Contents

1 Probabilities

1.1 Covariance of Sum

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

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

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

$$\mu_{x+y} = \mu_x + \mu_y \tag{2}$$

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

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

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 equation $(3)-(5)$, we have

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

By definition of variance and covariance, we obtain

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

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

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

Then, take $(4)-(6)$ into (3), we derive the desired result

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

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_1 = {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

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

Because of iid, we can easily work out the joint distribution of $P(S_1,S_2)$

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

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

Next, we use another way to describe the distribution

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

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

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

Obviously, the most likely gender condition of the neibhours 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 $P(NB = 1|NG \geq 1)$.

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

$$P(NB = 1|NG \geq 1) \ = \frac{P(NB = 1|NG = 1)}{P(NG \geq 1)} \ = \frac{\frac{1}{2}}{\frac{3}{4}} \ = \frac{2}{3} \tag{19}$$

That is to say, the probability of one child being boy is $\frac{2}{3}$, given the condition 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. Hence, we have the probability of the other child being a boy, given seeing one girl, to be

$$P(S_2 = b|S_1 = g) = \frac{P(S_2 = b, S_1 = g)}{P(S_1 = g)} \ = \frac{P(S_2 = b)P(S_1 = g)}{P(S_1 = g)} \ = P(S_2 = b) \ = \frac{1}{2} \tag{20}$$

Even we happend to see gender of one child, the probability of other one child being boy is $\frac{1}{2}$.

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

$$p(x|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{D}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \; exp[-\frac{1}{2}(x-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(x-\boldsymbol{\mu})] \tag{21}$$

Then we derive the likelihood function

$$\begin{aligned} L(\boldsymbol{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{n=1}^{N} \frac{1}{(2\pi)^{\frac{D}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} exp[-\frac{1}{2}(\boldsymbol{x_n}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_n}-\boldsymbol{\mu})] \\ &= \frac{1}{(2\pi)^{\frac{ND}{2}}|\boldsymbol{\Sigma}|^{\frac{N}{2}}} exp\Big[\sum_{n=1}^{N}\Big(-\frac{1}{2}(\boldsymbol{x_n}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_n}-\boldsymbol{\mu})\Big)\Big] \end{aligned} \tag{22}$$

1.3.2 Extrenum Solution

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

$$ln(L(\boldsymbol{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma})) = -\frac{ND}{2}ln(2\pi) - \frac{N}{2}ln|\boldsymbol{\Sigma}| + \sum_{n=1}^{N}[-\frac{1}{2}(\boldsymbol{x_n}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_n}-\boldsymbol{\mu})] \tag{23}$$

Then we take derivative to the log likelihood function with regard to $\boldsymbol{\mu}$

$$\frac{\partial L(\boldsymbol{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}} = -\sum_{n=1}^{N}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_n}-\boldsymbol{\mu}) \tag{24}$$

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 $\boldsymbol{\mu}$. See proof of derivation for third term at appendix B.1.

Next, solve the (4), we get the result

$$\boldsymbol{\mu}_{ML} = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x_n} \tag{25}$$

Finally, we work out $\boldsymbol{\Sigma}_{ML}$

$$\boldsymbol{\Sigma}_{ML} = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x_n}-\boldsymbol{\mu}_{ML})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_n}-\boldsymbol{\mu}_{ML}) \tag{26}$$

1.3.3 The found extremum is maximum

We take the second derivative of likelihood function with regard to mean of Gaussian distribution, see specific derivation at Appendix B.2.

$$\frac{\partial L(\boldsymbol{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}^2} = \tag{27}$$

As presented by (7), the second derivative of likelihood function is no greater than zero at its whole defined scope. Therefore, the likelihood function is a convex function.

1.3.4 Arbitrary Order

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

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 $L(X_1, X_2.., X_N, \theta)$

$$L(X_1, X_2.., X_N, \theta) = \prod_{n=1}^{N} P(x_n|\theta) = \prod_{n=1}^{N} \theta e^{-\theta x_n} \tag{28}$$

Then take the logarithm of likelihood

$$ln(L(X_1, X_2.., X_N, \theta)) = ln(\prod_{n=1}^{N} \theta e^{-\theta x_n}) = \sum_{n=1}^{N}(-\theta x_n + ln(\theta)) \tag{29}$$

Next, take derivative of log likelihood with regard to θ

$$\frac{d(ln(L(X_1, X_2.., X_N, \theta)))}{d\theta} = \frac{d(\sum_{n=1}^{N}(-\theta x_n + ln(\theta)))}{d\theta} = \sum_{n=1}^{N}(-x_n + \frac{1}{\theta}) = \frac{N}{\theta} - \sum_{n=1}^{N} X_n \tag{30}$$

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

$$\frac{N}{\hat{\theta}} - \sum_{n=1}^{N} X_n = 0 \tag{31}$$

Solve the equation above, we have

$$\hat{\theta} = \frac{N}{\sum_{n=1}^{N} x_n} \tag{32}$$

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

By definition, we have

$$mean = \frac{\sum_{n=1}^{N} x_n}{N} \tag{33}$$

Based on the (1), we can easily figure out the relation between mean of datasets and $\hat{\theta}$, that is

$$mean \cdot \hat{\theta} = 1 \tag{34}$$

1.4.3 One instantiation

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

$$mean = \frac{\sum_{n=1}^{4} x_n}{4} = \frac{5+4+3+4}{4} = 4 \tag{35}$$

Based on the (2), we have $\hat{\theta}$

$$\hat{\theta} = \frac{1}{mean} = \frac{1}{4} \tag{36}$$

Therefore, the MLE solution for given datasets is

$$P(x|\theta) = \frac{1}{4} e^{-\frac{1}{4}x} \qquad x \geq 0, \theta > 0 \tag{37}$$

2 Decision Theory

2.1 Lower Bound for the Correct Classification

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

We start proof from pre-condition of the implication.

$$a \le b \tag{38}$$

Since a is non-negative number, we have

$$a^2 \le ab \tag{39}$$

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 inequation)

$$|a| \le (ab)^{\frac{1}{2}} \tag{40}$$

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

$$|a| = a \tag{41}$$

Hence, we derive the post-condition.

$$a \le (ab)^{\frac{1}{2}} \tag{42}$$

2.1.2 Proof of lower bound

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

$$\begin{aligned} p(mistake) &= p(\mathbf{x} \in \mathcal{R}_1, \mathcal{C}_2) + p(\mathbf{x} \in \mathcal{R}_2, \mathcal{C}_1) \\ &= \int_{\mathbf{x} \in \mathcal{R}_1} p(\mathbf{x}, \mathcal{C}_2) d\mathbf{x} + \int_{\mathbf{x} \in \mathcal{R}_2} p(\mathbf{x}, \mathcal{C}_1) d\mathbf{x} \end{aligned} \tag{43}$$

Since the decision region was chosen to minimise the probability of misclassification, say, we are about to minimize the intergrand in (43), that is

$$p(\mathbf{x}, \mathcal{C}_2) \le p(\mathbf{x}, \mathcal{C}_1), \mathbf{x} \in \mathcal{R}_1 \tag{44}$$

$$p(\mathbf{x}, \mathcal{C}_1) \le p(\mathbf{x}, \mathcal{C}_2), \mathbf{x} \in \mathcal{R}_2 \tag{45}$$

Next, according to the inequality (4), we can derive the followings from (7) and (8),

$$p(\mathbf{x}, \mathcal{C}_2) \le \big(p(\mathbf{x}, \mathcal{C}_1) \cdot p(\mathbf{x}, \mathcal{C}_2)\big)^{1/2}, \mathbf{x} \in \mathcal{R}_1 \tag{46}$$

$$p(\mathbf{x}, \mathcal{C}_1) \le \big(p(\mathbf{x}, \mathcal{C}_1) \cdot p(\mathbf{x}, \mathcal{C}_2)\big)^{1/2}, \mathbf{x} \in \mathcal{R}_2 \tag{47}$$

Then, based on (6),(9) and (10), we have,

$$\begin{aligned} p(mistake) &\le \int_{\mathbf{x} \in \mathcal{R}_1} \big(p(\mathbf{x}, \mathcal{C}_1) \cdot p(\mathbf{x}, \mathcal{C}_2)\big)^{1/2} d\mathbf{x} + \int_{\mathbf{x} \in \mathcal{R}_2} \big(p(\mathbf{x}, \mathcal{C}_1) \cdot p(\mathbf{x}, \mathcal{C}_2)\big)^{1/2} d\mathbf{x} \\ &= \int_{\mathbf{x} \in \mathcal{R}_1 \cup \mathcal{R}_2} \big(p(\mathbf{x}, \mathcal{C}_1) \cdot p(\mathbf{x}, \mathcal{C}_2)\big)^{1/2} d\mathbf{x} \end{aligned} \tag{48}$$

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

$$\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \tag{49}$$

Apply equation (12) to equation (11), we have

$$p(mistake) \le \int_{\mathbf{x} \in \mathcal{R}} \sqrt{p(\mathbf{x}, \mathcal{C}_1) \cdot p(\mathbf{x}, \mathcal{C}_2)} d\mathbf{x} \tag{50}$$

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,

$$S_W = \sum_{k}^{K} \sum_{\boldsymbol{x_n} \in \mathcal{C}_k} (\boldsymbol{x_n} - \mathbf{m}_k)(\boldsymbol{x_n} - \mathbf{m}_k)^T \tag{51}$$

$$S_B = \sum_{k}^{K} N_k (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})^T \tag{52}$$

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 normalized 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,

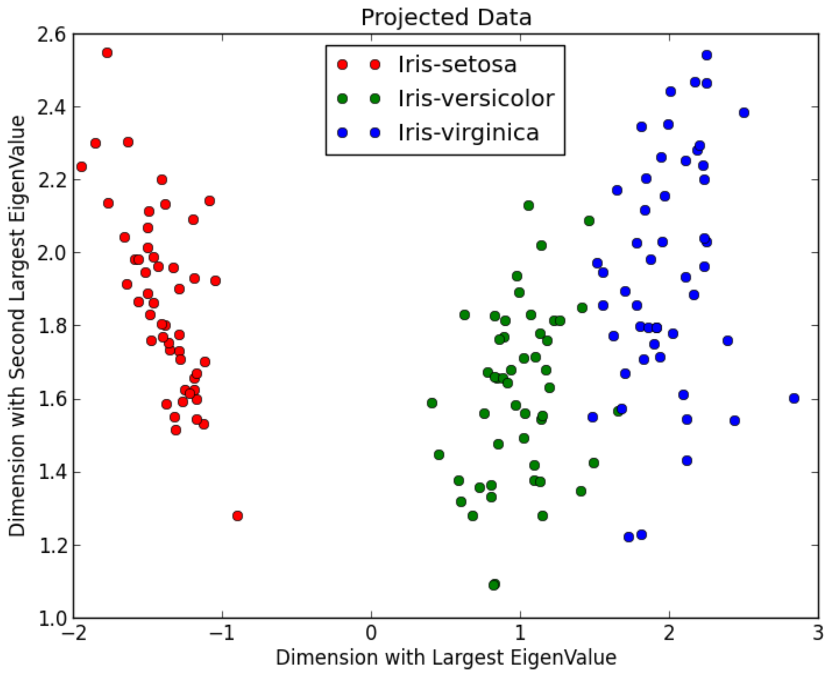
$$\lambda_1 = 32.1919292, \mathcal{V}_1 = (-0.20874183, -0.38620368, 0.5540117, 0.70735037)$$

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

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 believe in 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 projection.

4 Cross Validation and Classification

4.1 K-Nearest Neighbours Algorithm

4.1.1 Implement K-NN algorithm

See the codes for implementing the K-NN algorithm.

I also implement the min-max normalization for the original data for the reason that the KNN algorithm is vulnerable to bad scaling. (see the scaling function in KNN.py)

In the KNN.py, the function getSolution_4_1_1 is a test function for the KNN Implementation. It uses the last data of Iris datasets as the only testing data object and the rest as training dataset.

4.1.2 Apply Cross Validation

For 2-fold cross validation, we pick up $k = 11$, whose average cross validation test error is 2.66% . Click here to see the whole tabular result.

For 5-fold cross validation, we pick up $k = 16$, whose average cross validation test error is 2.66% . Click here to see the whole tabular result.

For 10-fold cross validation, we pick up $k = 30$, whose average cross validation test error is 2.66% . Click here to see the whole tabular result.

4.1.3 Report result of cross validation

2NN:
2-fold CV, average error: 6.00% (4.00%, 8.00%)
5-fold CV, average error: 4.00% (3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
10-fold CV, average error: 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
4NN:
2-fold CV, average error: 3.33% (4.00%, 2.67%)
5-fold CV, average error: 3.33% (3.33%, 3.33%, 0.00%, 10.00%, 0.00%)
10-fold CV, average error: 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
6NN:
2-fold CV, average error: 4.67% (5.33%, 4.00%)
5-fold CV, average error: 3.33% (3.33%, 3.33%, 0.00%, 10.00%, 0.00%)
10-fold CV, average error: 3.33% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
8NN:
2-fold CV, average error: 4.00% (4.00%, 4.00%)
5-fold CV, average error: 4.00% (3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
10-fold CV, average error: 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
10NN:
2-fold CV, average error: 4.00% (4.00%, 4.00%)
5-fold CV, average error: 4.00% (3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
10-fold CV, average error: 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
12NN:
2-fold CV, average error: 3.33% (4.00%, 2.67%)
5-fold CV, average error: 4.00% (3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
10-fold CV, average error: 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
14NN:
2-fold CV, average error: 5.33% (5.33%, 5.33%)
5-fold CV, average error: 3.33% (3.33%, 3.33%, 0.00%, 10.00%, 0.00%)
10-fold CV, average error: 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
16NN:
2-fold CV, average error: 6.67% (5.33%, 8.00%)
5-fold CV, average error: 2.67% (3.33%, 3.33%, 0.00%, 6.67%, 0.00%)
10-fold CV, average error: 3.33% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 0.00%, 13.33%, 6.67%, 0.00%, 0.00%)
18NN:
2-fold CV, average error: 4.67% (2.67%, 6.67%)
5-fold CV, average error: 3.33% (3.33%, 3.33%, 3.33%, 6.67%, 0.00%)
10-fold CV, average error: 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
20NN:
2-fold CV, average error: 5.33% (5.33%, 5.33%)
5-fold CV, average error: 4.00% (6.67%, 3.33%, 0.00%, 10.00%, 0.00%)
10-fold CV, average error: 4.67% (0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 6.67%, 20.00%, 0.00%, 0.00%, 0.00%)
22NN:
2-fold CV, average error: 6.00% (5.33%, 6.67%)
5-fold CV, average error: 4.67% (6.67%, 3.33%, 3.33%, 10.00%, 0.00%)

10-fold CV, average error: 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
24NN:
2-fold CV, average error: 6.67% (5.33%, 8.00%)
5-fold CV, average error: 4.67% (6.67%, 3.33%, 3.33%, 10.00%, 0.00%)
10-fold CV, average error: 3.33% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
26NN:
2-fold CV, average error: 8.00% (8.00%, 8.00%)
5-fold CV, average error: 4.00% (6.67%, 3.33%, 3.33%, 6.67%, 0.00%)
10-fold CV, average error: 4.67% (6.67%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
28NN:
2-fold CV, average error: 9.33% (9.33%, 9.33%)
5-fold CV, average error: 4.00% (6.67%, 3.33%, 3.33%, 6.67%, 0.00%)
10-fold CV, average error: 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 6.67%, 0.00%, 20.00%, 0.00%, 0.00%, 0.00%)
30NN:
2-fold CV, average error: 11.33% (12.00%, 10.67%)
5-fold CV, average error: 4.00% (3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
10-fold CV, average error: 2.67% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 0.00%, 13.33%, 0.00%, 0.00%, 0.00%)
32NN:
2-fold CV, average error: 11.33% (12.00%, 10.67%)
5-fold CV, average error: 4.00% (3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
10-fold CV, average error: 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 6.67%, 0.00%, 0.00%)
34NN:
2-fold CV, average error: 12.00% (13.33%, 10.67%)
5-fold CV, average error: 4.00% (3.33%, 3.33%, 3.33%, 10.00%, 0.00%)
10-fold CV, average error: 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 6.67%, 0.00%, 0.00%)
36NN:
2-fold CV, average error: 12.00% (13.33%, 10.67%)
5-fold CV, average error: 5.33% (10.00%, 3.33%, 3.33%, 10.00%, 0.00%)
10-fold CV, average error: 4.00% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 6.67%, 13.33%, 6.67%, 0.00%, 0.00%)
38NN:
2-fold CV, average error: 12.00% (12.00%, 12.00%)
5-fold CV, average error: 4.67% (6.67%, 3.33%, 3.33%, 10.00%, 0.00%)
10-fold CV, average error: 5.33% (0.00%, 6.67%, 0.00%, 6.67%, 0.00%, 20.00%, 13.33%, 6.67%, 0.00%, 0.00%)

4.1.4 Explain optimal error decreases with fold number

4.1.5 Explain optimal k decreases with fold number

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 for 1.1

A.1 Proof of linearity of μ_{x+y}

Based on definition of expectation, we have

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

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

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

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

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

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

$$\mu_{x+y} = \mu_x + \mu_y \tag{5}$$

B Supplementary proofs for 1.3

B.1 Specifics of first derivative to likelihood

B.2 Specifics of second derivative to likelihood

C Cross Validation result

C.1 2-fold cross validation

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%)

C.2 5-fold cross validation

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%)

C.3 10-fold cross validation

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