Question 1.

1.1. The solutions of the linear equations are $x_1 = -1, x_2 = 0, x_3 = 1$.

1.2. The matrix form (Ax=b) is $\begin{bmatrix} 2 & 2 & 3 \\ 1 & -1 & 0 \\ -1 & 2 & 1 \end{bmatrix} x = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$.

1.3. The rank of A is 3, since its row echelon form is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

1.4. $det(A) = \sum_{j=1}^{3} (-1)^{j+k} a_{jk} \det(A_{jk}) = -1.$

$A^{-1} = \frac{1}{det(A)} \begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{bmatrix} = \begin{bmatrix} 1 & -4 & -3 \\ 1 & -5 & -3 \\ -1 & 6 & 4 \end{bmatrix}$, where $A_{jk}$is the algebraic cofactor of a_{jk}.

1.5. $A^{-1}\mathrm{Ax} = A^{-1}b \quad \Rightarrow \quad x = \begin{bmatrix} 1 & -4 & -3 \\ 1 & -5 & -3 \\ -1 & 6 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$.

1.6. Inner product $\langle x, b \rangle = \sum_{j=1}^{3} x_j b_j = 1.$

Outer product $x \otimes b = x b^T = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} [1 \quad -1 \quad 2] = \begin{bmatrix} -1 & 1 & -2 \\ 0 & 0 & 0 \\ 1 & -1 & 2 \end{bmatrix}$.

1.7. $L_1\ norm$: $\|b\|_1 = \sum_{i=1}^{3} |b_i| = 4$.

$L_2\ norm$: $\|b\|_2 = \sqrt{\sum_{i=1}^{3} b_i^2} = \sqrt{6}$.

$L_\infty\ norm$: $\|b\|_\infty = \max_i |b_i| = 2$.

1.8. $y^T A y = [y_1 \quad y_2 \quad y_3] \begin{bmatrix} 2 & 2 & 3 \\ 1 & -1 & 0 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$

$= 2y_1^2 - y_2^2 + y_3^2 + 3y_1y_2 + 2y_1y_3 + 2y_2y_3$.

$\nabla_y y^T A y = \begin{bmatrix} \frac{\partial y^T A y}{\partial y_1} & \frac{\partial y^T A y}{\partial y_2} & \frac{\partial y^T A y}{\partial y_3} \end{bmatrix}$

$$= [4y_1 + 3y_2 + 2y_3 \quad 3y_1 - 2y_2 + 2y_3 \quad 2y_1 + 2y_2 + 2y_3]$$

$$= \begin{bmatrix} 4 & 3 & 2 \\ 3 & -2 & 2 \\ 2 & 2 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$

OR. $\nabla_y y^T A y = (A + A^T)y = \begin{bmatrix} 4 & 3 & 2 \\ 3 & -2 & 2 \\ 2 & 2 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$.

1.9. The matrix form $A_1 x = b$ is $\begin{bmatrix} 2 & 2 & 3 \\ 1 & -1 & 0 \\ -1 & 2 & 1 \\ -1 & 2 & 1 \end{bmatrix} x = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 1 \end{bmatrix}$.

1.10. The rank of A_1 is 3, because its row echelon form is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$.

1.11. No, it can't be solved. Because it is an overdetermined system, and the matrix A_1 can't be inversed.

1.12. The pseudo-inverse of A_1 is

$$A_1^{+} = (A_1^T A_1)^{-1} A_1^T$$

$$= \frac{1}{2} \begin{bmatrix} 43 & 51 & -62 \\ 51 & 61 & -74 \\ -62 & -74 & 90 \end{bmatrix} \begin{bmatrix} 2 & -1 & -1 & -1 \\ 2 & -1 & 2 & 2 \\ 3 & 0 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -4 & -3/2 & -3/2 \\ 1 & -5 & -3/2 & -3/2 \\ -1 & 6 & 2 & 2 \end{bmatrix}.$$

1.13. $BB^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. It means that $u_j \cdot u_k = 0$, for every $j \neq k$. Therefore matrix B is an orthogonal matrix.

Question 2.

A. Three linear model equations from the training data sets via leave-one-out cross validation(LOOCV) are listed below:

$\hat{y}_1 = 4 - \mathrm{x}$, (leave (0, 2) out)

$\hat{y}_2 = 2 - \mathrm{x}/3\,,$ (leave (2, 2) out)

$\hat{y}_3 = 2\,,$ (leave (3, 1) out)

The mean squared error of the LOOCV is

$$CV_{(3)} = \frac{1}{3}\sum_{i=1}^{3} MSE_i = \frac{1}{3}\sum_{i=1}^{3}\left(y_i - \hat{y}_i(x_i)\right)^2 = \frac{1}{3}\left(4 + \frac{4}{9} + 1\right) = \frac{49}{27}\,.$$

B. If using prediction model y = c , the three equations fit by the training sets via LOOCV are listed below:

$\hat{y}_1 = 3/2\,,$ (leave (0, 2) out)

$\hat{y}_2 = 3/2\,,$ (leave (2, 2) out)

$\hat{y}_3 = 2\,,$ (leave (3, 1) out)

The mean squared error of the LOOCV is

$$CV_{(3)} = \frac{1}{3}\left(\frac{1}{4} + \frac{1}{4} + 1\right) = \frac{1}{2}\,.$$

Question 3.

3.1. The screenshot pictures are listed in their question number below.

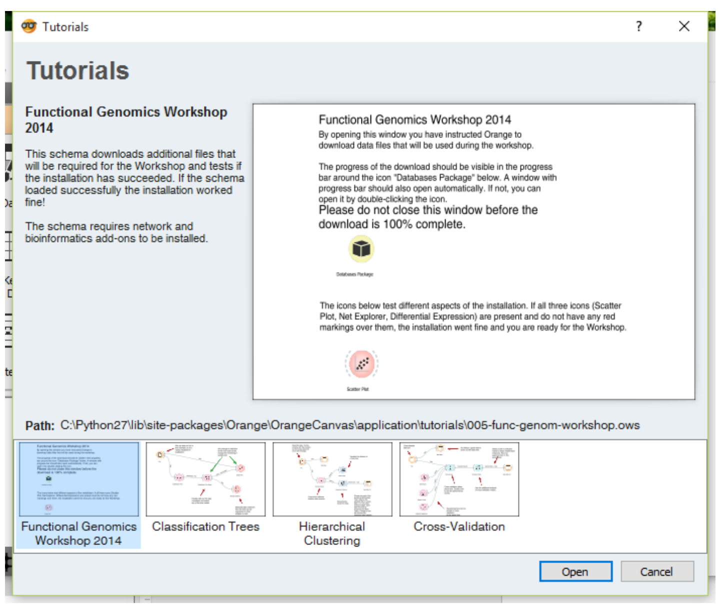


Fig. 3.1.1

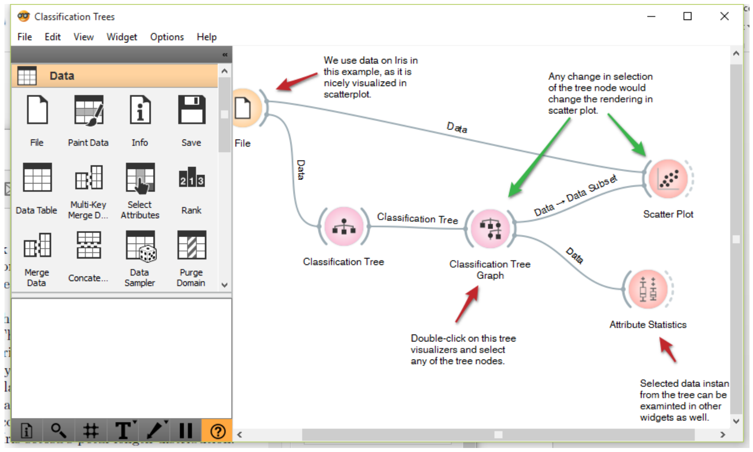


Fig. 3.1.2

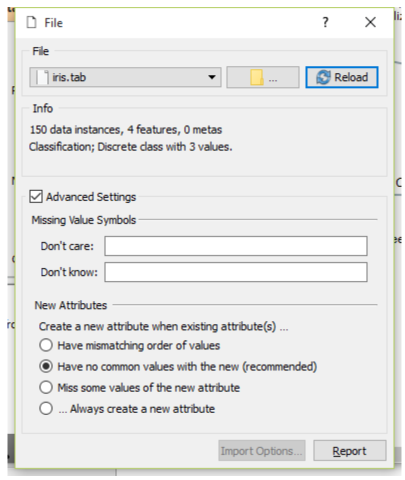


Fig. 3.1.3

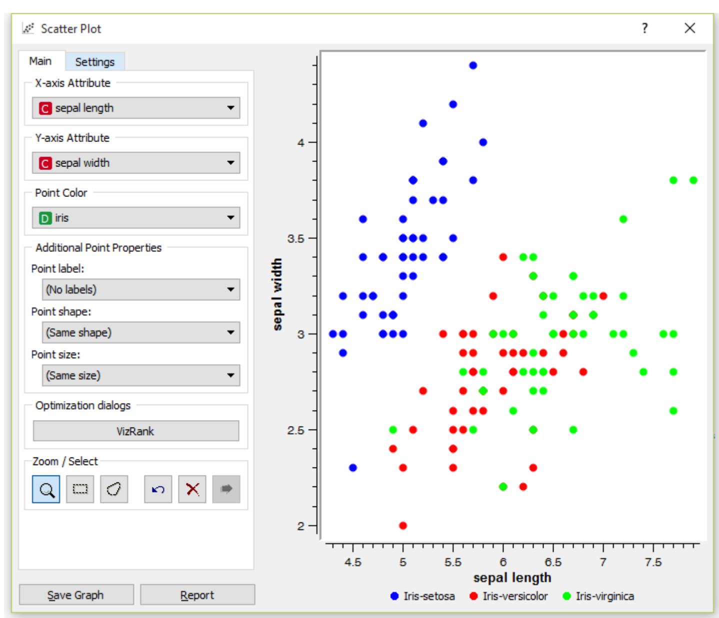


Fig. 3.1.4

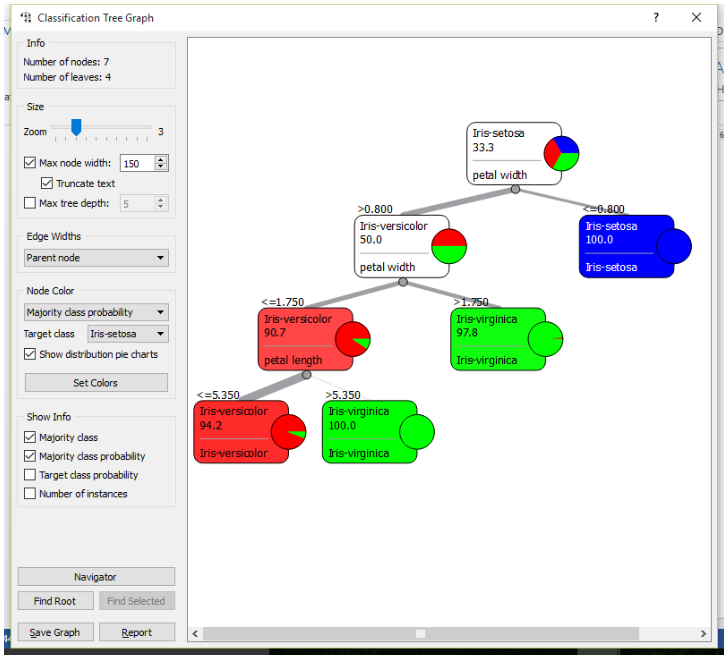


Fig. 3.1.5

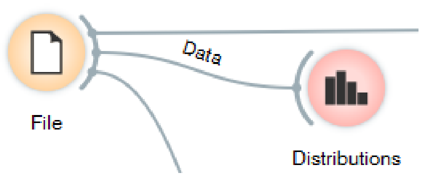


Fig. 3.1.6

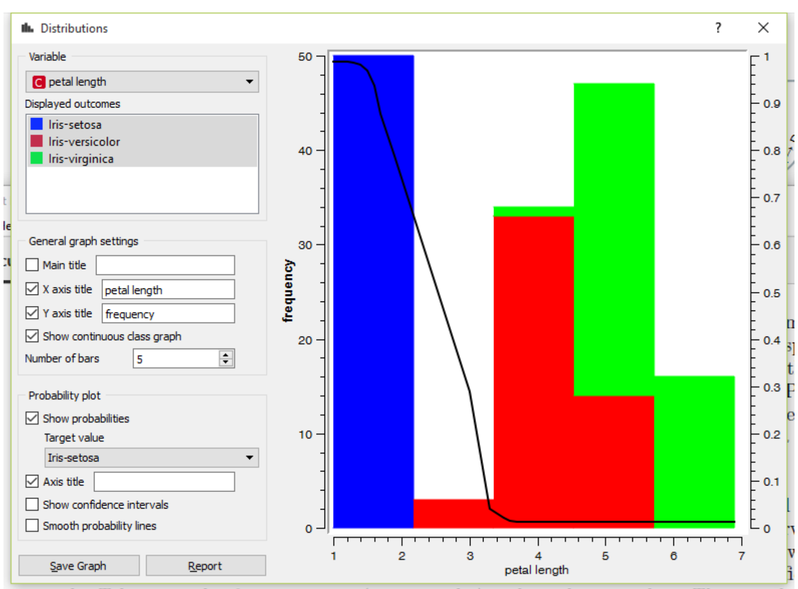


Fig. 3.1.7

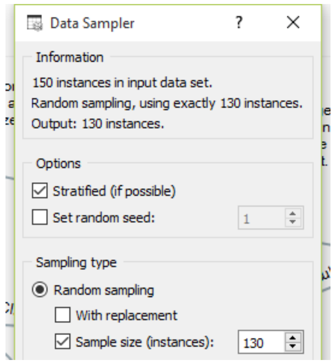


Fig. 3.1.8

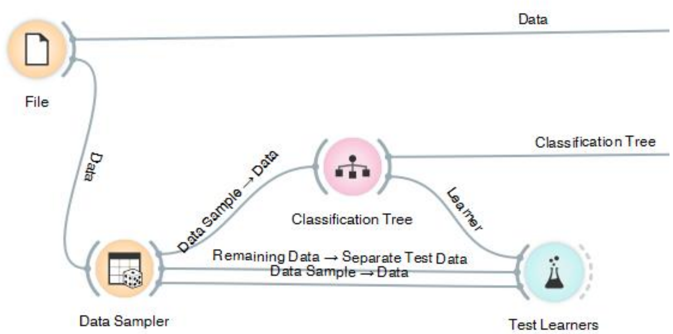
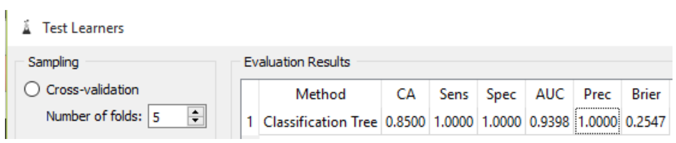


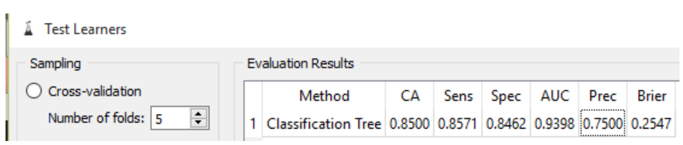
Fig. 3.1.9

Iris-setosa



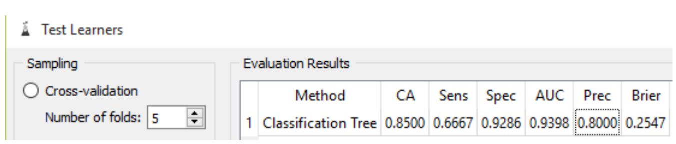
	Method	CA	Sens	Spec	AUC	Prec	Brier
1	Classification Tree	0.8500	1.0000	1.0000	0.9398	1.0000	0.2547

Iris-versicolar



	Method	CA	Sens	Spec	AUC	Prec	Brier
1	Classification Tree	0.8500	0.8571	0.8462	0.9398	0.7500	0.2547

Iris-virginica



	Method	CA	Sens	Spec	AUC	Prec	Brier
1	Classification Tree	0.8500	0.6667	0.9286	0.9398	0.8000	0.2547

Fig. 3.1.10

3.2. Firstly, the whole data set (no attributes reduce) are analyzed by three types of regression model (Linear regression, SVM regression and Random forest regression) via **5-folders Cross-Validation**. The process graph is shown in Fig. 3.2.1 and the results are list in Fig. 3.2.2. The best performing method is the Linear Regression.

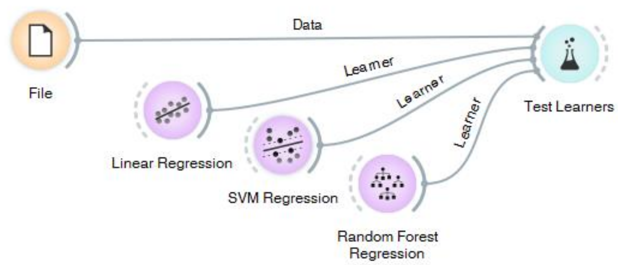


Fig. 3.2.1

Evaluation Results

	Method	MSE	MAE	RSE	RAE	R2
1	Linear Regression	0.1883	0.3741	0.7532	0.7481	0.2468
2	SVM Regression	0.1909	0.3829	0.7635	0.7658	0.2365
3	Random Forest	0.2101	0.4219	0.8403	0.8438	0.1597

Fig. 3.2.2

Secondly, the "Rank" node is used to select 10 attributes that are mostly related with the class to reduce the number of attributes (process graph is shown in Fig. 3.2.3). If using "ReliefF" to sort the rank, the list of selected attributes are shown in Fig. 3.2.4 and the results are shown in Fig. 3.2.5. By comparing the MSE, the reduced dataset performs worse than the whole dataset. The Linear Regression also performs best among the three methods.

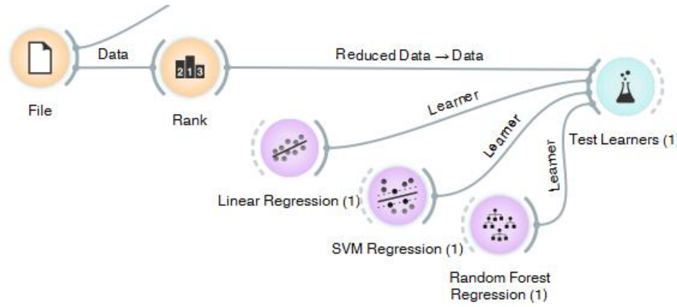
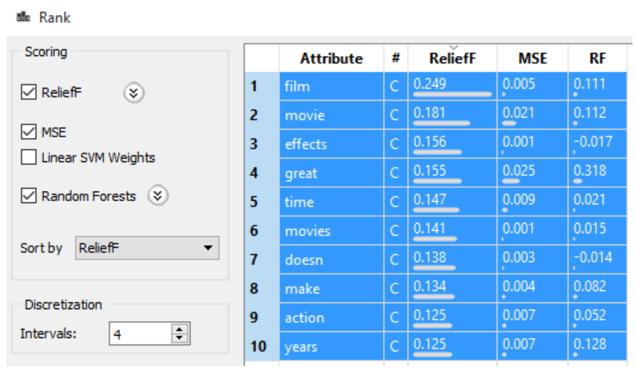


Fig. 3.2.3



	Attribute	#	ReliefF	MSE	RF
1	film	C	0.249	0.005	0.111
2	movie	C	0.181	0.021	0.112
3	effects	C	0.156	0.001	-0.017
4	great	C	0.155	0.025	0.318
5	time	C	0.147	0.009	0.021
6	movies	C	0.141	0.001	0.015
7	doesn	C	0.138	0.003	-0.014
8	make	C	0.134	0.004	0.082
9	action	C	0.125	0.007	0.052
10	years	C	0.125	0.007	0.128

Fig. 3.2.4

	Method	MSE	MAE	RSE	RAE	R2
1	Linear Regression	0.2383	0.4737	0.9531	0.9474	0.0469
2	SVM Regression	0.2664	0.4545	1.0656	0.9090	-0.0656
3	Random Forest	0.2471	0.4756	0.9884	0.9513	0.0116

Fig. 3.2.5

However, if using “MSE” to sort the rank, the 10 selected attributes and the results are shown in Fig. 3.2.6 and 3.2.7, which are very close to the results of the whole dataset. It means that using “MSE” can distinguish the main attributes that have high correlation with the class.

Rank

Scoring

☑ ReliefF

☑ MSE

☐ Linear SVM Weights

☑ Random Forests

Sort by MSE

Discretization

Intervals: 4

	Attribute	#	ReliefF	MSE	RF
1	bad	C	0.068	0.088	1.983
2	worst	C	0.025	0.056	0.909
3	life	C	0.077	0.032	0.634
4	supposed	C	0.041	0.032	0.481
5	plot	C	0.086	0.029	0.463
6	great	C	0.155	0.025	0.318
7	TRUE	C	0.057	0.023	0.149
8	script	C	0.040	0.021	0.302
9	movie	C	0.181	0.021	0.112
10	performances	C	0.040	0.021	0.275

Fig. 3.2.6

	Method	MSE	MAE	RSE	RAE	R2
1	Linear Regression	0.1964	0.4035	0.7854	0.8069	0.2146
2	SVM Regression	0.2080	0.3873	0.8321	0.7746	0.1679
3	Random Forest	0.2052	0.4159	0.8210	0.8317	0.1790

Fig. 3.2.7

Question 4.

The distribution of the original data from “Q4data.txt” is shown in Fig. 4.1. Here, the method of least squares via the normal equation is utilized to optimize the linear regression. The calculus equation is shown below:

$$\theta^* = (X^T X)^{-1} X^T y \,.$$

The equation of the linear regression is

$$\hat{y} = \theta_0^* + \theta_1^* x$$

Where $\theta_0^* = \mathbf{3.0077}$ *and* $\theta_1^* = \mathbf{1.6953}$.

It is plotted as blue solid line with the original data distribution together in Fig. 4.2 below.

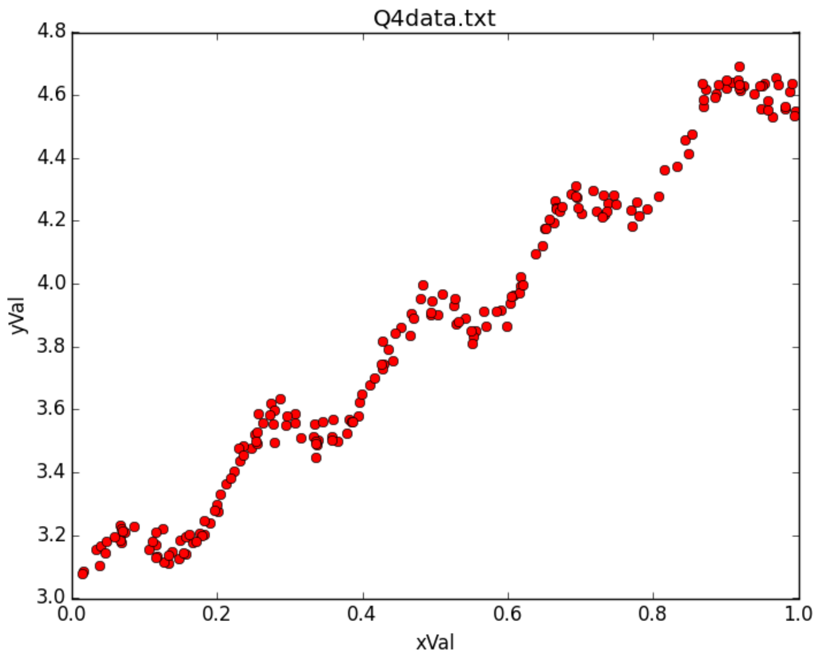

Q4data.txt
4.8
4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
yVal
0.0
0.2
0.4
0.6
0.8
1.0
xVal

Fig. 4.1

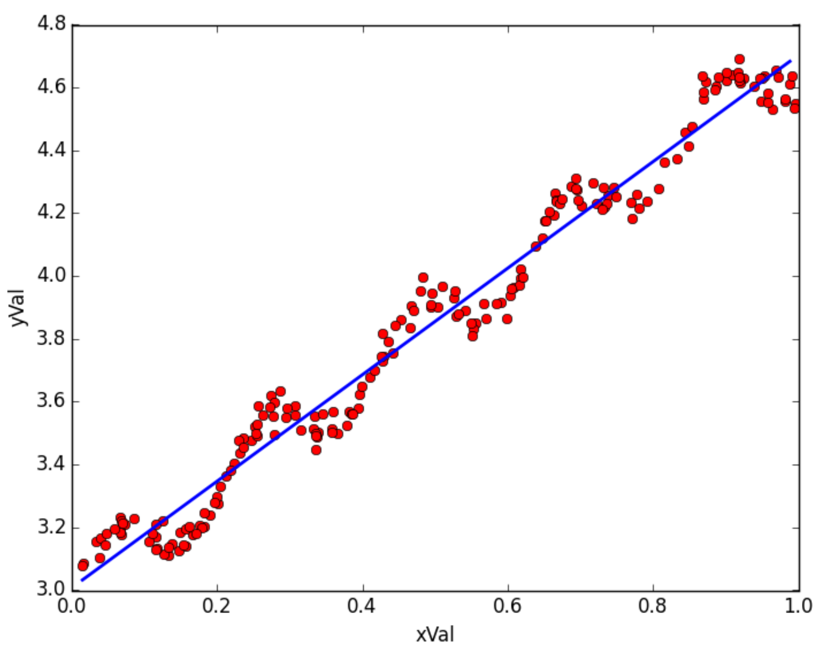

4.8
4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
yVal
0.0
0.2
0.4
0.6
0.8
1.0
xVal

Fig. 4.2