilon)f_n^{kNN}(X) + f_n^{kNN}(X)^2 \mid X = x\right]$$

$$= \mathbb{E}\left[f(X)^2 + \varepsilon^2 - 2f(X)f_n^{kNN}(X) + f_n^{kNN}(X)^2 \mid X = x\right]$$

$\mathbb{E}(\varepsilon^2 \mid X=x) = \sigma_\varepsilon^2$

$$= \sigma_\varepsilon^2 + \mathbb{E}\left[(f_n^{kNN}(X) - f(X))^2 \mid X = x\right]$$

$$= \sigma_\varepsilon^2 + \mathbb{E}\left[(f_n^{kNN}(X) - \mathbb{E} f_n^{kNN}(X) + \mathbb{E} f_n^{kNN}(X) - f(X))^2 \mid X = x\right]$$

cross-term is zero

$$= \sigma_\varepsilon^2 + \mathbb{E}\left[(f_n^{kNN}(X) - \mathbb{E} f_n^{kNN}(X))^2 + (\mathbb{E} f_n^{kNN}(X) - f(X))^2 \mid X = x\right]$$

$$= \sigma_\varepsilon^2 + \mathbb{E}\left[(f_n^{kNN}(X) - \mathbb{E} f_n^{kNN}(X))^2 \mid X = x\right] + \mathbb{E}\left[(\mathbb{E} f_n^{kNN}(X) - f(X))^2 \mid X = x\right]$$

$$= \underbrace{\sigma_\varepsilon^2}_{\text{irreducable error}} + \underbrace{(\mathbb{E} f_n^{kNN}(X) - f(X))^2}_{\text{sq. bias}} + \underbrace{\mathbb{E}\left[(f_n^{kNN}(X) - \mathbb{E} f_n^{kNN}(X))^2 \mid X = x\right]}_{\text{variance}}$$

where for the variance term we can further simplify:

$$\mathrm{Var}(f_n^{kNN}(x)) = \mathrm{Var}\left(\frac{1}{k}\sum_{\ell=1}^{k} Y_{(\ell)}(x)\right)$$

$$= \frac{1}{k^2}\,\mathrm{var}\left(\sum_{\ell=1}^{k} Y_{(\ell)}(x)\right)$$

$$\overset{\perp\!\!\!\perp}{=} \frac{1}{k^2}\sum_{\ell=1}^{k} \mathrm{var}(Y_{(\ell)}(x))$$

$$\overset{\text{ident. distr.}}{=} \frac{1}{k}\,\sigma_\varepsilon$$

Then asymptotically we can show:

$$\lim_{n\to\infty} \mathbb{E} R(f_n^{KNN}) = \sigma^2_\varepsilon + \lim_{n\to\infty} \left(\mathbb{E} f_n^{KNN}(x) - f(x)\right)^2 + \frac{\sigma_\varepsilon^2}{k}$$

$$= \sigma^2_\varepsilon \left(\frac{k+1}{k}\right) + \lim_{n\to\infty} \left(\mathbb{E}\left[\frac{1}{k}\sum_{\ell=1}^{k} Y_{(\ell)}(x)\right] - Y_1\right)^2$$

$$= \sigma^2_\varepsilon \left(\frac{k+1}{k}\right) + \lim_{n\to\infty} \left(d\left(\mathbb{E}\left[\frac{1}{k}\sum_{\ell=1}^{k} Y_{(\ell)}(x)\right], Y_1\right)^2 \right)$$

Note that $d\left(\mathbb{E}\left[\frac{1}{k}\sum_{\ell=1}^{k} Y_{(\ell)}(x)\right], Y_1\right) \to 0$ as $n \to \infty$

Then it is finally easy to see the asymptotic risks of the 1NN and KNN regressor:

For 1NN

$$\lim_{n\to\infty} \mathbb{E} R(f_n^{1NN}) = 2\sigma^2_\varepsilon = 2R^*$$

For KNN

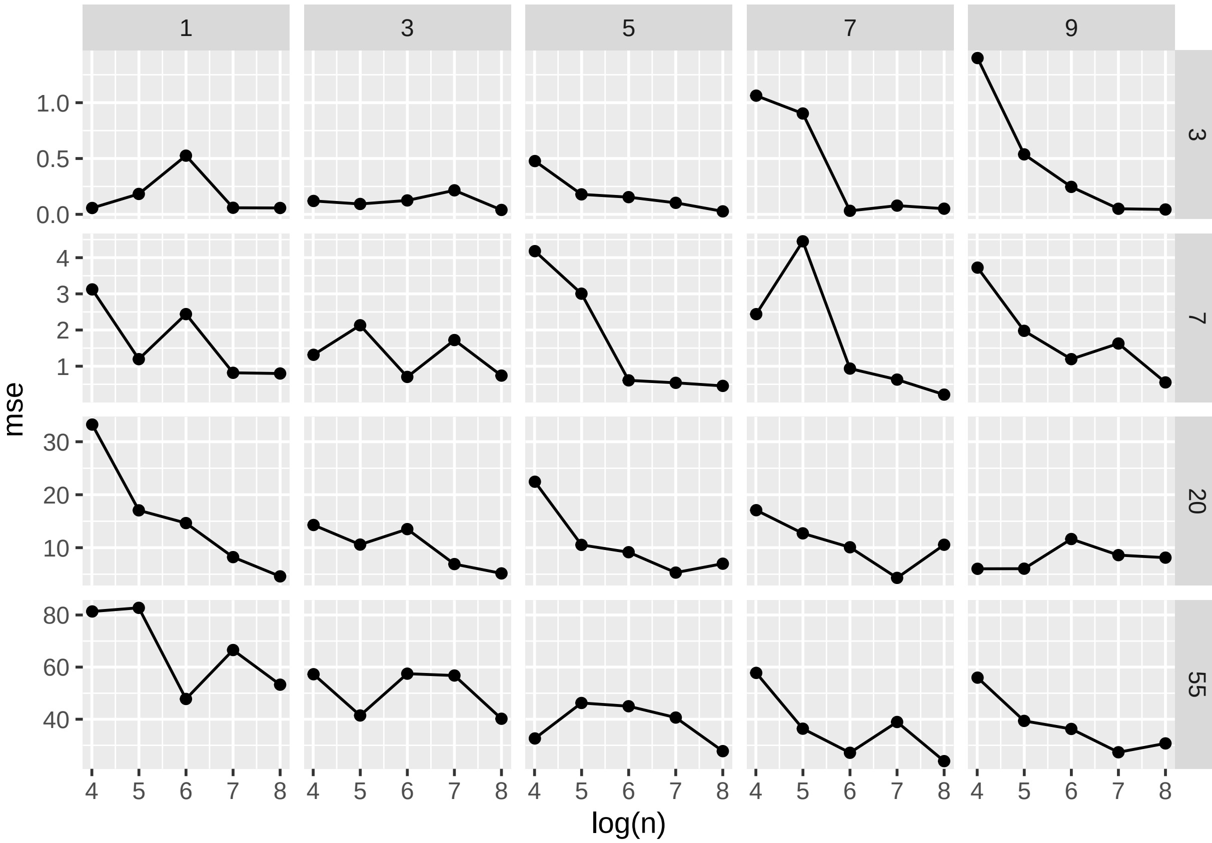
$$\lim_{n\to\infty} \mathbb{E} R(f_n^{KNN}) = \left(\frac{k+1}{k}\right)\sigma^2_\varepsilon = \left(\frac{k+1}{k}\right) R^*$$

2.2 Program

The KNN regressor can be implemented in R as follows:

```
knn_regressor <- function(X,y,k, ...) {
  row_idx <- 1:nrow(X)
  distances <- data.table(t(combn(row_idx,2)))
  distances[,dist:=c(dist(X,...))]
  distances_rev <- copy(distances)
  setnames(distances_rev, c("V1", "V2"), c("V2", "V1"))
  distances <- rbind(distances, distances_rev)
  setorder(distances, V1, dist)
  setnames(distances, c("V1", "V2"), c("X", "neighbour"))
  distances[,y_neighbour:=y[neighbour]]
  fitted <- distances[,mean(y_neighbour[1:k]),by=X]$V1
  return(fitted)
}
```

Using this regressor I simulate data multiple times, fit and predict from the regressor and compute the mean squared error each time. I vary the dimensions d, the number of neighbours k and the sample size n.



3 Bayes risk

4 NN for binary classification

4.1 Maths

Bayes risk and the asymptotic risks of the different KNN classifiers involves an expectation with respect to $\eta(\mathbf{X}) = \frac{x^{(1)}+x^{(2)}}{2}$. To solve for that we first need to know the density of $z = x^{(1)} + x^{(2)}$. It turns out that z follows a *triangular* distribution with density

$$f(z) = \begin{cases} z & \text{if} \quad 0 < z < 1 \\ 2 - z & \text{if} \quad 1 \leq z < 2 \\ 0 & \text{otherwise} \end{cases}$$

This can be shown as follows:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx$$

$$= \int_{-\infty}^{\infty} f(x)\, f(z-x)\, dx \qquad \text{since } f_X = f_Y$$

Since $X, Y \sim unif(0,1)$, this implies that

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

So the integrand $f(x) f(z-x) \in \{0, 1\}$

More specifically,

$$f(x)\, f(z-x) = \begin{cases} 1 & \text{if } 0 < x < 1 \text{ and } 0 < z-x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Clearly $f(z) = 0$ if $z < 0$ or $z \geq 2$. So we are dealing with

$$f_Z(z) = \int_0^z f(x)\, f(z-x)\, dx$$

Then consider the following two cases: (1) $0 < z \leq 1$ and (2) $1 < z < 2$. For these cases we have the following:

(1) we need $x \leq z$ for $0 < z - x < 1$, $f(x)\, f(z-x) = 1$

Then:

$$f_z(z) = \int_0^z 1 \, dx = z$$

(2) We need $x \geq 1 - z$ for $z - x \leq 1$

Then:

$$f_z(z) = \int_{z-1}^1 1 \, dx = 2 - z$$

Now for our specific example we have

$$f(x^{(1)} + x^{(2)}) = \begin{cases} x^{(1)} + x^{(2)} & \text{if } 0 < x^{(1)} + x^{(2)} < 1 \\ 2 - (x^{(1)} + x^{(2)}) & \text{if } 1 \leq x^{(1)} + x^{(2)} < 2 \\ 0 & \text{otherwise} \end{cases}$$

Using this density function we can now solve for the expectation.

4.1.1 Bayes risk

The Bayes risk is $R^* = \frac{1}{3}$ which can be computed as follows:

$$R^* = \mathbb{E} \min \left(\frac{x^{(1)} + x^{(2)}}{2}, 1 - \frac{x^{(1)} + x^{(2)}}{2} \right)$$

Let $x^{(1)} + x^{(2)} = z$

$$= \int_{-\infty}^{\infty} \min \left(\frac{z}{2}, 1 - \frac{z}{2} \right) f(z) \, dz$$

$$= 0 + \int_0^2 \min \left(\frac{z}{2}, 1 - \frac{z}{2} \right) f(z) \, dz$$

$$= \int_0^1 \min \left(\frac{z}{2}, 1 - \frac{z}{2} \right) f(z) \, dz + \int_1^2 \min \left(\frac{z}{2}, 1 - \frac{z}{2} \right) f(z) \, dz$$

$$= \int_0^1 \frac{z}{2} \, z \, dz + \int_1^2 \left(1 - \frac{z}{2} \right) (2 - z) \, dz$$

$$= \int_0^1 \frac{z^2}{2} \, dz + \frac{1}{2} \int_1^2 (z^2 - 4z + 4) \, dz$$

$$= \left[\frac{1}{6} z^3 \right]_0^1 + \frac{1}{2} \left[\frac{1}{3} z^3 - 2z^2 + 4z \right]_1^2$$

$$= \frac{1}{6} + \frac{1}{2} \left(\left[\frac{1}{3} 8 - \overset{\frac{8}{3}}{\cancel{8}} + 8 \right] - \left[\overset{\frac{7}{3}}{\cancel{\frac{1}{3} - 2 + 4}} \right] \right)$$

$$= \frac{1}{6} + \frac{1}{2} \, \frac{1}{3}$$

4.1.2 1NN

The risk of the 1NN classifier is $R^{1NN} = \frac{5}{12}$ which can be shown as follows:

$$R^{1NN} = 2\,\mathbb{E}\left[\frac{z}{2}\left(1-\frac{z}{2}\right)\right]$$

$$= 2\int_0^2 \frac{z}{2}\left(1-\frac{z}{2}\right) f(z)\,dz$$

$$= 2\left[\int_0^1 \frac{z}{2}\left(1-\frac{z}{2}\right) z\,dz + \int_1^2 \frac{z}{2}\left(1-\frac{z}{2}\right)(2-z)\,dz\right]$$

$$= 2\left[\frac{1}{2}\int_0^1 z^2 - \frac{z^3}{2}\,dz + \frac{1}{2}\int_1^2 \left(\frac{z^3}{2} - 2z^2 + 2z\right) dz\right]$$

$$= \left[\frac{1}{3}z^3 - \frac{1}{8}z^4\right]_0^1 + \left[\frac{1}{8}z^4 - \frac{2}{3}z^3 + z^2\right]_1^2$$

$$= \frac{5}{24} + \left[\overbrace{\left(2 - \frac{16}{3} + 4\right)}^{\frac{2}{3} \quad \frac{16}{24}} - \overbrace{\left(\frac{1}{8} - \frac{2}{3} + 1\right)}^{\frac{11}{24}}\right]$$

$$R^{1NN} = \frac{10}{24} = \frac{5}{12}$$

A quick sanity check shows that

$$R^{1NN} = \frac{5}{12} < \frac{4}{9} = 2R^*(1-R^*)$$

4.1.3 3NN

The risk of the 3NN classifier is $R^{3NN} = \frac{47}{120}$ which can be shown as follows:

$$
\begin{aligned}
R^{3NN} &= \mathbb{E}\left[\overset{\frac{1}{2}R^{1NN}}{\cancel{\eta(x)(1-\eta(x))}}\right] + 4\,\mathbb{E}\left[\eta(x)^2 (1-\eta(x))^2\right] \\
&= \frac{5}{24} + 4 \int_0^2 \frac{z^2}{4} \left(\frac{2-z}{2}\right)^2 f(z)\,dz \\
&= \frac{5}{24} + \frac{1}{4} \int_0^2 z^2 (2-z)^2 f(z)\,dz \\
&= \frac{5}{24} + \frac{1}{4}\left[\int_0^1 z^3 (2-z)^2\,dz + \int_1^2 z^2 (2-z)^3\,dz\right] \\
&= \frac{5}{24} + \frac{1}{4}\left[\int_0^1 z^5 - 4z^4 + 4z^3\,dz + \int_1^2 z^2 \left(2^3 - 3\cdot 2^2\cdot z + 3\cdot 2\cdot z^2 - z^3\right) dz\right] \\
&= \frac{5}{24} + \frac{1}{4}\left[\left[\frac{1}{6}z^6 - \frac{4}{5}z^5 + z^4\right]_0^1 + \int_1^2 \left(-z^5 + 6z^4 - 12z^3 + 8z^2\right) dz\right] \\
&= \frac{5}{24} + \frac{1}{4}\left[\left[\frac{1}{6} - \frac{4}{5} + 1\right] + \left[-\frac{1}{6}z^6 + \frac{6}{5}z^5 - 3z^4 + \frac{8}{3}z^3\right]_1^2\right] \\
&= \frac{5}{24} + \frac{1}{4}\left[\frac{11}{30} + \left[\underbrace{\cancel{\left(-\frac{64}{6} + \frac{32\cdot 6}{5} - 3\cdot 16 + \frac{64}{3}\right)}}_{\frac{32}{30}} - \underbrace{\cancel{\left(-\frac{1}{6} + \frac{6}{5} - 3 + \frac{8}{3}\right)}}_{\frac{21}{30}}\right]\right] \\
&= \frac{5}{24} + \frac{1}{4}\left[\frac{11}{30} + \frac{11}{30}\right] \\
&= \frac{5}{24} + \frac{1}{2}\,\frac{11}{30} = \frac{5}{24} + \frac{11}{60} = \frac{300}{1440} + \frac{264}{1440} = \frac{564}{1440} = \frac{47}{120} \\
R^{3NN} &= \frac{47}{120}
\end{aligned}
$$

In conclusion we have that:

$$R^* = \frac{40}{120} < R^{3NN} = \frac{47}{120} < R^{1NN} = \frac{50}{120}$$

4.2 Program

The KNN classifier can be implemented in R as follows:

```r
knn_classifier <- function(X,y,k,...) {
  row_idx <- 1:nrow(X)
  distances <- data.table(t(combn(row_idx,2)))
  distances[,dist:=c(dist(X,...))]
  distances_rev <- copy(distances)
  setnames(distances_rev, c("V1", "V2"), c("V2", "V1"))
  distances <- rbind(distances, distances_rev)
  setorder(distances, V1, dist)
  setnames(distances, c("V1", "V2"), c("X", "neighbour"))
  distances[,y_neighbour:=y[neighbour]]
  fitted <- distances[,median(y_neighbour[1:k]),by=X]$V1
  return(fitted)
}
```

To simulate the data I use the following helper function:

```r
sim_data <- function(n,d) {
  X <- matrix(runif(n*d),n) # uniform 0,1
  p_y <- rowSums(X[,1:2])/2 # probabilities of each Bernoulli trial
  y <- rbinom(n, 1, p_y)
  return(list(X=X,y=y))
}
```

I then run the programm multiple times varying the dimension d, the number of neighbours k and the sample size n. Each time I compute the frequency of error and finally average over those frequencies to obtain an estimate of the probability of error.