AUSTRALIAN NATIONAL UNIVERSITY

COMP4670 THEORY OF COMPUTATION

Assignment 01

Edited by LaTeX

College of Engineering and Computer Science

©*Author*

Jimmy Lin

u5223173

¶ *Lecturer*

Christfried Webers

¶ *Tutor*

Wen Shao

† *Release Date*

March. 22 2013

‡ *Due Date*

April. 22 2013

τ *Time Spent*

40 hours

April 22, 2013

Contents

1 Probabilities

1.1 Covariance of Sum

In order to prove the equation, we start from definition of $var[X+Y]$

$$var[X+Y] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Big[(x+y)-E(x+y)\Big]^2 P(x,y)dxdy \tag{1}$$

Based on the linearity of expectation of random variable (see proof at Appendix A.1), we have

$$E(x+y) = E(x) + E(y) \tag{2}$$

Hence, we can continue our manipulation for $var[X+Y]$

$$\begin{aligned}
var[X+Y] &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Big[x+y-(E(x)+E(y))\Big]^2 P(x,y)dxdy \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Big[(x-E(x))+(y-E(y))\Big]^2 P(x,y)dxdy \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Big[(x-E(x))^2+(y-E(y))^2+2(x-E(x))(y-E(y))\Big] P(x,y)dxdy \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-E(x))^2 P(x,y)dxdy + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(y-E(y))^2 P(x,y)dxdy \\
&\quad + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} 2(x-E(x))(y-E(y))P(x,y)dxdy \\
&= \int_{-\infty}^{+\infty}(x-E(x))^2 (\int_{-\infty}^{+\infty} P(x,y)dy)dx + \int_{-\infty}^{+\infty}(y-E(y))^2 (\int_{-\infty}^{+\infty} P(x,y)dx)dy \\
&\quad + 2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-E(x))(y-E(y))P(x,y)dxdy
\end{aligned} \tag{3}$$

Based on sum rule of probability, we have

$$P(x) = \int_{-\infty}^{+\infty} P(x,y)dy \tag{4}$$

$$P(y) = \int_{-\infty}^{+\infty} P(x,y)dx \tag{5}$$

By using the result of sum rule (4) and (5), we further induce $var[X+Y]$ from (3)

$$\begin{aligned}
var[X+Y] &= \int_{-\infty}^{+\infty}(x-E(x))^2 P(x)dx + \int_{-\infty}^{+\infty}(y-E(y))^2 P(y)dy \\
&\quad + 2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-E(x))(y-E(y))P(x,y)dxdy
\end{aligned} \tag{6}$$

By definition of variance and covariance, we obtain

$$var[X] = \int_{-\infty}^{+\infty}(x-E(x))^2 P(x)dx \tag{7}$$

$$var[Y] = \int_{-\infty}^{+\infty}(y-E(y))^2 P(y)dx \tag{8}$$

$$cov[X,Y] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-E(x))(y-E(y))P(x,y)dxdy \tag{9}$$

Then, by using (7), (8) and (9), we derive the result from (3)

$$var[X+Y] = var[X] + var[Y] + 2cov[X,Y] \tag{10}$$

1.2 Probability of Babies

Notational declaration: here, we use $S_1 = \{b,g\}$ to denote the first children being Boy and Girl respectively, and similarly, we use $S_2 = \{b,g\}$ to represent the gender of second children. Then, we utilize NB = {0,1,2} to denote the number of boys and NG = {0,1,2} to denote the number of girls.

1.2.1 Number of girls most likely to be

First, we denote the marginal distribution of S_1 and S_2

$$P(S_1 = b) = P(S_1 = g) = \frac{1}{2} \qquad P(S_2 = b) = P(S_2 = g) = \frac{1}{2} \tag{11}$$

Based on (11) and independence of genders of two children (iid), the joint distribution of $P(S_1,S_2)$ are

$$P(S_1 = b, S_2 = b) = \frac{1}{4} \qquad P(S_1 = b, S_2 = g) = \frac{1}{4} \tag{12}$$

$$P(S_1 = g, S_2 = b) = \frac{1}{4} \qquad P(S_1 = g, S_2 = g) = \frac{1}{4} \tag{13}$$

Next, from (12), (13), we can describe the problem using NB and NG,

$$P(NB = 0, NG = 2) = P(S_1 = g, S_2 = g) = \frac{1}{4} \tag{14}$$

$$P(NB = 2, NG = 0) = P(S_1 = b, S_2 = b) = \frac{1}{4} \tag{15}$$

$$P(NB = 1, NG = 1) = P(S_1 = b, S_2 = b) + P(S_1 = b, S_2 = g) = \frac{1}{2} \tag{16}$$

Having observed (14), (15) and (16), the most likely gender condition of the neibhours' children is

$$NB = 1, NG = 1 \tag{17}$$

That is to say, **one boy and one girl** is the most likely scenario.

1.2.2 Probability of one child being a boy

Since we have known the number of girls cannot be zero, and we want to know existence of boys, we need to work out $P(NB \geq 1|NG \geq 1)$, that is $P(NB = 1|NG \geq 1)$.

$$P(NG \geq 1) \ = P(NB = 1, NG = 1) + P(NB = 0, NG = 2) \ = \frac{3}{4} \tag{18}$$

Since the total number of children is 2, we have

$$P(NB = 1, NG \geq 1) = P(NB = 1, NG = 1) = \frac{1}{2} \tag{19}$$

From (18) and (19), we obtain the objective probability,

$$P(NB = 1|NG \geq 1) \ = \frac{P(NB = 1, NG \geq 1)}{P(NG \geq 1)} \ = \frac{P(NB = 1, NG = 1)}{P(NG \geq 1)} \ = \frac{\frac{1}{2}}{\frac{3}{4}} \ = \frac{2}{3} \tag{20}$$

That is, the probability of one child being boy is $\frac{2}{3}$, given that there is at least one girl.

1.2.3 Probability of the other child being a boy

Based on the iid of two children's gender, pre-knowledge of first child has no effect on the gender of sencond child. Suppose we have seen S_1 to be a girl, the probability of S_2 given $S_1 = g$ is

$$P(S_2 = b|S_1 = g) = \frac{P(S_2 = b, S_1 = g)}{P(S_1 = g)} \ = \frac{P(S_2 = b)P(S_1 = g)}{P(S_1 = g)} \ = P(S_2 = b) \ = \frac{1}{2} \tag{21}$$

Even we know the gender of one child, the probability of other one child being boy remains to be $\frac{1}{2}$.

1.3 Maximum Likelihood for Multivariate Gaussian Distribution

1.3.1 Likelihood of all data

To get the likelihood of all data, we first present the Gaussian Distribution for single data object

$$p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) \;= \frac{1}{(2\pi)^{\frac{D}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\; exp[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})] \tag{22}$$

Since each data are collected independently on each other (iid), we can derive the following likelihood function in the given data collection $\boldsymbol{X} = \{\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, ..., \boldsymbol{x}_n^T\}$

$$\begin{aligned} L(\boldsymbol{\mu},\boldsymbol{\Sigma}) &= p(\boldsymbol{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) \\ &= \prod_{n=1}^{N} p(\boldsymbol{x}_n|\boldsymbol{\mu},\boldsymbol{\Sigma}) \\ &= \prod_{n=1}^{N} \frac{1}{(2\pi)^{\frac{D}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} exp[-\frac{1}{2}(\boldsymbol{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu})] \\ &= \frac{1}{(2\pi)^{\frac{ND}{2}}|\boldsymbol{\Sigma}|^{\frac{N}{2}}} exp\Big[-\frac{1}{2}\sum_{n=1}^{N}\Big((\boldsymbol{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu})\Big)\Big] \end{aligned} \tag{23}$$

1.3.2 Extrenum Solution

To obtain the MLE solution for parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, we take logarithm of likelihood (23) first

$$ln(L(\boldsymbol{\mu},\boldsymbol{\Sigma})) \;= -\frac{ND}{2}ln(2\pi) - \frac{N}{2}ln|\boldsymbol{\Sigma}| \;-\frac{1}{2}\sum_{n=1}^{N}[(\boldsymbol{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu})] \tag{24}$$

Then we take directional derivative to the log likelihood function (24) with regard to $\boldsymbol{\mu}$, here we assume an arbitrary vector ζ in $\boldsymbol{\mu}$'s space $\mathcal{R}^D$,

$$\mathcal{D}\{L(\boldsymbol{\mu},\boldsymbol{\Sigma})\}(\zeta) = \mathcal{D}\{-\frac{1}{2}\sum_{n=1}^{N}[(\boldsymbol{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu})]\}(\zeta) \tag{25}$$

Note that the first and second term in (3) is irrelavant to $\boldsymbol{\mu}$, hence they are removed after being taken derivative with regard to $\boldsymbol{\mu}$.
Based on the linearity of directional derivative, we have,

$$\mathcal{D}\{L(\boldsymbol{\mu},\boldsymbol{\Sigma})\}(\zeta) = -\frac{1}{2}\sum_{n=1}^{N}\mathcal{D}\{(\boldsymbol{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu})\}(\zeta) \tag{26}$$

Since the $(\boldsymbol{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu})$ is a scalar value, its trace equals to itself,

$$\mathcal{D}\{L(\boldsymbol{\mu},\boldsymbol{\Sigma})\}(\zeta) = -\frac{1}{2}\sum_{n=1}^{N}\mathcal{D}\{tr\big((\boldsymbol{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu})\big)\}(\zeta) \tag{27}$$

Then we cite a significant formula introduced in lecture, that is why trace in formula (27) is introduced

$$if\;\, f(\boldsymbol{X}) = tr\big(\boldsymbol{X}^T\boldsymbol{C}\boldsymbol{X}\big), then\;\, \mathcal{D}\{f(\boldsymbol{X})\}(\xi) = tr\big(\boldsymbol{X}^T(\boldsymbol{C}^T+\boldsymbol{C})\xi\big) \tag{28}$$

Here, we use $\Rightarrow$ notation to present the instantiation of symbols in this question, i.e. $\boldsymbol{X}$ is instantiated to be $\boldsymbol{x}_n-\boldsymbol{\mu}$ in the case of our problem

$$\boldsymbol{X} \Rightarrow \boldsymbol{x}_n-\boldsymbol{\mu} \tag{29}$$

$$\boldsymbol{C} \Rightarrow \boldsymbol{\Sigma}^{-1} \tag{30}$$

$$\xi \Rightarrow \zeta \tag{31}$$

$$f(\boldsymbol{X}) \Rightarrow tr\Big((\boldsymbol{x}_n - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu})\Big) \tag{32}$$

Then, by applying the cited formula (28), we have the following from (27)

$$\mathcal{D}\{L(\boldsymbol{\mu}, \boldsymbol{\Sigma})\}(\zeta) = -\frac{1}{2}\sum_{n=1}^{N} tr\Big((\boldsymbol{x}_n - \boldsymbol{\mu})^T ((\boldsymbol{\Sigma}^{-1})^T + \boldsymbol{\Sigma}^{-1})\zeta\Big) \tag{33}$$

Since the symmetry of covariance matrix $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Sigma}^T = \boldsymbol{\Sigma} \tag{34}$$

Take inverse for both side of (34)

$$(\boldsymbol{\Sigma}^T)^{-1} = (\boldsymbol{\Sigma})^{-1} \tag{35}$$

Change the order of inverse and transpose of leftside term,

$$(\boldsymbol{\Sigma}^{-1})^T = (\boldsymbol{\Sigma})^{-1} \tag{36}$$

Next we apply the symmetry of $\boldsymbol{\Sigma}^{-1}$, derive the following from (33),

$$\mathcal{D}\{L(\boldsymbol{\mu}, \boldsymbol{\Sigma})\}(\zeta) = -\sum_{n=1}^{N} tr\Big((\boldsymbol{x}_n - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}\zeta\Big) \tag{37}$$

Since the $(\boldsymbol{x}_n - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}\zeta$ is a scalar value, we can directly ignore its trace,

$$\mathcal{D}\{L(\boldsymbol{\mu}, \boldsymbol{\Sigma})\}(\zeta) = -\sum_{n=1}^{N} \Big((\boldsymbol{x}_n - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}\zeta\Big) \tag{38}$$

Set the directional derivative in (38) to zero for deriving extrenum, we have

$$\sum_{n=1}^{N} \Big((\boldsymbol{x}_n - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}\zeta\Big) = 0 \tag{39}$$

Since ζ is constant vector, we have (note the change of position of parenthesis)

$$0 = \sum_{n=1}^{N} \Big((\boldsymbol{x}_n - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}\Big)\zeta \tag{40}$$

And since ζ can be arbitrary and (40) holds for any assignment of ζ,

$$0 = \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} \tag{41}$$

Multiply one $\boldsymbol{\Sigma}$ in the rightmost position of both side of formula (41),

$$0 = \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{\mu})^T \tag{42}$$

Next, solve the formula (42), we get the desired $\boldsymbol{\mu}_{ML}$,

$$\boldsymbol{\mu}_{ML} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n \tag{43}$$

Now we have figured out the parameter $\boldsymbol{\mu}$ of the extrenum. Next, we are about to compute the $\boldsymbol{\Sigma}$ at that point. First, revisit the log likelihood function (24),

$$ln(L(\boldsymbol{\mu}, \boldsymbol{\Sigma})) = -\frac{ND}{2}ln(2\pi) - \frac{N}{2}ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{n=1}^{N}[(\boldsymbol{x}_n - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu})] \quad (44)$$

Take first-order directional derivative to log likelihood with regard to $\boldsymbol{\Sigma}$. The η is arbitrary vector in the $\boldsymbol{\Sigma}$'s space $\mathcal{R}^{D\times D}$,

$$\mathcal{D}\{ln(L(\boldsymbol{\mu}, \boldsymbol{\Sigma}))\}(\eta) = \mathcal{D}\{-\frac{N}{2}ln|\boldsymbol{\Sigma}|\}(\eta) + \mathcal{D}\{-\frac{1}{2}\sum_{n=1}^{N}[(\boldsymbol{x}_n - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu})]\}(\eta) \quad (45)$$

Then, we consider the first term in (45) for now,

$$\begin{aligned}
\mathcal{D}\{-\frac{N}{2}ln|\boldsymbol{\Sigma}|\}(\eta) &= \frac{\partial(-\frac{N}{2}ln|\boldsymbol{\Sigma}|)}{\partial|\boldsymbol{\Sigma}|}\mathcal{D}\{|\boldsymbol{\Sigma}|\}(\eta) && (46)\\
&= -\frac{N}{2}\frac{1}{|\boldsymbol{\Sigma}|}\mathcal{D}\{|\boldsymbol{\Sigma}|\}(\eta) && (47)\\
&= -\frac{N}{2}\frac{1}{|\boldsymbol{\Sigma}|}|\boldsymbol{\Sigma}|\ tr\big(\boldsymbol{\Sigma}^{-1}\eta\big) && (48)\\
&= tr\big(-\frac{N}{2}\boldsymbol{\Sigma}^{-1}\eta\big) && (49)
\end{aligned}$$

where chain rule of directional derivative are utilized for induction. Besides, we use the linear property of trace operation for the final step. On top of that, the penultimate step involves in using the provided rule,

$$\mathcal{D}\{|A|\}(B) = |A|\ tr\big(A^{-1}B\big) \quad (50)$$

Again, we use the property of directional derivative,

$$\mathcal{D}\{f(X)\}(\xi) = \langle grad\ f(X),\ \xi\rangle = tr\Big(\big(grad\ f(X)\big)^T\ \xi\Big) \quad (51)$$

where definition of inner product is $\langle A,\ B\rangle = tr(A^T\ B)$.
According to the convention (51) and what we derived previously (49), we have

$$grad\ \Big(-\frac{N}{2}ln|\boldsymbol{\Sigma}|\Big) = \Big(-\frac{N}{2}\boldsymbol{\Sigma}^{-1}\Big)^T = -\frac{N}{2}\boldsymbol{\Sigma}^{-1} \quad (52)$$

Note that here the symmetric property of precision matrix $\boldsymbol{\Sigma}^{-1}$ (36) is used.
Then, consider the second term in (45), and we use $W = \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu})(\boldsymbol{x}_n - \boldsymbol{\mu})^T$ for abbreviation,

$$\mathcal{D}\{-\frac{1}{2}\sum_{n=1}^{N}[(\boldsymbol{x}_n - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu})]\}(\eta) = -\frac{1}{2}\sum_{n=1}^{N}\mathcal{D}\{[(\boldsymbol{x}_n - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu})]\}(\eta) \quad (53)$$

$$(54)$$

Similarly, work out the gradient of the second term. (In fact, I did not work it out by using the way shown above and given formula. Instead, I directly utilize the gradient following formula to get the result

$$\frac{\partial a^T X^{-1} b^b}{\partial X} = -X^{-T}ab^TX^{-T}$$

The gradient of second term is,

$$\begin{aligned}
grad\ \Big(-\frac{1}{2}\sum_{n=1}^{N}(\boldsymbol{x}_n - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu})\Big) &= \frac{1}{2}\sum_{n=1}^{N}\big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu})(\boldsymbol{x}_n - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\big) && (55)\\
&= \boldsymbol{\Sigma}^{-1}\Big(\frac{1}{2}\sum_{n=1}^{N}(\boldsymbol{x}_n - \boldsymbol{\mu})(\boldsymbol{x}_n - \boldsymbol{\mu})^T\Big)\boldsymbol{\Sigma}^{-1} && (56)
\end{aligned}$$

Next, we take the gradient of likelihood function, and set it to zero

$$grad\left(ln(L(\boldsymbol{\mu}, \boldsymbol{\Sigma}))\right) = -\frac{N}{2}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\Big(\frac{1}{2}\sum_{n=1}^{N}(\boldsymbol{x}_n - \boldsymbol{\mu})(\boldsymbol{x}_n - \boldsymbol{\mu})^T\Big)\boldsymbol{\Sigma}^{-1} = 0 \tag{57}$$

Solve the equation above, we have

$$\boldsymbol{\Sigma}_{ML} = \frac{1}{N}\sum_{n=1}^{N}\Big((\boldsymbol{x}_n - \boldsymbol{\mu}_{ML})(\boldsymbol{x}_n - \boldsymbol{\mu}_{ML})^T\Big) \tag{58}$$

Note that since the Multivariate Gaussian distribution has only one extrenum, the extreme parameter $\boldsymbol{\mu}$ must occur with the extreme $\boldsymbol{\Sigma}$. So here we use ML to subscript the $\boldsymbol{\mu}$ in the solution of $\boldsymbol{\Sigma}_{ML}$.

Finally, we substitute $\boldsymbol{\mu}_{ML}$ by the solution we work out in (43)

$$\boldsymbol{\Sigma}_{ML} = \frac{1}{N}\sum_{n=1}^{N}\Big((\boldsymbol{x}_n - \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n)(\boldsymbol{x}_n - \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n)^T\Big) \tag{59}$$

1.3.3 The found extremum is maximum

We take the second-order directional derivative of likelihood function with regard to mean of Gaussian distribution based on the what we derived in (38). Here, the definition of ζ follows what we specified in 1.3.2.

$$\mathcal{D}^2\{L(\boldsymbol{\mu},\boldsymbol{\Sigma})\}(\zeta) = \mathcal{D}\{\mathcal{D}\{L(\boldsymbol{\mu},\boldsymbol{\Sigma})\}(\zeta)\}(\zeta) = \mathcal{D}\{-\sum_{n=1}^{N}\Big((\boldsymbol{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\zeta\Big)\}(\zeta) \tag{60}$$

Out of linearity of directional derivative, we have

$$\mathcal{D}^2\{L(\boldsymbol{\mu},\boldsymbol{\Sigma})\}(\zeta) = -\sum_{n=1}^{N}\mathcal{D}\{\Big((\boldsymbol{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\zeta\Big)\}(\zeta) \tag{61}$$

Here, another formula presented in lecture should be introduced,

$$\mathcal{D}\{f(\boldsymbol{X})\}(\xi) = \langle\ grad\Big(f(\boldsymbol{X})\Big),\ \xi\ \rangle \tag{62}$$

where, the inner product $\langle A, B\rangle$ is defined as,

$$\langle A, B\rangle = tr(A^TB) \tag{63}$$

Following the (61), we present the instantiation by using the notation $\Rightarrow$,

$$\xi \Rightarrow \zeta \tag{64}$$

$$f(\boldsymbol{X}) \Rightarrow (\boldsymbol{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\zeta \tag{65}$$

Next, we calculate the gradient of the $(\boldsymbol{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\zeta$ with regard to $\boldsymbol{\mu}$

$$grad\Big((\boldsymbol{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\zeta\Big) = \boldsymbol{\Sigma}^{-1}\zeta \tag{66}$$

Then, derive the form of second-order derivative,

$$\begin{aligned}\mathcal{D}^2\{L(\boldsymbol{\mu},\boldsymbol{\Sigma})\}(\zeta) &= -\sum_{n=1}^{N}\langle\ grad\Big((\boldsymbol{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\zeta\Big),\zeta\ \rangle \\ &= -\sum_{n=1}^{N}\langle\ \boldsymbol{\Sigma}^{-1}\zeta,\zeta\ \rangle \\ &= -\sum_{n=1}^{N} tr\Big((\boldsymbol{\Sigma}^{-1}\zeta)^T\zeta\Big) = -N\ tr\Big(\zeta^T(\boldsymbol{\Sigma}^{-1})^T\zeta\Big)\end{aligned} \tag{67}$$

Based on the symmetry of inverse of covariance matrix (36), the final form of second-order directional derivative was determined from (67)

$$\mathcal{D}^2\{L(\boldsymbol{\mu},\boldsymbol{\Sigma})\}(\zeta) = -N\ tr\Big(\zeta^T\boldsymbol{\Sigma}^{-1}\zeta\Big) \tag{68}$$

Since the $\zeta^T\boldsymbol{\Sigma}^{-1}\zeta$ is scalar value, we have

$$\mathcal{D}^2\{L(\boldsymbol{\mu},\boldsymbol{\Sigma})\}(\zeta) = -N\ \zeta^T\boldsymbol{\Sigma}^{-1}\zeta \tag{69}$$

Because of positive-definiteness of covariance matrix $\boldsymbol{\Sigma}$, we can easily derive the positive-definiteness of precision matrix $\boldsymbol{\Sigma}^{-1}$

$$\zeta^T\boldsymbol{\Sigma}^{-1}\zeta > 0,\ for\ any\ \zeta \tag{70}$$

Therefore, we get the negativity of second derivative of likelihood function,

$$\mathcal{D}^2\{L(\boldsymbol{\mu},\boldsymbol{\Sigma})\}(\zeta) = -N\ \zeta^T\boldsymbol{\Sigma}^{-1}\zeta < 0,\ for\ any\ \zeta \tag{71}$$

As presented by (71), the second-order directional derivative of likelihood function is **less than zero at any direction**. Therefore, the first-order directional derivative at any direction is decreasing all the time and **the only extrenum derived is maixmum** rather than minimum.

1.3.4 Arbitrary Order

The order of maximizing the parameters are **not arbitrary**. We have to first work out the optimal $\boldsymbol{\mu}_{ML}$, which is the point where the Gaussian distribution centered. And then compute the optimal covariance matrix $\boldsymbol{\Sigma}_{ML}$ by using the previously derived $\boldsymbol{\mu}_{ML}$. However, the procedure **cannot be reversed**. This is because even we take derivative towards the logarithm of likelihood $ln(L(\boldsymbol{\mu}, \boldsymbol{\Sigma}))$ with regard to covariance matrix $\boldsymbol{\Sigma}$, shown in the (59), the result $\boldsymbol{\Sigma}$ contains one term with $\boldsymbol{\mu}$.

$$\boldsymbol{\Sigma}_{ML} = \frac{1}{N}\sum_{n=1}^{N}\Big((\boldsymbol{x}_n - \boldsymbol{\mu}_{ML})(\boldsymbol{x}_n - \boldsymbol{\mu}_{ML})^T\Big) \tag{72}$$

That is to say, we cannot obtain the $\boldsymbol{\Sigma}_{ML}$ unless we derive its corresponding $\boldsymbol{\mu}$ in advance. Therefore, **the order of maximizing the parameter could not be arbitrary**.

1.3.5 Parameters change w.r.t order of input data

There should be **no change to parameters** if the order of input data was changed. This is for the reason that the $\boldsymbol{\mu}_{ML}$, as suggested by (43), is computed by simply summing over all the input data, without involving in any information about the order of the input data. Since the $\boldsymbol{\mu}_{ML}$ is a constant after being computed, the latter $\boldsymbol{\Sigma}_{ML}$, suggested by (59) does not have anything to do with the order of input data. .

Therefore, if order of input data changed, **no change would happen to the resulted parameters $\boldsymbol{\mu}_{ML}$ and $\boldsymbol{\Sigma}_{ML}$**. But if the value if input data, the resulted parameters may change.

1.4 Lifetime of Equipment

1.4.1 Calculate MLE $\hat{\theta}$

For a set of iid data X_1, X_2 .. , X_N, first work out its likelihood $L(X_1, X_2.., X_N, \theta)$

$$L(X_1, X_2.., X_N, \theta) = \prod_{n=1}^{N} P(x_n|\theta) = \prod_{n=1}^{N} \theta e^{-\theta x_n} \tag{73}$$

Then take the logarithm of likelihood

$$ln(L(X_1, X_2.., X_N, \theta)) = ln(\prod_{n=1}^{N} \theta e^{-\theta x_n}) = \sum_{n=1}^{N}(-\theta x_n + ln(\theta)) \tag{74}$$

Next, take derivative of log likelihood with regard to θ

$$\frac{d(ln(L(X_1, X_2.., X_N, \theta)))}{d\theta} = \frac{d(\sum_{n=1}^{N}(-\theta x_n + ln(\theta)))}{d\theta} = \sum_{n=1}^{N}(-x_n + \frac{1}{\theta}) = \frac{N}{\theta} - \sum_{n=1}^{N} X_n \tag{75}$$

To get the $\hat{\theta}$, set the derivative to zero

$$\frac{N}{\hat{\theta}} - \sum_{n=1}^{N} X_n = 0 \tag{76}$$

Solve the equation above, we have

$$\hat{\theta} = \frac{N}{\sum_{n=1}^{N} X_n} \tag{77}$$

1.4.2 Relation to mean of X_1, X_2 .. , X_N

By definition, we have

$$mean = \frac{\sum_{n=1}^{N} X_n}{N} \tag{78}$$

Based on the (77) and (78), we can easily figure out the relation between mean of datasets and MLE solution $\hat{\theta}$, that is

$$mean \cdot \hat{\theta} = 1 \tag{79}$$

1.4.3 One instantiation

First, derive the value of mean of $X_1 = 5, X_2 = 4, X_3 = 3, X_4 = 4$

$$mean = \frac{\sum_{n=1}^{4} X_n}{4} = \frac{5+4+3+4}{4} = 4 \tag{80}$$

Based on the (79), we have the MLE solution $\hat{\theta}$

$$\hat{\theta} = \frac{1}{mean} = \frac{1}{4} \tag{81}$$

Correspondingly, the MLE probability distribution for given datasets is

$$P(x) = \frac{1}{4} e^{-\frac{1}{4}x} \qquad x \geq 0 \tag{82}$$

2 Decision Theory

2.1 Lower Bound for the Correct Classification

2.1.1 Show that: if $a \leq b$, then $a \leq (ab)^{\frac{1}{2}}$

We start proof from pre-condition of the implication.

$$a \leq b \tag{83}$$

Since a is non-negative number, we have

$$a^2 \leq ab \tag{84}$$

Then we obtain the following by taking root of two side, (since the root function is monotonically increasing in its scope, we keep the direction of inequality)

$$|a| \leq (ab)^{\frac{1}{2}} \tag{85}$$

The step above is valid because the b is non-negative. Besides, due to the non-negativity of a, we have,

$$|a| = a \tag{86}$$

Hence, we derive the post-condition.

$$a \leq (ab)^{\frac{1}{2}} \tag{87}$$

2.1.2 Proof of lower bound

We start from the definition of probability of mistake in the problem of binary classification,

$$\begin{aligned} p(mistake) &= p(\mathbf{x} \in \mathcal{R}_1, \mathcal{C}_2) + p(\mathbf{x} \in \mathcal{R}_2, \mathcal{C}_1) \\ &= \int_{\mathbf{x} \in \mathcal{R}_1} p(\mathbf{x}, \mathcal{C}_2) d\mathbf{x} + \int_{\mathbf{x} \in \mathcal{R}_2} p(\mathbf{x}, \mathcal{C}_1) d\mathbf{x} \end{aligned} \tag{88}$$

Since the decision region was chosen to minimise the probability of misclassification, say, we are about to minimize the intergrand in (88), that is

$$p(\mathbf{x}, \mathcal{C}_2) \leq p(\mathbf{x}, \mathcal{C}_1), \mathbf{x} \in \mathcal{R}_1 \tag{89}$$

$$p(\mathbf{x}, \mathcal{C}_1) \leq p(\mathbf{x}, \mathcal{C}_2), \mathbf{x} \in \mathcal{R}_2 \tag{90}$$

Next, according to the inequality (87), we can derive the followings from (89) and (90),

$$p(\mathbf{x}, \mathcal{C}_2) \leq \big(p(\mathbf{x}, \mathcal{C}_1) \cdot p(\mathbf{x}, \mathcal{C}_2)\big)^{1/2}, \mathbf{x} \in \mathcal{R}_1 \tag{91}$$

$$p(\mathbf{x}, \mathcal{C}_1) \leq \big(p(\mathbf{x}, \mathcal{C}_1) \cdot p(\mathbf{x}, \mathcal{C}_2)\big)^{1/2}, \mathbf{x} \in \mathcal{R}_2 \tag{92}$$

Then, based on (88), (91) and (92), we have,

$$\begin{aligned} p(mistake) &\leq \int_{\mathbf{x} \in \mathcal{R}_1} \big(p(\mathbf{x}, \mathcal{C}_1) \cdot p(\mathbf{x}, \mathcal{C}_2)\big)^{1/2} d\mathbf{x} + \int_{\mathbf{x} \in \mathcal{R}_2} \big(p(\mathbf{x}, \mathcal{C}_1) \cdot p(\mathbf{x}, \mathcal{C}_2)\big)^{1/2} d\mathbf{x} \\ &= \int_{\mathbf{x} \in \mathcal{R}_1 \cup \mathcal{R}_2} \big(p(\mathbf{x}, \mathcal{C}_1) \cdot p(\mathbf{x}, \mathcal{C}_2)\big)^{1/2} d\mathbf{x} \end{aligned} \tag{93}$$

Since the total region $\mathcal{R}$ is only separated to be R_1 and R_2 in binary classification, we have

$$\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \tag{94}$$

Apply the above equation about the union of region to inequation (93), we have

$$p(mistake) \leq \int_{\mathbf{x} \in \mathcal{R}} \sqrt{p(\mathbf{x}, \mathcal{C}_1) \cdot p(\mathbf{x}, \mathcal{C}_2)} d\mathbf{x} \tag{95}$$

3 Dimensionality Reduction

3.1 Projection with Fisher's Discriminant

3.1.1 Calculate S_W, S_B and Find W

Two equations we will use for computation are shown in the following,

$$S_W = \sum_{k}^{K} \sum_{\boldsymbol{x}_n \in \mathcal{C}_k} (\boldsymbol{x}_n - \mathbf{m}_k)(\boldsymbol{x}_n - \mathbf{m}_k)^T \tag{96}$$

$$S_B = \sum_{k}^{K} N_k(\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})^T \tag{97}$$

See programs for computation. Next are the result presentations.
Within-Class Scatter Matrix:

$$S_W = \begin{pmatrix} 38.9562 & 13.63 & 24.6246 & 5.645 \\ 13.63 & 16.962 & 8.1208 & 4.8084 \\ 24.6246 & 8.1208 & 27.2226 & 6.2718 \\ 5.645 & 4.8084 & 6.2718 & 6.1566 \end{pmatrix}$$

Between-Class Scatter Matrix:

$$S_B = \begin{pmatrix} 63.21213333 & -19.95266667 & 165.2484 & 71.27933333 \\ -19.95266667 & 11.34493333 & -57.2396 & -22.93266667 \\ 165.2484 & -57.2396 & 437.1028 & 186.774 \\ 71.27933333 & -22.93266667 & 186.774 & 80.41333333 \end{pmatrix}$$

Then, we derive:

$$S_W^{-1} S_B = \begin{pmatrix} -3.05836939 & 1.08138264 & -8.1119227 & -3.45864987 \\ -5.56163926 & 2.17821866 & -14.96461194 & -6.30773951 \\ 8.07743878 & -2.94271854 & 21.5115909 & 9.14206468 \\ 10.49708187 & -3.41985449 & 27.54852482 & 11.84588007 \end{pmatrix}$$

After eigenvalue decomposition, we have the pairs of eigenvalue and eigenvector:

Eigenvalue	Corresponding scaled eigenvector
32.1919292	(-0.20874183, -0.38620368, 0.5540117, 0.70735037)
0.285391043	(0.00653196, 0.58661056, -0.25256154, 0.76945311)
1.22e-15 +4.95e-15j	(-0.061-0.570j, -0.228+0.249j, -0.282+0.248j, 0.64441626)
1.22-15 -4.9515j	(-0.061+0.570j, -0.228-0.249j, -0.282-0.248j, 0.64441626)

Finally, we work out the matrix W associated with D' largest eigenvector:

$$W = \begin{pmatrix} -0.20874183 & 0.00653196 & -0.061 - 0.570j & -0.061 + 0.570j \\ -0.38620368 & 0.58661056 & -0.228 + 0.249j & -0.228 - 0.249j \\ 0.5540117 & -0.25256154 & -0.282 + 0.248j & -0.282 - 0.248j \\ 0.70735037 & 0.76945311 & 0.64441626 & 0.64441626 \end{pmatrix}$$

3.1.2 When $D' = 2$

(a) Report the two eigenvalues and eigenvectors found.
Two eigenvalues and corresponding eigenvectors we found are as follows,

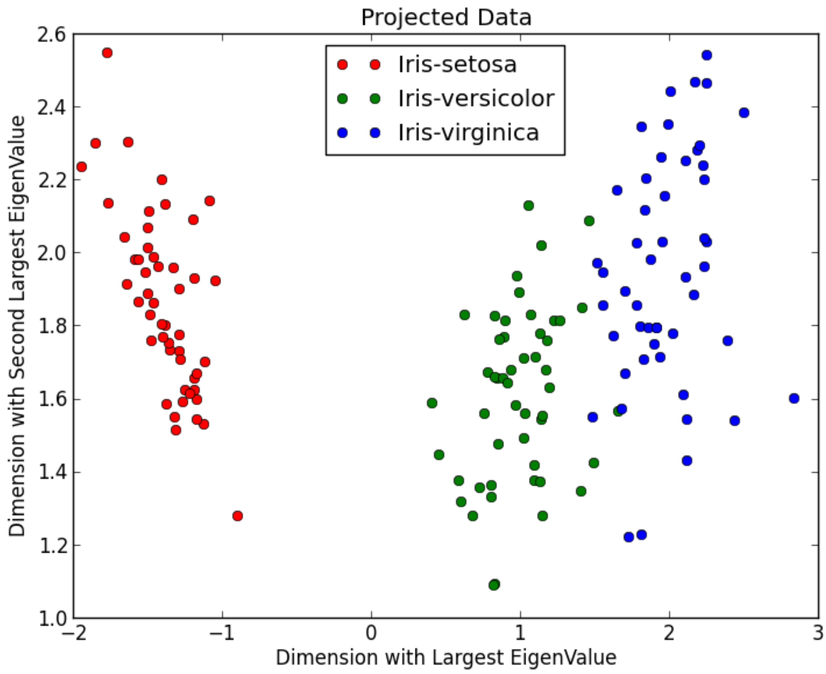
$$\lambda_1 = 32.1919292, \mathcal{V}_1 = (-0.20874183, -0.38620368, 0.5540117, 0.70735037)$$

$$\lambda_2 = 0.285391043, \mathcal{V}_2 = (0.00653196, 0.58661056, -0.25256154, 0.76945311)$$

Therefore, we have $W_{D'=2}$

$$W_{D'=2} = \begin{pmatrix} -0.20874183 & 0.00653196 \\ -0.38620368 & 0.58661056 \\ 0.5540117 & -0.25256154 \\ 0.70735037 & 0.76945311 \end{pmatrix}$$

(b) Provide a plot of the projected data using different colours for each class.



(c) Discuss the ratio of the two eigenvalues found with respect to the task of classifying the data in the projected 2-dimensional space.
Again, present the the two eigenvalues found first,

$$\lambda_1 = 32.1919292, \lambda_2 = 0.285391043$$

And the ratio of those two eigenvalues are as follows,

$$r = \frac{\lambda_1}{\lambda_2} = 112.799367851$$

Note that the horizontal coordinate axis X_1 corresponds to the dimension with λ_1, the vertical coordinate axis X_2 corresponds to the dimension with λ_2.

From the result of data projection in two dimension, we can see that **those three classes has tremendous distinction in the horizontal coordinate** (with large eigenvalue), while the differences in the vertical coordinate (with the smaller eigenvalue). Therefore, I commit to the intuition that **the ratio of two eigenvalues r shows the relative distinguishability (between two dimensions) of data objects from each other class**. And by the way, such distinguisability may come from the large between-class variance and small within-class variance.

3.1.3 Codes to Compute criteria J

We have accomplished the codes to compute S_w, S_B and then criteria J for projecting original data into $\mathcal{V}_1$ $\mathcal{V}_2$, please see getSolution_3_3_,

Within-Class Scatter Matrix:

$$S_W = \begin{pmatrix} 9.31175619 & -2.18194354 \times 10^{-08} \\ -2.18194354 \times 10^{-08} & 10.7967585 \end{pmatrix}$$

Between-Class Scatter Matrix:

$$S_B = \begin{pmatrix} 299.763396 & 2.58546748 \times 10^{-06} \\ 2.58546748 \times 10^{-6} & 3.08129816 \end{pmatrix}$$

Then, we derive:

$$S_W^{-1} S_B = \begin{pmatrix} 32.1919292 & 2.78325001 \times 10^{-07} \\ 3.04524474 \times 10^{-07} & 0.285391043 \end{pmatrix}$$

Finally, we have trace of $S_W^{-1} S_B$,

$$J = tr\{S_W^{-1} S_B\} = 32.4773202409$$

3.1.4 Compare to other projection

We have achieved automatically derive all combinations of two-axes projection. And use the pre-defined framework to work out criteria J for each combination, please see getSolution_3_4_,

projecting combo	criteria J
$\mathcal{V}_1$, $\mathcal{V}_2$	32.4773202409
Sepal Length , Sepal Width	4.3327944087
Sepal Length , Petal Length	23.3646503713
Sepal Length , Petal Width	13.0644017179
Sepal Width , Petal Length	21.8610096544
Sepal Width , Petal Width	20.3468961554
Petal Length , Petal Width	19.7820503322

Note that the Sepal Length, Sepal Width, Petal Length, Petal Width are the 1st, 2nd, 3rd, 4th feature in the original input Iris data, respectively.

It is obvious that **the practice of projecting data into $\mathcal{V}_1$,$\mathcal{V}_2$ derived from fisher algorithm has the best outcome in maximizing between-class variance the and minimizing the within-class variance in the meanwhile**, comparing to projecting into any combination of 2-dimensional axes in original space.

It may be generalized to a wider conclusion that for any dimension $D'(D' < n)$ (n is the number of input features), the projecting matrix $W_{n \times D'}$ figured out by fisher's algorithm are the best projection to classify data objects among all possible D'-dimension projection. This is for the reason that the indicator J of projection from fisher algorithm's would be largest among all D'-dimension projections.

4 Cross Validation and Classification

4.1 K-Nearest Neighbours Algorithm

4.1.1 Implement K-NN algorithm

See the code file "KNN.py" for implementing the K-NN algorithm.

In the KNN.py, the function getSolution_4_1_1 is a test function for the KNN Implementation. It uses the last data object of Iris datasets as the only testing data object and the rest as training dataset. I also implement the min-max scaling to the original data since the KNN classifier is **vulnerable to bad scaling**. (see the scaling function in KNN.py) Specifically, preprocessing towards data, called min-max scaling, is to project each dimension of input x into the range of $[0, 1]$,

$$x := \frac{x - x_{min}}{x_{max} - x_{min}} \tag{98}$$

4.1.2 Apply Cross Validation

Here we provide result of cross validation for scaled input and original input. You can see both results by switching corresponding argument of all "getSolution" functions.

Note that the result reported in the following is derived not through shuffling the original input. Because Unshuffled grouping is one particular shuffled grouping in cross validation, I did not implement the shuffling in my code. (in fact, only several lines of codes calling numpy.random is sufficient to achieve that functionality)

For original input data:

For 2-fold cross validation, we pick up $k = 8$, whose average cross validation test error is 4.0% . See the tabular result of unscaled 2-fold Cross Validation.

For 5-fold cross validation, we pick up $k = 12$, whose average cross validation test error is 2.0% . See the tabular result of unscaled 5-fold Cross Validation.

For 10-fold cross validation, we pick up $k = 20$, whose average cross validation test error is 2.0% . See the tabular result of unscaled 10-fold Cross Validation.

For scaled input data:

For 2-fold cross validation, we pick up $k = 11$, whose average cross validation test error is 2.67% . See the tabular result of scaled 2-fold Cross Validation.

For 5-fold cross validation, we pick up $k = 16$, whose average cross validation test error is 2.67% . See the tabular result of scaled 5-fold Cross Validation.

For 10-fold cross validation, we pick up $k = 30$, whose average cross validation test error is 2.67%. See the tabular result of scaled 10-fold Cross Validation.

Why choose larger k if several k has the same lowest error rate:

Under the circumstance that increasing k will not lead to rise of error rate (since the same lowest error we have observed in the experiment), we can use higher value of k to enhance the robustness of classifier. If more nearest neighbours are considered for majority voting without decreasing the error rate, we can determine the result of classification from a larger amount of information (neighbours) by using a greater k.

4.1.3 Report result of cross validation

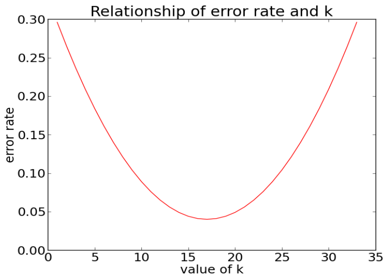
Considering the beautification of the type setting, the report of cross validation for various $k = 2, 4, .., 40$ was put in the appendix.

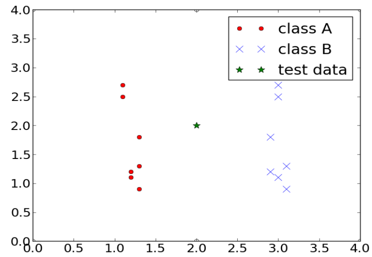
See report of various k with ORIGINAL data.

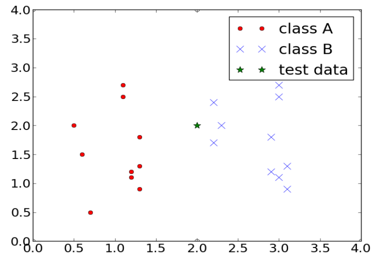
See report of various k with SCALED data.

4.1.4 Explain optimal error decreases with fold number

First, let us analyze the source of the error in KNN algorithm. The overall relationship of error and k could be represented by the the first graph below. In the first stage (before it comes to the optima) when k is small, inprecision of decision regions mainly comes from the reason that very few neighbours are considered while classifying a testing data object. Error will be arised in those ambiguous regions like the green testing data obejct in second figure).



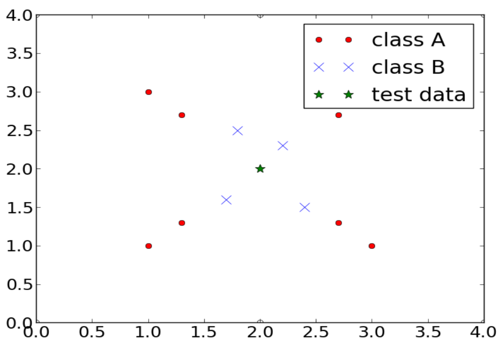


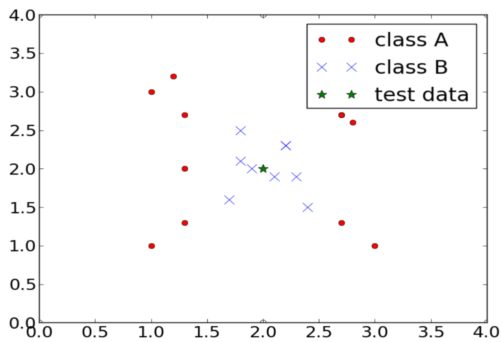


As the fold number rises, **the number of data objects as training data also increases**. (In the case of Iris-dataset, 2-fold has 75 objects; 5-fold has 120 objects; 10-fold has 135 objects). Since more data objects are used in KNN algorithm, which use training data (rather than a parametric model) as classifier, prediction of testing data object in ambiguous region will be more deterministic and the bias will be effectively decreased. Hence, accuracy of KNN classifier will be enhanced (at least non-decreased). And correspondingly, the optimal errors will decrease with the fold number.

4.1.5 Explain optimal k decreases with fold number

However, in the second stage when k is large, the error of algorithm is mainly coming from the reason that too many neighbours are considered. See the first image below where the classification of middle testing data is not desirable when $k = 10$.





But if more training data objects are given (casued by the increase of fold number), as shown in the second graph, the k must be large enough to make the same classification mistake as in the first case. Here, only when $k = 18$, the green test data will be mistaken as of class A. It is concluded from the comparison between two cases above that, **more training data (or equivalently, higher fold number) in KNN tolerate larger k for optimal error.** Geometrically, the minima would shift considerably rightwards if fold number increase and thereby more training data are used.

4.1.6 Listing of programs and solutions

```
KNN.py
    class
        KNNClassifier
        CrossValidation

    member
        __init__ [KNNClassifier]
        getFeatureVector [KNNClassifier]
        getLabel [KNNClassifier]
        getEuclidianDistance [KNNClassifier]
        getDistance [KNNClassifier]
        getKNearestNeighbours [KNNClassifier]
        majorityVoting [KNNClassifier]
        run [KNNClassifier]

        __init__ [CrossValidation]
        getRandomGroup [CrossValidation]
        getError [CrossValidation]
        run [CrossValidation]

    function
        printMatrix
        classMapping
        readMatrix
        scaling
        getSolution_4_1_1_
        getSolution_4_1_2_
        getSolution_4_1_3_
        main
```

A Supplementary proof

A.1 Proof of linearity of $E(x+y)$

Based on definition of expectation, we have

$$E(x+y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x+y)P(x,y)dxdy \tag{1}$$

$$E(x) = \int_{-\infty}^{+\infty} xP(x)dx \tag{2}$$

$$E(y) = \int_{-\infty}^{+\infty} yP(y)dy \tag{3}$$

Note that we use μ as abbreviation of expectation of certain random variable.
Then we start manipulating $E(x+y)$ from (1)

$$\begin{aligned} E(x+y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Big[xP(x,y) + yP(x,y)\Big] dxdy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xP(x,y)dxdy + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} yP(x,y)dxdy \\ &= \int_{-\infty}^{+\infty} xP(x)dx + \int_{-\infty}^{+\infty} yP(y)dy \end{aligned} \tag{4}$$

Note that last derivation above is based on sum rule of probability.
By taking (2) and (3) for (4), we solve the proof

$$E(x+y) = E(x) + E(y) \tag{5}$$

B Result for Cross Validation

B.1 Cross Validation for KNN with ORIGINAL input

B.1.1 2-fold

item name	average error	error of each group
2-NN 2-fold CV	8.00%	(8.00%, 8.00%)
3-NN 2-fold CV	6.00%	(4.00%, 8.00%)
4-NN 2-fold CV	4.67%	(4.00%, 5.33%)
5-NN 2-fold CV	6.00%	(4.00%, 8.00%)
6-NN 2-fold CV	5.33%	(2.67%, 8.00%)
7-NN 2-fold CV	6.00%	(5.33%, 6.67%)
8-NN 2-fold CV	4.00%	(2.67%, 5.33%)
9-NN 2-fold CV	4.67%	(4.00%, 5.33%)
10-NN 2-fold CV	4.67%	(2.67%, 6.67%)
11-NN 2-fold CV	6.00%	(5.33%, 6.67%)
12-NN 2-fold CV	6.00%	(5.33%, 6.67%)
13-NN 2-fold CV	6.00%	(6.67%, 5.33%)
14-NN 2-fold CV	6.00%	(5.33%, 6.67%)
15-NN 2-fold CV	6.67%	(6.67%, 6.67%)
16-NN 2-fold CV	6.00%	(6.67%, 5.33%)
17-NN 2-fold CV	5.33%	(6.67%, 4.00%)
18-NN 2-fold CV	4.67%	(5.33%, 4.00%)
19-NN 2-fold CV	6.00%	(6.67%, 5.33%)
20-NN 2-fold CV	6.00%	(6.67%, 5.33%)
21-NN 2-fold CV	6.00%	(6.67%, 5.33%)
22-NN 2-fold CV	6.67%	(5.33%, 8.00%)
23-NN 2-fold CV	5.33%	(5.33%, 5.33%)
24-NN 2-fold CV	5.33%	(5.33%, 5.33%)
25-NN 2-fold CV	6.00%	(5.33%, 6.67%)
26-NN 2-fold CV	7.33%	(6.67%, 8.00%)
27-NN 2-fold CV	7.33%	(6.67%, 8.00%)
28-NN 2-fold CV	8.00%	(8.00%, 8.00%)
29-NN 2-fold CV	8.67%	(8.00%, 9.33%)
30-NN 2-fold CV	9.33%	(9.33%, 9.33%)
31-NN 2-fold CV	9.33%	(9.33%, 9.33%)
32-NN 2-fold CV	9.33%	(9.33%, 9.33%)
33-NN 2-fold CV	10.00%	(9.33%, 10.67%)
34-NN 2-fold CV	9.33%	(9.33%, 9.33%)
35-NN 2-fold CV	8.67%	(8.00%, 9.33%)
36-NN 2-fold CV	10.67%	(10.67%, 10.67%)
37-NN 2-fold CV	10.67%	(10.67%, 10.67%)
38-NN 2-fold CV	11.33%	(10.67%, 12.00%)
39-NN 2-fold CV	11.33%	(10.67%, 12.00%)

B.1.2 5-fold

item name	average error	error of each group
2-NN 5-fold CV	5.33%	(3.33%, 6.67%, 6.67%, 10.00%, 0.00%)
3-NN 5-fold CV	3.33%	(3.33%, 3.33%, 6.67%, 3.33%, 0.00%)
4-NN 5-fold CV	2.67%	(3.33%, 3.33%, 3.33%, 3.33%, 0.00%)
5-NN 5-fold CV	2.67%	(3.33%, 0.00%, 6.67%, 3.33%, 0.00%)
6-NN 5-fold CV	2.00%	(3.33%, 0.00%, 3.33%, 3.33%, 0.00%)
7-NN 5-fold CV	2.00%	(3.33%, 0.00%, 3.33%, 3.33%, 0.00%)
8-NN 5-fold CV	3.33%	(3.33%, 0.00%, 6.67%, 6.67%, 0.00%)
9-NN 5-fold CV	2.67%	(3.33%, 0.00%, 3.33%, 6.67%, 0.00%)
10-NN 5-fold CV	2.00%	(3.33%, 0.00%, 0.00%, 6.67%, 0.00%)
11-NN 5-fold CV	2.00%	(6.67%, 0.00%, 0.00%, 3.33%, 0.00%)
12-NN 5-fold CV	2.00%	(6.67%, 0.00%, 0.00%, 3.33%, 0.00%)
13-NN 5-fold CV	2.67%	(6.67%, 0.00%, 3.33%, 3.33%, 0.00%)
14-NN 5-fold CV	3.33%	(6.67%, 0.00%, 3.33%, 6.67%, 0.00%)
15-NN 5-fold CV	3.33%	(6.67%, 0.00%, 6.67%, 3.33%, 0.00%)
16-NN 5-fold CV	3.33%	(6.67%, 0.00%, 6.67%, 3.33%, 0.00%)
17-NN 5-fold CV	3.33%	(6.67%, 0.00%, 6.67%, 3.33%, 0.00%)
18-NN 5-fold CV	2.67%	(6.67%, 0.00%, 3.33%, 3.33%, 0.00%)
19-NN 5-fold CV	3.33%	(6.67%, 0.00%, 6.67%, 3.33%, 0.00%)
20-NN 5-fold CV	4.00%	(6.67%, 0.00%, 6.67%, 6.67%, 0.00%)
21-NN 5-fold CV	3.33%	(6.67%, 0.00%, 6.67%, 3.33%, 0.00%)
22-NN 5-fold CV	4.00%	(6.67%, 0.00%, 6.67%, 6.67%, 0.00%)
23-NN 5-fold CV	4.00%	(6.67%, 0.00%, 6.67%, 6.67%, 0.00%)
24-NN 5-fold CV	5.33%	(6.67%, 3.33%, 10.00%, 6.67%, 0.00%)
25-NN 5-fold CV	4.67%	(10.00%, 3.33%, 6.67%, 3.33%, 0.00%)
26-NN 5-fold CV	5.33%	(10.00%, 3.33%, 6.67%, 6.67%, 0.00%)
27-NN 5-fold CV	5.33%	(10.00%, 3.33%, 6.67%, 6.67%, 0.00%)
28-NN 5-fold CV	5.33%	(10.00%, 3.33%, 6.67%, 6.67%, 0.00%)
29-NN 5-fold CV	6.00%	(10.00%, 3.33%, 10.00%, 6.67%, 0.00%)
30-NN 5-fold CV	6.00%	(10.00%, 3.33%, 6.67%, 10.00%, 0.00%)
31-NN 5-fold CV	6.00%	(10.00%, 3.33%, 10.00%, 6.67%, 0.00%)
32-NN 5-fold CV	6.00%	(10.00%, 3.33%, 6.67%, 10.00%, 0.00%)
33-NN 5-fold CV	6.67%	(10.00%, 3.33%, 10.00%, 10.00%, 0.00%)
34-NN 5-fold CV	4.67%	(6.67%, 3.33%, 6.67%, 6.67%, 0.00%)
35-NN 5-fold CV	5.33%	(10.00%, 3.33%, 6.67%, 6.67%, 0.00%)
36-NN 5-fold CV	5.33%	(10.00%, 3.33%, 6.67%, 6.67%, 0.00%)
37-NN 5-fold CV	5.33%	(10.00%, 3.33%, 6.67%, 6.67%, 0.00%)
38-NN 5-fold CV	7.33%	(10.00%, 6.67%, 13.33%, 6.67%, 0.00%)
39-NN 5-fold CV	5.33%	(10.00%, 6.67%, 6.67%, 3.33%, 0.00%)

B.1.3 10-fold

item	AVG	error of each group
2-NN	4.67%	(0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 0.00%, 13.33%, 6.67%, 0.00%, 0.00%)
3-NN	3.33%	(0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 0.00%, 6.67%, 0.00%, 0.00%, 0.00%)
4-NN	3.33%	(0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 0.00%, 6.67%, 0.00%, 0.00%, 0.00%)
5-NN	3.33%	(0.00%, 6.67%, 0.00%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%, 0.00%)
6-NN	3.33%	(0.00%, 6.67%, 0.00%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%, 0.00%)
7-NN	3.33%	(0.00%, 6.67%, 0.00%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%, 0.00%)
8-NN	3.33%	(0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
9-NN	2.67%	(0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 6.67%, 0.00%, 0.00%)
10-NN	3.33%	(0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
11-NN	3.33%	(0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
12-NN	2.67%	(0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 6.67%, 0.00%, 0.00%)
13-NN	2.00%	(0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 0.00%, 0.00%, 0.00%)
14-NN	2.67%	(0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 13.33%, 6.67%, 0.00%, 0.00%, 0.00%)
15-NN	2.67%	(6.67%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 0.00%, 0.00%, 0.00%)
16-NN	2.67%	(0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 6.67%, 0.00%, 0.00%)
17-NN	2.67%	(6.67%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 0.00%, 0.00%, 0.00%)
18-NN	2.00%	(0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 0.00%, 0.00%, 0.00%)
19-NN	2.67%	(6.67%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 0.00%, 0.00%, 0.00%)
20-NN	2.00%	(0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 0.00%, 0.00%, 0.00%)
21-NN	3.33%	(6.67%, 6.67%, 0.00%, 0.00%, 6.67%, 6.67%, 6.67%, 0.00%, 0.00%, 0.00%)
22-NN	3.33%	(6.67%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 6.67%, 0.00%, 0.00%)
23-NN	2.67%	(6.67%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 0.00%, 0.00%, 0.00%)
24-NN	4.00%	(6.67%, 6.67%, 0.00%, 0.00%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
25-NN	3.33%	(6.67%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 6.67%, 0.00%, 0.00%)
26-NN	4.00%	(6.67%, 6.67%, 0.00%, 0.00%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
27-NN	3.33%	(6.67%, 6.67%, 0.00%, 0.00%, 0.00%, 13.33%, 6.67%, 0.00%, 0.00%, 0.00%)
28-NN	4.67%	(6.67%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
29-NN	5.33%	(13.33%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
30-NN	5.33%	(13.33%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
31-NN	5.33%	(13.33%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
32-NN	5.33%	(13.33%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
33-NN	5.33%	(13.33%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
34-NN	4.67%	(6.67%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
35-NN	4.67%	(6.67%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
36-NN	5.33%	(6.67%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 13.33%, 6.67%, 0.00%, 0.00%)
37-NN	5.33%	(6.67%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 13.33%, 6.67%, 0.00%, 0.00%)
38-NN	5.33%	(6.67%, 6.67%, 0.00%, 6.67%, 0.00%, 20.00%, 6.67%, 6.67%, 0.00%, 0.00%)
39-NN	5.33%	(13.33%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)

B.2 Cross Validation for KNN with SCALED input

B.2.1 2-fold

item name	average error	error of each group
2NN 2-fold CV	6.00%	(4.00%, 8.00%)
3NN 2-fold CV	4.67%	(5.33%, 4.00%)
4NN 2-fold CV	3.33%	(4.00%, 2.67%)
5NN 2-fold CV	4.67%	(4.00%, 5.33%)
6NN 2-fold CV	4.67%	(5.33%, 4.00%)
7NN 2-fold CV	4.67%	(5.33%, 4.00%)
8NN 2-fold CV	4.00%	(4.00%, 4.00%)
9NN 2-fold CV	4.67%	(4.00%, 5.33%)
10NN 2-fold CV	4.00%	(4.00%, 4.00%)
11NN 2-fold CV	2.67%	(4.00%, 1.33%)
12NN 2-fold CV	3.33%	(4.00%, 2.67%)
13NN 2-fold CV	4.00%	(5.33%, 2.67%)
14NN 2-fold CV	5.33%	(5.33%, 5.33%)
15NN 2-fold CV	4.67%	(5.33%, 4.00%)
16NN 2-fold CV	6.67%	(5.33%, 8.00%)
17NN 2-fold CV	6.67%	(6.67%, 6.67%)
18NN 2-fold CV	4.67%	(2.67%, 6.67%)
19NN 2-fold CV	5.33%	(5.33%, 5.33%)
20NN 2-fold CV	5.33%	(5.33%, 5.33%)
21NN 2-fold CV	5.33%	(5.33%, 5.33%)
22NN 2-fold CV	6.00%	(5.33%, 6.67%)
23NN 2-fold CV	6.00%	(5.33%, 6.67%)
24NN 2-fold CV	6.67%	(5.33%, 8.00%)
25NN 2-fold CV	6.67%	(5.33%, 8.00%)
26NN 2-fold CV	8.00%	(8.00%, 8.00%)
27NN 2-fold CV	8.67%	(8.00%, 9.33%)
28NN 2-fold CV	9.33%	(9.33%, 9.33%)
29NN 2-fold CV	9.33%	(10.67%, 8.00%)
30NN 2-fold CV	11.33%	(12.00%, 10.67%)
31NN 2-fold CV	11.33%	(12.00%, 10.67%)
32NN 2-fold CV	11.33%	(12.00%, 10.67%)
33NN 2-fold CV	12.00%	(13.33%, 10.67%)
34NN 2-fold CV	12.00%	(13.33%, 10.67%)
35NN 2-fold CV	12.00%	(13.33%, 10.67%)
36NN 2-fold CV	12.00%	(13.33%, 10.67%)
37NN 2-fold CV	12.00%	(13.33%, 10.67%)
38NN 2-fold CV	12.00%	(12.00%, 12.00%)
39NN 2-fold CV	12.67%	(13.33%, 12.00%)

B.2.2 5-fold

item name	average error	error of each group
2NN 5-fold CV	4.00%	(3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
3NN 5-fold CV	4.67%	(3.33%, 3.33%, 6.67%, 10.00%, 0.00%)
4NN 5-fold CV	3.33%	(3.33%, 3.33%, 0.00%, 10.00%, 0.00%)
5NN 5-fold CV	4.00%	(3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
6NN 5-fold CV	3.33%	(3.33%, 3.33%, 0.00%, 10.00%, 0.00%)
7NN 5-fold CV	4.00%	(3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
8NN 5-fold CV	4.00%	(3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
9NN 5-fold CV	4.00%	(3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
10NN 5-fold CV	4.00%	(3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
11NN 5-fold CV	4.00%	(3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
12NN 5-fold CV	4.00%	(3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
13NN 5-fold CV	4.00%	(3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
14NN 5-fold CV	3.33%	(3.33%, 3.33%, 0.00%, 10.00%, 0.00%)
15NN 5-fold CV	2.67%	(3.33%, 3.33%, 0.00%, 6.67%, 0.00%)
16NN 5-fold CV	2.67%	(3.33%, 3.33%, 0.00%, 6.67%, 0.00%)
17NN 5-fold CV	4.00%	(3.33%, 3.33%, 6.67%, 6.67%, 0.00%)
18NN 5-fold CV	3.33%	(3.33%, 3.33%, 3.33%, 6.67%, 0.00%)
19NN 5-fold CV	5.33%	(6.67%, 3.33%, 6.67%, 10.00%, 0.00%)
20NN 5-fold CV	4.00%	(6.67%, 3.33%, 0.00%, 10.00%, 0.00%)
21NN 5-fold CV	5.33%	(6.67%, 3.33%, 6.67%, 10.00%, 0.00%)
22NN 5-fold CV	4.67%	(6.67%, 3.33%, 3.33%, 10.00%, 0.00%)
23NN 5-fold CV	4.67%	(6.67%, 3.33%, 6.67%, 6.67%, 0.00%)
24NN 5-fold CV	4.67%	(6.67%, 3.33%, 3.33%, 10.00%, 0.00%)
25NN 5-fold CV	5.33%	(6.67%, 3.33%, 6.67%, 10.00%, 0.00%)
26NN 5-fold CV	4.00%	(6.67%, 3.33%, 3.33%, 6.67%, 0.00%)
27NN 5-fold CV	4.00%	(6.67%, 3.33%, 3.33%, 6.67%, 0.00%)
28NN 5-fold CV	4.00%	(6.67%, 3.33%, 3.33%, 6.67%, 0.00%)
29NN 5-fold CV	4.00%	(6.67%, 3.33%, 3.33%, 6.67%, 0.00%)
30NN 5-fold CV	4.00%	(3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
31NN 5-fold CV	4.00%	(3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
32NN 5-fold CV	4.00%	(3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
33NN 5-fold CV	4.00%	(3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
34NN 5-fold CV	4.00%	(3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
35NN 5-fold CV	5.33%	(10.00%, 3.33%, 3.33%, 10.00%, 0.00%)
36NN 5-fold CV	5.33%	(10.00%, 3.33%, 3.33%, 10.00%, 0.00%)
37NN 5-fold CV	5.33%	(10.00%, 3.33%, 3.33%, 10.00%, 0.00%)
38NN 5-fold CV	4.67%	(6.67%, 3.33%, 3.33%, 10.00%, 0.00%)
39NN 5-fold CV	4.67%	(6.67%, 3.33%, 3.33%, 10.00%, 0.00%)

B.2.3 10-fold

item	AVE	error of each group
2NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
3NN	4.67%	(0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
4NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
5NN	4.67%	(0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
6NN	3.33%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
7NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
8NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
9NN	4.00%	(0.00%, 6.67%, 0.00%, 0.00%, 6.67%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
10NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
11NN	4.67%	(0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
12NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
13NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
14NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
15NN	3.33%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 0.00%, 0.00%, 0.00%)
16NN	3.33%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 0.00%, 13.33%, 6.67%, 0.00%, 0.00%)
17NN	3.33%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 0.00%, 0.00%, 0.00%)
18NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
19NN	4.67%	(0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
20NN	4.67%	(0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
21NN	4.67%	(0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
22NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
23NN	4.67%	(0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
24NN	3.33%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
25NN	5.33%	(6.67%, 6.67%, 0.00%, 6.67%, 6.67%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
26NN	4.67%	(6.67%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
27NN	5.33%	(6.67%, 6.67%, 0.00%, 6.67%, 6.67%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
28NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
29NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
30NN	2.67%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 0.00%, 13.33%, 0.00%, 0.00%, 0.00%)
31NN	4.00%	(6.67%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 0.00%, 0.00%, 0.00%)
32NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 6.67%, 0.00%, 0.00%)
33NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 6.67%, 0.00%, 0.00%)
34NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 6.67%, 0.00%, 0.00%)
35NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 6.67%, 0.00%, 0.00%)
36NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 6.67%, 0.00%, 0.00%)
37NN	4.00%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 6.67%, 0.00%, 0.00%)
38NN	5.33%	(0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 20.00%, 13.33%, 6.67%, 0.00%, 0.00%)
39NN	5.33%	(13.33%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 6.67%, 0.00%, 0.00%)

B.3 Report CV result for various k

B.3.1 with ORIGINAL input

2NN
2-NN 2-fold CV 8.00% (8.00%, 8.00%)
2-NN 5-fold CV 5.33% (3.33%, 6.67%, 6.67%, 10.00%, 0.00%)
2-NN 10-fold CV 4.67% (0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 0.00%, 13.33%, 6.67%, 0.00%, 0.00%)
4NN
4-NN 2-fold CV 4.67% (4.00%, 5.33%)
4-NN 5-fold CV 2.67% (3.33%, 3.33%, 3.33%, 3.33%, 0.00%)
4-NN 10-fold CV 3.33% (0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 0.00%, 6.67%, 0.00%, 0.00%, 0.00%)
6NN
6-NN 2-fold CV 5.33% (2.67%, 8.00%)
6-NN 5-fold CV 2.00% (3.33%, 0.00%, 3.33%, 3.33%, 0.00%)
6-NN 10-fold CV 3.33% (0.00%, 6.67%, 0.00%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%, 0.00%)
8NN
8-NN 2-fold CV 4.00% (2.67%, 5.33%)
8-NN 5-fold CV 3.33% (3.33%, 0.00%, 6.67%, 6.67%, 0.00%)
8-NN 10-fold CV 3.33% (0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
10NN
10-NN 2-fold CV 4.67% (2.67%, 6.67%)
10-NN 5-fold CV 2.00% (3.33%, 0.00%, 0.00%, 6.67%, 0.00%)
10-NN 10-fold CV 3.33% (0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
12NN
12-NN 2-fold CV 6.00% (5.33%, 6.67%)
12-NN 5-fold CV 2.00% (6.67%, 0.00%, 0.00%, 3.33%, 0.00%)
12-NN 10-fold CV 2.67% (0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 6.67%, 0.00%, 0.00%)
14NN
14-NN 2-fold CV 6.00% (5.33%, 6.67%)
14-NN 5-fold CV 3.33% (6.67%, 0.00%, 3.33%, 6.67%, 0.00%)
14-NN 10-fold CV 2.67% (0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 13.33%, 6.67%, 0.00%, 0.00%, 0.00%)
16NN
16-NN 2-fold CV 6.00% (6.67%, 5.33%)
16-NN 5-fold CV 3.33% (6.67%, 0.00%, 6.67%, 3.33%, 0.00%)
16-NN 10-fold CV 2.67% (0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 6.67%, 0.00%, 0.00%)
18NN
18-NN 2-fold CV 4.67% (5.33%, 4.00%)
18-NN 5-fold CV 2.67% (6.67%, 0.00%, 3.33%, 3.33%, 0.00%)
18-NN 10-fold CV 2.00% (0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 0.00%, 0.00%, 0.00%)
20NN
20-NN 2-fold CV 6.00% (6.67%, 5.33%)
20-NN 5-fold CV 4.00% (6.67%, 0.00%, 6.67%, 6.67%, 0.00%)
20-NN 10-fold CV 2.00% (0.00%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 0.00%, 0.00%, 0.00%)
22NN
22-NN 2-fold CV 6.67% (5.33%, 8.00%)
22-NN 5-fold CV 4.00% (6.67%, 0.00%, 6.67%, 6.67%, 0.00%)
22-NN 10-fold CV 3.33% (6.67%, 6.67%, 0.00%, 0.00%, 0.00%, 6.67%, 6.67%, 6.67%, 0.00%, 0.00%)
24NN
24-NN 2-fold CV 5.33% (5.33%, 5.33%)
24-NN 5-fold CV 5.33% (6.67%, 3.33%, 10.00%, 6.67%, 0.00%)
24-NN 10-fold CV 4.00% (6.67%, 6.67%, 0.00%, 0.00%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)

26NN
26-NN 2-fold CV 7.33% (6.67%, 8.00%)
26-NN 5-fold CV 5.33% (10.00%, 3.33%, 6.67%, 6.67%, 0.00%)
26-NN 10-fold CV 4.00% (6.67%, 6.67%, 0.00%, 0.00%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
28NN
28-NN 2-fold CV 8.00% (8.00%, 8.00%)
28-NN 5-fold CV 5.33% (10.00%, 3.33%, 6.67%, 6.67%, 0.00%)
28-NN 10-fold CV 4.67% (6.67%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
30NN
30-NN 2-fold CV 9.33% (9.33%, 9.33%)
30-NN 5-fold CV 6.00% (10.00%, 3.33%, 6.67%, 10.00%, 0.00%)
30-NN 10-fold CV 5.33% (13.33%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
32NN
32-NN 2-fold CV 9.33% (9.33%, 9.33%)
32-NN 5-fold CV 6.00% (10.00%, 3.33%, 6.67%, 10.00%, 0.00%)
32-NN 10-fold CV 5.33% (13.33%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
34NN
34-NN 2-fold CV 9.33% (9.33%, 9.33%)
34-NN 5-fold CV 4.67% (6.67%, 3.33%, 6.67%, 6.67%, 0.00%)
34-NN 10-fold CV 4.67% (6.67%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)
36NN
36-NN 2-fold CV 10.67% (10.67%, 10.67%)
36-NN 5-fold CV 5.33% (10.00%, 3.33%, 6.67%, 6.67%, 0.00%)
36-NN 10-fold CV 5.33% (6.67%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 13.33%, 6.67%, 0.00%, 0.00%)
38NN
38-NN 2-fold CV 11.33% (10.67%, 12.00%)
38-NN 5-fold CV 7.33% (10.00%, 6.67%, 13.33%, 6.67%, 0.00%)
38-NN 10-fold CV 5.33% (6.67%, 6.67%, 0.00%, 6.67%, 0.00%, 20.00%, 6.67%, 6.67%, 0.00%, 0.00%)
40NN
40-NN 2-fold CV 11.33% (10.67%, 12.00%)
40-NN 5-fold CV 6.67% (10.00%, 6.67%, 13.33%, 3.33%, 0.00%)
40-NN 10-fold CV 4.67% (6.67%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 6.67%, 6.67%, 0.00%, 0.00%)

B.3.2 with SCALED input

Then, we present the result for the SCALED data,
2NN
2-NN 2-fold CV 6.00% (4.00%, 8.00%)
2-NN 5-fold CV 4.00% (3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
2-NN 10-fold CV 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
4NN
4-NN 2-fold CV 3.33% (4.00%, 2.67%)
4-NN 5-fold CV 3.33% (3.33%, 3.33%, 0.00%, 10.00%, 0.00%)
4-NN 10-fold CV 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
6NN
6-NN 2-fold CV 4.67% (5.33%, 4.00%)
6-NN 5-fold CV 3.33% (3.33%, 3.33%, 0.00%, 10.00%, 0.00%)
6-NN 10-fold CV 3.33% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
8NN
8-NN 2-fold CV 4.00% (4.00%, 4.00%)
8-NN 5-fold CV 4.00% (3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
8-NN 10-fold CV 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
10NN
10-NN 2-fold CV 4.00% (4.00%, 4.00%)
10-NN 5-fold CV 4.00% (3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
10-NN 10-fold CV 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
12NN
12-NN 2-fold CV 3.33% (4.00%, 2.67%)
12-NN 5-fold CV 4.00% (3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
12-NN 10-fold CV 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
14NN
14-NN 2-fold CV 5.33% (5.33%, 5.33%)
14-NN 5-fold CV 3.33% (3.33%, 3.33%, 0.00%, 10.00%, 0.00%)
14-NN 10-fold CV 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
16NN
16-NN 2-fold CV 6.67% (5.33%, 8.00%)
16-NN 5-fold CV 2.67% (3.33%, 3.33%, 0.00%, 6.67%, 0.00%)
16-NN 10-fold CV 3.33% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 0.00%, 13.33%, 6.67%, 0.00%, 0.00%)
18NN
18-NN 2-fold CV 4.67% (2.67%, 6.67%)
18-NN 5-fold CV 3.33% (3.33%, 3.33%, 3.33%, 6.67%, 0.00%)
18-NN 10-fold CV 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
20NN
20-NN 2-fold CV 5.33% (5.33%, 5.33%)
20-NN 5-fold CV 4.00% (6.67%, 3.33%, 0.00%, 10.00%, 0.00%)
20-NN 10-fold CV 4.67% (0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
22NN
22-NN 2-fold CV 6.00% (5.33%, 6.67%)
22-NN 5-fold CV 4.67% (6.67%, 3.33%, 3.33%, 10.00%, 0.00%)
22-NN 10-fold CV 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
24NN
24-NN 2-fold CV 6.67% (5.33%, 8.00%)
24-NN 5-fold CV 4.67% (6.67%, 3.33%, 3.33%, 10.00%, 0.00%)
24-NN 10-fold CV 3.33% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
26NN
26-NN 2-fold CV 8.00% (8.00%, 8.00%)
26-NN 5-fold CV 4.00% (6.67%, 3.33%, 3.33%, 6.67%, 0.00%)
26-NN 10-fold CV 4.67% (6.67%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)

28NN
28-NN 2-fold CV 9.33% (9.33%, 9.33%)
28-NN 5-fold CV 4.00% (6.67%, 3.33%, 3.33%, 6.67%, 0.00%)
28-NN 10-fold CV 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
30NN
30-NN 2-fold CV 11.33% (12.00%, 10.67%)
30-NN 5-fold CV 4.00% (3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
30-NN 10-fold CV 2.67% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 0.00%, 13.33%, 0.00%, 0.00%, 0.00%)
32NN
32-NN 2-fold CV 11.33% (12.00%, 10.67%)
32-NN 5-fold CV 4.00% (3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
32-NN 10-fold CV 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 6.67%, 0.00%, 0.00%)
34NN
34-NN 2-fold CV 12.00% (13.33%, 10.67%)
34-NN 5-fold CV 4.00% (3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
34-NN 10-fold CV 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 6.67%, 0.00%, 0.00%)
36NN
36-NN 2-fold CV 12.00% (13.33%, 10.67%)
36-NN 5-fold CV 5.33% (10.00%, 3.33%, 3.33%, 10.00%, 0.00%)
36-NN 10-fold CV 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 6.67%, 0.00%, 0.00%)
38NN
38-NN 2-fold CV 12.00% (12.00%, 12.00%)
38-NN 5-fold CV 4.67% (6.67%, 3.33%, 3.33%, 10.00%, 0.00%)
38-NN 10-fold CV 5.33% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 20.00%, 13.33%, 6.67%, 0.00%, 0.00%)
40NN
40-NN 2-fold CV 12.67% (13.33%, 12.00%)
40-NN 5-fold CV 6.00% (6.67%, 6.67%, 3.33%, 10.00%, 3.33%)
40-NN 10-fold CV 5.33% (6.67%, 6.67%, 0.00%, 6.67%, 0.00%, 13.33%, 13.33%, 6.67%, 0.00%, 0.00%)