1. (i) $\min_{w_1, w_0 \in R} \frac{1}{N} \sum_{t=1}^{N} (r^t - (w_1 x^t + w_0))^2$

error $= e$ $= \sum_{t=1}^{N} \| r^t - (w_1 x^t + w_0) \|^2$

$\frac{\partial e}{\partial w_1} = 0 \Leftrightarrow \sum_{t=1}^{N} 2(r^t - (w_1 x^t + w_0)) x^t = 0 \Leftrightarrow \left(\sum_{t=1}^{N} r^t x^t - (w_1 \sum_{t=1}^{N} (x^t)^2 + w_0 \sum_{t=1}^{N} x^t) \right) = 0$

$\Leftrightarrow w_1 \sum_{t=1}^{N} (x^t)^2 + w_0 \sum_{t=1}^{N} x^t = \sum_{t=1}^{N} r^t x^t$

$\frac{\partial e}{w_0} = 0 \Leftrightarrow \sum_{t=1}^{N} 2(r^t - (w_1 x^t + w_0)) = 0 \Leftrightarrow \sum_{t=1}^{N} r^t - (w_1 \sum_{t=1}^{N} x^t + N w_0) = 0$

$\Leftrightarrow w_1 \sum_{t=1}^{N} x^t + N w_0 = \sum_{t=1}^{N} r^t$

$$\begin{pmatrix} \sum_{t=1}^{N} (x^t)^2 & \sum_{t=1}^{N} x^t \\ \sum_{t=1}^{N} x^t & N \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_0 \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^{N} r^t x^t \\ \sum_{t=1}^{N} r^t \end{pmatrix}$$

Let $A = \begin{pmatrix} \sum_{t=1}^{N} (x^t)^2 & \sum_{t=1}^{N} x^t \\ \sum_{t=1}^{N} x^t & N \end{pmatrix}$ $\det A = \sum_{t=1}^{N} (x^t)^2 \cdot (N-1)$ if $N \neq 1$, $\begin{pmatrix} w_1 \\ w_0 \end{pmatrix} = A^{-1} \begin{pmatrix} \sum_{t=1}^{N} r^t x^t \\ \sum_{t=1}^{N} r^t \end{pmatrix}$

if $N=1$ $\det A = 0$ $A \xrightarrow{\text{elimination}} \begin{pmatrix} \sum_{t=1}^{N} (x^t)^2 & \sum_{t=1}^{N} x^t \\ 0 & 0 \end{pmatrix}$

Thus $0 = r$, $\sum_{t=1}^{N} (x^t)^2 \cdot w_1 + \sum_{t=1}^{N} x^t w_0 = 0 \Leftrightarrow x \cdot w_1 + w_0 = 0$

$\begin{pmatrix} w_1 \\ w_0 \end{pmatrix} = \begin{pmatrix} 1 \\ -x \end{pmatrix}$ where $x \in R$

(ii) error $= e = \sum_{t=1}^{N} (r^t - (v_2 (x^t)^2 + v_1 x^t + v_0))^2$

$\frac{\partial e}{\partial v_2} = 0 \Leftrightarrow \sum_{t=1}^{N} 2(r^t - (v_2 (x^t)^2 + v_1 x^t + v_0)) \cdot (x^t)^2 = 0$

$\Leftrightarrow v_2 \sum_{t=1}^{N} (x^t)^4 + v_1 \sum_{t=1}^{N} (x^t)^3 + v_0 \sum_{t=1}^{N} (x^t)^2 = \sum_{t=1}^{N} r^t (x^t)^2$

$\frac{\partial e}{\partial v_1} = 0 \Leftrightarrow \sum_{t=1}^{N} 2(r^t - (v_2 (x^t)^2 + v_1 x^t + v_0)) x^t = 0$

$\Leftrightarrow v_2 \sum_{t=1}^{N} (x^t)^3 + v_1 \sum_{t=1}^{N} (x^t)^2 + v_0 \sum_{t=1}^{N} x^t = \sum_{t=1}^{N} r^t x^t$

$$\frac{\partial e}{v_0}=0 \Leftrightarrow \sum_{t=1}^{N} 2\cdot(r^t-(v_2(x^t)^2+v_1x^t+v_0))=0$$

$$\Leftrightarrow v_2\sum_{t=1}^{N}(x^t)^2+v_1\sum_{t=1}^{N}x^t+N\,v_0=\sum_{t=1}^{N}r^t$$

In summary,

$$\text{let } A=\begin{pmatrix}\sum_{t=1}^{N}(x^t)^4 & \sum_{t=1}^{N}(x^t)^3 & \sum_{t=1}^{N}(x^t)^2\\ \sum_{t=1}^{N}(x^t)^3 & \sum_{t=1}^{N}(x^t)^2 & \sum_{t=1}^{N}x^t\\ \sum_{t=1}^{N}(x^t)^2 & \sum_{t=1}^{N}x^t & N\end{pmatrix},\quad A\cdot\begin{pmatrix}v_2\\ v_1\\ v_0\end{pmatrix}=\begin{pmatrix}\sum_{t=1}^{N}r^t(x^t)^2\\ \sum_{t=1}^{N}r^tx^t\\ \sum_{t=1}^{N}r^t\end{pmatrix}$$

$$A\xrightarrow{\text{elimination}}\begin{pmatrix}\sum_{t=1}^{N}(x^t)^4 & \sum_{t=1}^{N}(x^t)^3 & \sum_{t=1}^{N}(x^t)^2\\ 0 & 0 & 0\\ 0 & 0 & \sum_{t=1}^{N}x^t\cdot(N-1)\end{pmatrix}\qquad \det A=0$$

$$\Leftrightarrow \begin{aligned} v_2\sum_{t=1}^{N}(x^t)^4+v_1\sum_{t=1}^{N}(x^t)^3+v_0\sum_{t=1}^{N}(x^t)^2 &= \sum_{t=1}^{N}r^t(x^t)^2\\ 0 &= \sum_{t=1}^{N}r^tx^t\\ v_0\sum_{t=1}^{N}x^t(N-1) &= \sum_{t=1}^{N}r^t\end{aligned}$$

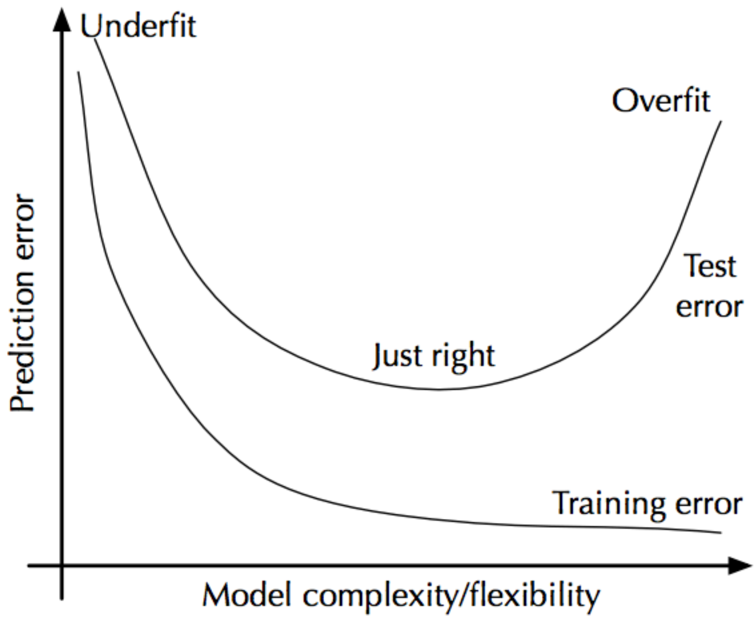
$$\Leftrightarrow v_2\sum_{t=1}^{N}(x^t)^4+v_1\sum_{t=1}^{N}(x^t)^3=0$$

$$\Leftrightarrow \begin{pmatrix}v_2\\ v_1\end{pmatrix}=\begin{pmatrix}1\\ -\sum_{t=1}^{N}(x^t)\end{pmatrix}$$

(iii) Yes. As the order of the polynomial model increases, the error in the training set decreases, since it better fits the training data

(iv) No. It's hard to tell which model has lower error since on the one hand the higher order model could better fit the model but it may overfit for some test sets.
So the overall performance is usually nondeterministic

The relationship between model complexity and prediction error can be described in the graph below:



http://www.stats.ox.ac.uk/~sejdinov/teaching/sdmml15/materials/HT15_lecture12-nup.pdf

2. (i) $tr(A) = 1+2+9+64+625 = 701$

$tr(A^T) = tr(A) = 701$

$tr(A^TA) = \sum_{i=0}^{4}\sum_{j=0}^{4} a_{ij}^2 = (\sum_{i=0}^{4} 1^i)^2 + (\sum_{i=0}^{4} 2^i)^2 + (\sum_{i=0}^{4} 3^i)^2 + (\sum_{i=0}^{4} 4^i)^2 + (\sum_{i=0}^{4} 5^i)^2 = 484533$

$tr(AA^T) = tr(A^TA) = 484533$

(ii)
detA=product of all of A's eigenvalues.
A can be projected in a way that only its eigenvectors stretch without rotating. And each eigenvector is stretched to an extent that can be measured by multiplying its eigenvalue and all the other vectors can be expressed as a linear combination of all the eigenvectors and thus the volume is only changed to a new one which is the certain amount times the original volume. And the change rate of volume can be calculated by multiplying all the eigenvalues as they measure the change rate of each eigenvector and all the other vector can be expressed as a linear combination of these eigenvectors.

(iii)

$$A \xrightarrow{\text{elimination}} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 7 & 15 \\ 0 & 1 & 5 & 19 & 61 \\ 0 & 1 & 7 & 37 & 175 \\ 0 & 1 & 9 & 61 & 369 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 7 & 15 \\ 1 & 5 & 19 & 65 \\ 1 & 7 & 37 & 175 \\ 1 & 9 & 61 & 369 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 7 & 15 \\ 0 & 3 & 12 & 50 \\ 0 & 2 & 18 & 110 \\ 0 & 2 & 24 & 194 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 12 & 50 \\ 2 & 18 & 110 \\ 2 & 24 & 194 \end{pmatrix} \rightarrow \begin{pmatrix} 3 & 12 & 50 \\ 0 & -6 & -84 \\ 2 & 24 & 194 \end{pmatrix}$$

after the elimination, we get a 3x3 matrix with the same determinant as $|A|$, and clearly we can see the three column vectors are linearly independent so $|A|$ is not zero.

(iv) Yes. $rank(A)=5$ since $|A| \neq 0$ in (iii), all of the five row vectors are linearly independent

Q3

(i)

LinearSVC with Boston50

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.24	0.2	0.14	0.7	0.14	0.08	0.34	0.3	0.1	0.58	0.282	0.197474048928

LinearSVC with Boston75

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.18	0.52	0.04	0.42	0.36	0.32	0.64	0.16	0.04	0.04	0.272	0.202820117345

LinearSVC with Digits

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.0949720670391	0.072625698324	0.156424581006	0.100558659218	0.072625698324	0.0502793296089	0.0335195530726	0.072625698324	0.122905027933	0.0111731843575	0.0787709497207	0.0403203364822

SVC with Boston50

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.66	0.28	0.84	0.84	0.88	0.9	0.46	0.46	0.8	0.0	0.612	0.287638662214

SVC with Boston75

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.18	0.26	0.04	0.6	0.46	0.66	0.2	0.16	0.04	0.02	0.262	0.221350400948

SVC with Digits

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.603351955307	0.491620111732	0.625698324022	0.581005586592	0.530726256983	0.586592178771	0.603351955307	0.519553072626	0.446927374302	0.00558659217877	0.499441340782	0.17316993993

LogisticRegression with Boston50

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.12	0.16	0.12	0.12	0.18	0.06	0.28	0.28	0.04	0.16	0.152	0.076

LogisticRegression with Boston75

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.08	0.08	0.04	0.16	0.14	0.12	0.12	0.14	0.04	0.04	0.096	0.0436348484585

LogisticRegression with Digits

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.0949720670391	0.0502793296089	0.106145251397	0.0949720670391	0.0558659217877	0.0335195530726	0.0167597765363	0.0502793296089	0.128491620112	0.0111731843575	0.0642458100559	0.0377456892326

(ii)
LinearSVC with Boston50

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.528	0.504	0.504	0.496	0.536	0.496	0.528	0.528	0.552	0.496	0.5168	0.0189989473393

LinearSVC with Boston75

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.296	0.256	0.264	0.272	0.712	0.256	0.648	0.232	0.264	0.464	0.3664	0.168979998816

LinearSVC with Digits

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.926339285714	0.924107142857	0.899553571429	0.901785714286	0.901785714286	0.883928571429	0.917410714286	0.890625	0.901785714286	0.904017857143	0.905133928571	0.0129867751855

SVC with Boston50

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.536	0.552	0.512	0.6	0.528	0.44	0.568	0.536	0.48	0.432	0.5184	0.0510748470384

SVC with Boston75

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.28	0.304	0.2	0.224	0.272	0.288	0.248	0.232	0.296	0.288	0.2632	0.0332649966181

SVC with Digits

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.928571428571	0.901785714286	0.924107142857	0.908482142857	0.90625	0.90625	0.90625	0.912946428571	0.919642857143	0.915178571429	0.912946428571	0.00835191380976

LogisticRegression with Boston50

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.488	0.432	0.48	0.52	0.512	0.488	0.44	0.464	0.504	0.44	0.4768	0.0299759903923

LogisticRegression with Boston75

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.248	0.256	0.248	0.272	0.28	0.28	0.288	0.192	0.208	0.296	0.2568	0.0324863048068

LogisticRegression with Digits

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.912946428571	0.881696428571	0.917410714286	0.895089285714	0.892857142857	0.919642857143	0.908482142857	0.926339285714	0.881696428571	0.895089285714	0.903125	0.0151457102265

Q4

LinearSVC with ~X1

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.156424581006	0.0335195530726	0.156424581006	0.122905027933	0.134078212291	0.0893854748603	0.0391061452514	0.0558659217877	0.134078212291	0.111731843575	0.103351955307	0.0440774688214

LinearSVC with ~X2

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.0502793296089	0.0	0.0670391061453	0.0335195530726	0.0223463687151	0.0223463687151	0.00558659217877	0.0111731843575	0.0391061452514	0.0	0.0251396648045	0.0212363997863

LinearSVC with ~X3

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.111731843575	0.0391061452514	0.0670391061453	0.100558659218	0.0614525139665	0.0670391061453	0.0446927374302	0.0558659217877	0.100558659218	0.0167597765363	0.0664804469274	0.0286445670241

SVC with ~X1

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.905027932961	0.916201117318	0.905027932961	0.899441340782	0.899441340782	0.91061452514	0.91061452514	0.905027932961	0.905027932961	0.00558659217877	0.816201117318	0.270248473415

SVC with ~X2

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.905027932961	0.916201117318	0.905027932961	0.748603351955	0.899441340782	0.91061452514	0.91061452514	0.905027932961	0.905027932961	0.00558659217877	0.801117318436	0.269389622485

SVC with ~X3

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.905027932961	0.916201117318	0.905027932961	0.899441340782	0.899441340782	0.91061452514	0.91061452514	0.905027932961	0.905027932961	0.00558659217877	0.816201117318	0.270248473415

LogisticRegression with ~X1

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.128491620112	0.0335195530726	0.145251396648	0.0782122905028	0.111731843575	0.128491620112	0.0558659217877	0.0782122905028	0.117318435754	0.0391061452514	0.0916201117318	0.0379888268156

LogisticRegression with ~X2

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.0502793296089	0.0	0.0558659217877	0.0335195530726	0.0279329608939	0.0167597765363	0.00558659217877	0.0111731843575	0.0391061452514	0.0	0.0240223463687	0.0193605848616

LogisticRegression with ~X3

K=0	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	Mean	Std
0.128491620112	0.0391061452514	0.106145251397	0.139664804469	0.072625698324	0.111731843575	0.0837988826816	0.0837988826816	0.106145251397	0.0111731843575	0.0882681564246	0.0375425645059