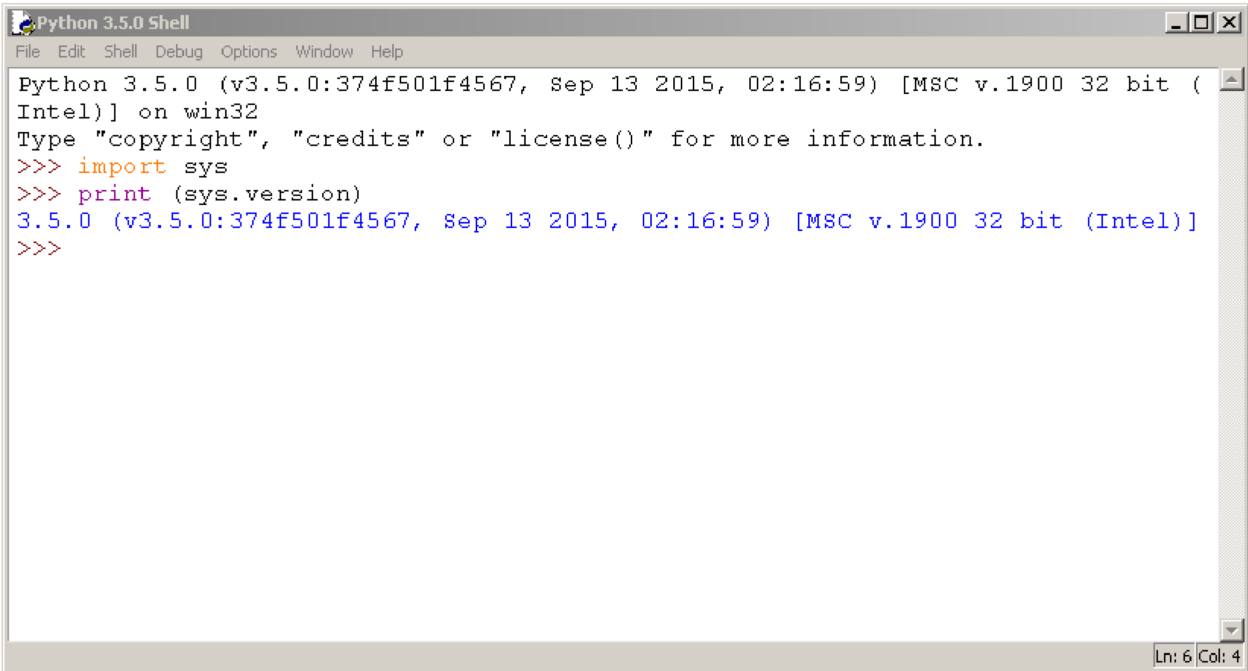
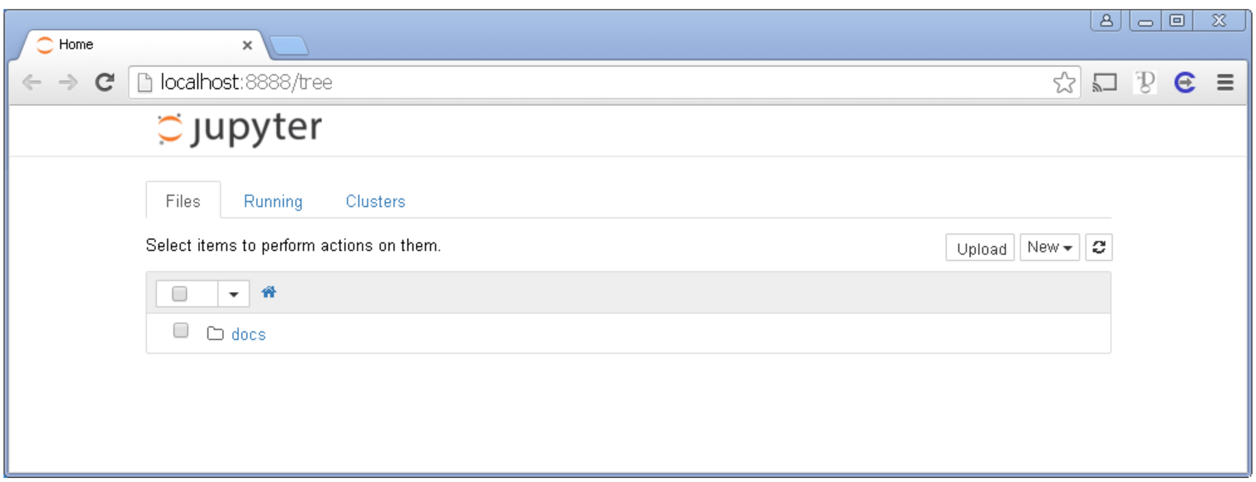
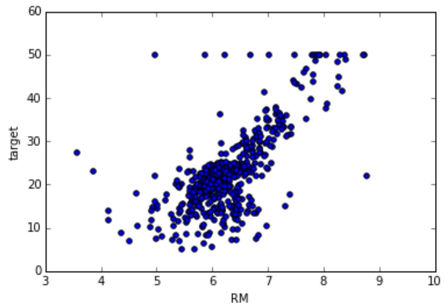
Chapter 1: Regression – The Workhorse of Data Science





In [1]:
```
import pandas as pd
from sklearn.datasets import load_boston
boston = load_boston()
dataset = pd.DataFrame(boston.data, columns=boston.feature_names)
dataset['target'] = boston.target
```

In [2]:
```
%matplotlib inline
# If you are using IPython, this will make the images available in the notebook
scatter = dataset.plot(kind='scatter', x='RM', y='target')
```



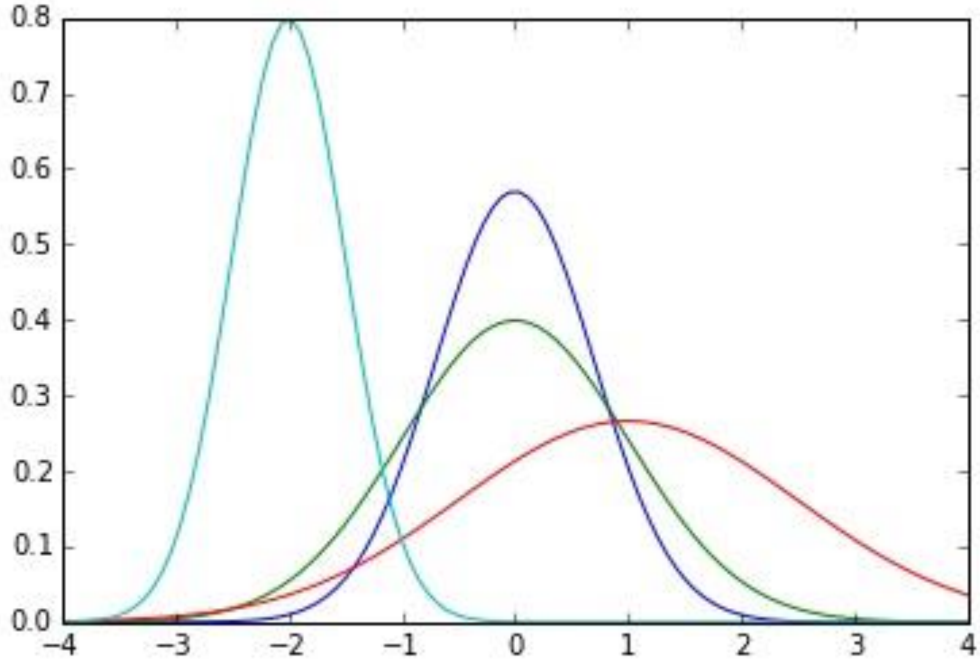
Chapter 2: Approaching Simple Linear Regression

$$y = h(X)$$

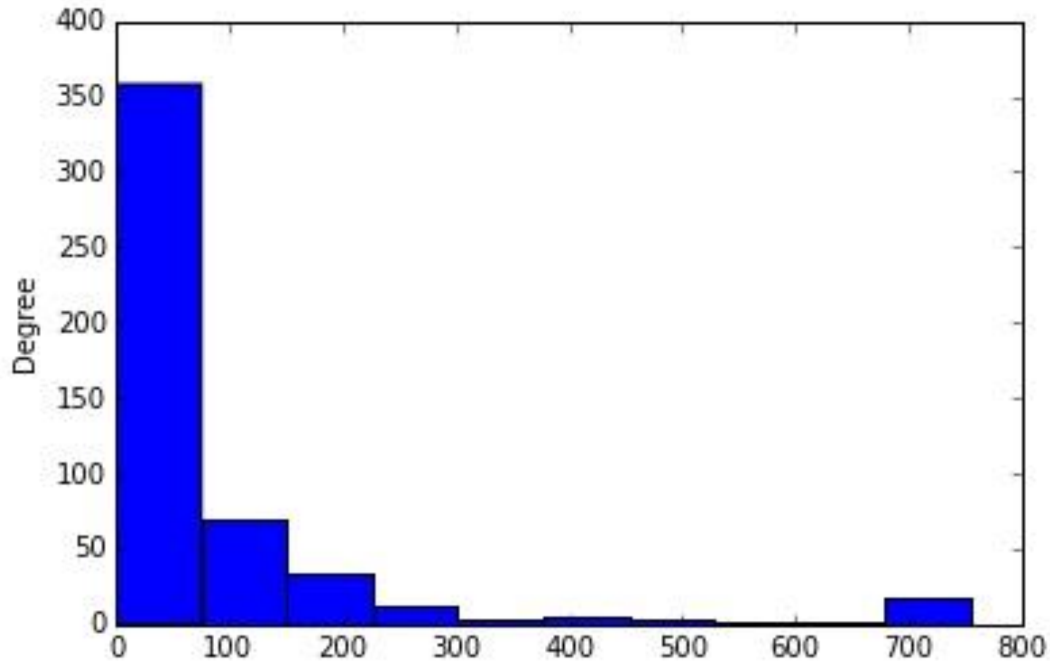
$$X = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}$$

$$X = \begin{bmatrix} x_{1,1}\, x_{1,2} - x_{1,p} \\ x_{2,1}\, x_{2,2} - x_{2,p} \\ . \\ . \\ . \\ x_{n,1}\, x_{n,2} - x_{n,p} \end{bmatrix}$$

$$f(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−4
−3
−2
−1
0
1
2
3
4

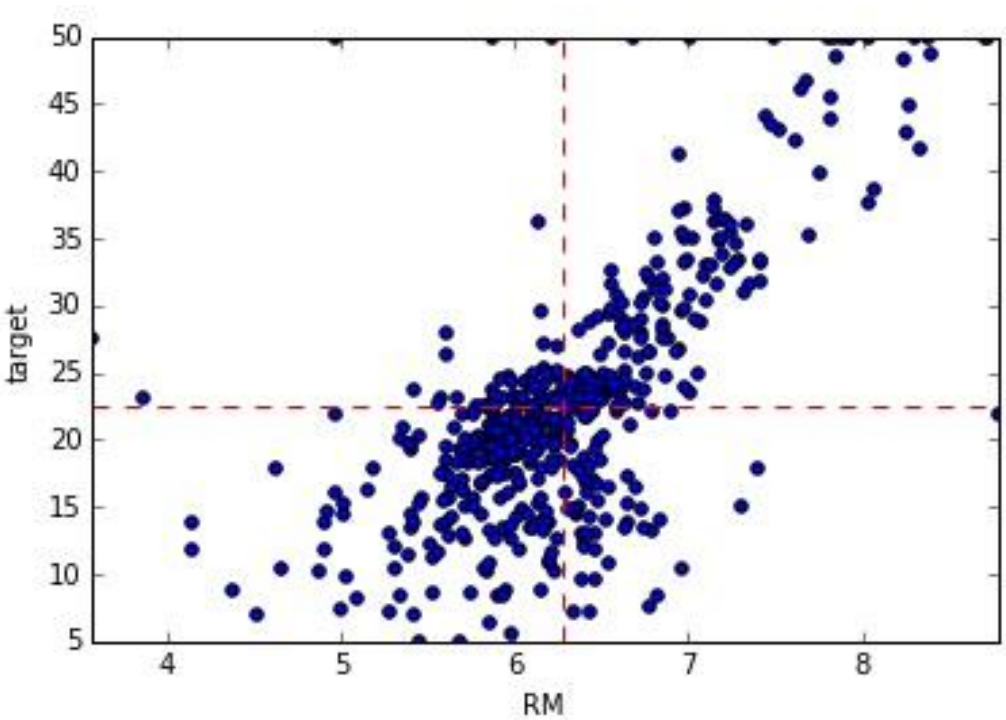
$$y = \bar{x}$$


Degree
400
350
300
250
200
150
100
50
0
0
100
200
300
400
500
600
700
800

$$x = \frac{x - \bar{x}}{\sigma x}$$

$$cov(x_i, y) = \frac{1}{n} * \sum (x_i - \bar{x}_i) * (y - \bar{y})$$

$$r = \frac{1}{n} * \frac{\sum (x_i - \bar{x}_i) * (y - \bar{y})}{\sigma_{x_i} * \sigma_y}$$



$$y = \beta X + \beta_0$$

$$y = \beta X + \beta_0$$

$$y = \beta X$$

	const	RM
0	1	6.575
1	1	6.421
2	1	7.185
3	1	6.998
4	1	7.147

OLS Regression Results

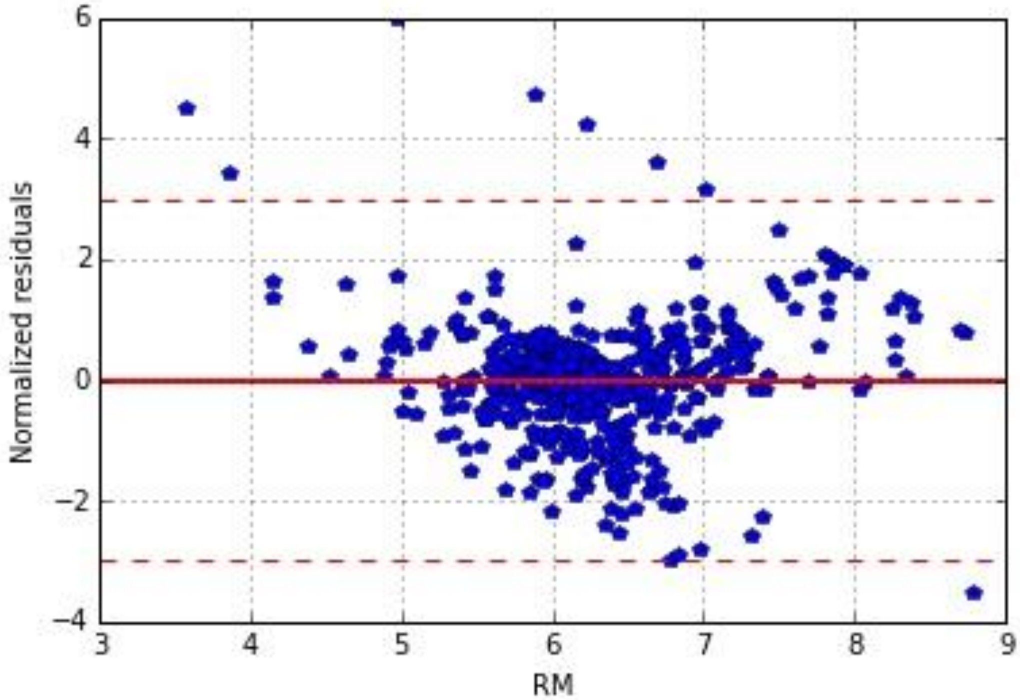
Dep. Variable:	target	**R-squared:**	0.484
Model:	OLS	**Adj. R-squared:**	0.483
Method:	Least Squares	**F-statistic:**	471.8
Date:	Sat, 28 Nov 2015	**Prob (F-statistic):**	2.49e-74
Time:	21:02:32	**Log-Likelihood:**	-1673.1
No. Observations:	506	**AIC:**	3350.
Df Residuals:	504	**BIC:**	3359.
Df Model:	1		
Covariance Type:	nonrobust		

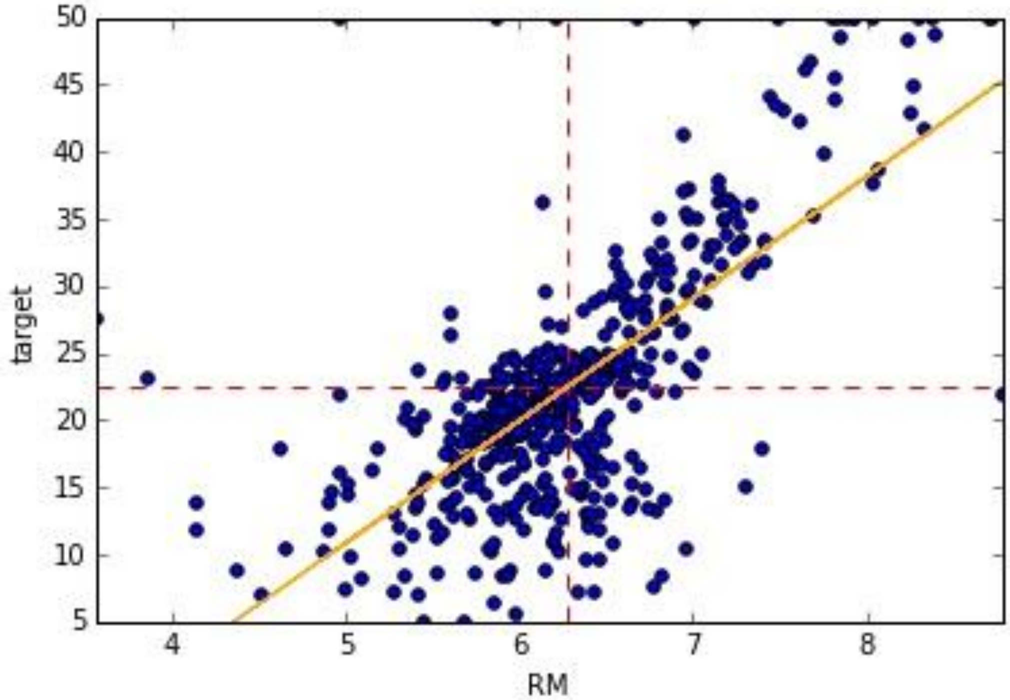
	coef	std err	t	P>|t|	[95.0% Conf. Int.]
const	-34.6706	2.650	-13.084	0.000	-39.877 -29.465
RM	9.1021	0.419	21.722	0.000	8.279 9.925

Omnibus:	102.585	**Durbin-Watson:**	0.684
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	612.449
Skew:	0.726	**Prob(JB):**	1.02e-133
Kurtosis:	8.190	**Cond. No.**	58.4

$$y = \beta X + \beta_0$$

$$y = 9.1021 * x_{RM} - 34.6706$$





$$y \approx h(X) = \beta X + \beta_0$$

$$\frac{1}{2n} * \sum (h(X) - y)^2$$

$$w = (X^T X)^{-1} X^T y$$

$$(X^T X)^{-1} * w = X^T y$$

$$J(w) = \frac{1}{2n} \sum (Xw - y)^2$$

$$w_j = w_j - \alpha * \frac{\partial}{\partial w} J(w)$$

$$w_j = w_j - \alpha * \frac{1}{n} \sum (Xw - y) * x_j$$

Chapter 3: Multiple Regression in Action

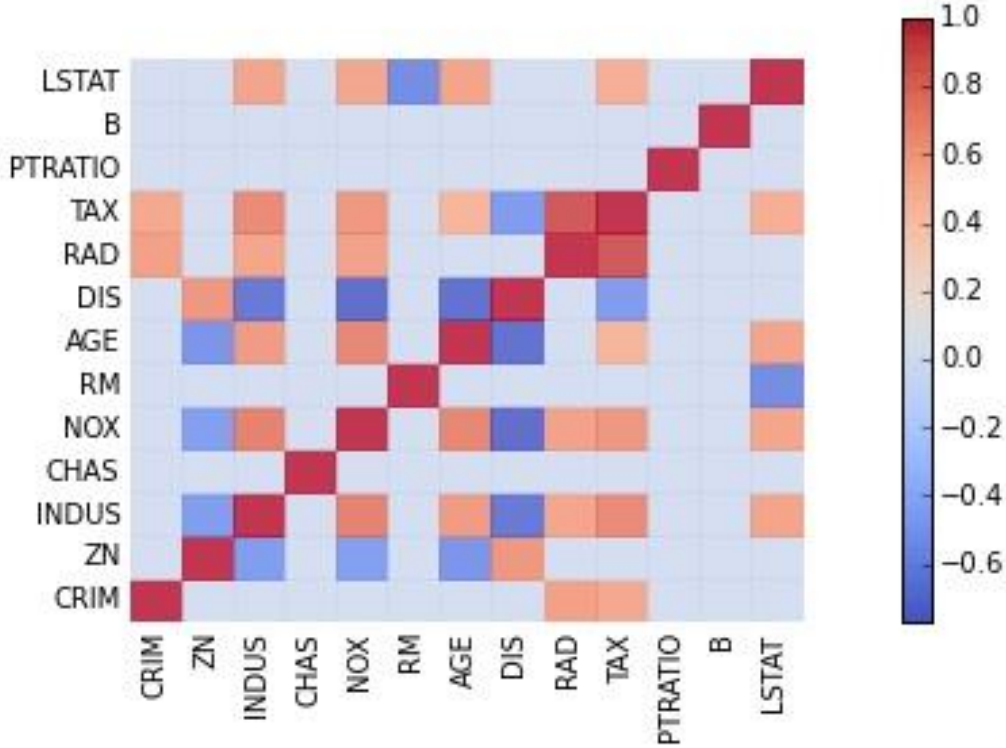
OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.741
Model:	OLS	**Adj. R-squared:**	0.734
Method:	Least Squares	**F-statistic:**	108.1
Date:	Tue, 29 Sep 2015	**Prob (F-statistic):**	6.95e-135
Time:	21:45:28	**Log-Likelihood:**	-1498.8
No. Observations:	506	**AIC:**	3026.
Df Residuals:	492	**BIC:**	3085.
Df Model:	13		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
const	36.4911	5.104	7.149	0.000	26.462 46.520
CRIM	-0.1072	0.033	-3.276	0.001	-0.171 -0.043
ZN	0.0464	0.014	3.380	0.001	0.019 0.073
INDUS	0.0209	0.061	0.339	0.735	-0.100 0.142
CHAS	2.6886	0.862	3.120	0.002	0.996 4.381
NOX	-17.7958	3.821	-4.658	0.000	-25.302 -10.289
RM	3.8048	0.418	9.102	0.000	2.983 4.626
AGE	0.0008	0.013	0.057	0.955	-0.025 0.027
DIS	-1.4758	0.199	-7.398	0.000	-1.868 -1.084
RAD	0.3057	0.066	4.608	0.000	0.175 0.436
TAX	-0.0123	0.004	-3.278	0.001	-0.020 -0.005
PTRATIO	-0.9535	0.131	-7.287	0.000	-1.211 -0.696
B	0.0094	0.003	3.500	0.001	0.004 0.015
LSTAT	-0.5255	0.051	-10.366	0.000	-0.625 -0.426

Omnibus:	178.029	**Durbin-Watson:**	1.078
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	782.015
Skew:	1.521	**Prob(JB):**	1.54e-170
Kurtosis:	8.276	**Cond. No.**	1.51e+04

```
             CRIM        ZN     INDUS      CHAS       NOX        RM       AGE  \
CRIM     1.000000 -0.199458  0.404471 -0.055295  0.417521 -0.219940  0.350784
ZN      -0.199458  1.000000 -0.533828 -0.042697 -0.516604  0.311991 -0.569537
INDUS    0.404471 -0.533828  1.000000  0.062938  0.763651 -0.391676  0.644779
CHAS    -0.055295 -0.042697  0.062938  1.000000  0.091203  0.091251  0.086518
NOX      0.417521 -0.516604  0.763651  0.091203  1.000000 -0.302188  0.731470
RM      -0.219940  0.311991 -0.391676  0.091251 -0.302188  1.000000 -0.240265
AGE      0.350784 -0.569537  0.644779  0.086518  0.731470 -0.240265  1.000000
DIS     -0.377904  0.664408 -0.708027 -0.099176 -0.769230  0.205246 -0.747881
RAD      0.622029 -0.311948  0.595129 -0.007368  0.611441 -0.209847  0.456022
TAX      0.579564 -0.314563  0.720760 -0.035587  0.668023 -0.292048  0.506456
PTRATIO  0.288250 -0.391679  0.383248 -0.121515  0.188933 -0.355501  0.261515
B       -0.377365  0.175520 -0.356977  0.048788 -0.380051  0.128069 -0.273534
LSTAT    0.452220 -0.412995  0.603800 -0.053929  0.590879 -0.613808  0.602339

              DIS       RAD       TAX   PTRATIO         B     LSTAT
CRIM    -0.377904  0.622029  0.579564  0.288250 -0.377365  0.452220
ZN       0.664408 -0.311948 -0.314563 -0.391679  0.175520 -0.412995
INDUS   -0.708027  0.595129  0.720760  0.383248 -0.356977  0.603800
CHAS    -0.099176 -0.007368 -0.035587 -0.121515  0.048788 -0.053929
NOX     -0.769230  0.611441  0.668023  0.188933 -0.380051  0.590879
RM       0.205246 -0.209847 -0.292048 -0.355501  0.128069 -0.613808
AGE     -0.747881  0.456022  0.506456  0.261515 -0.273534  0.602339
DIS      1.000000 -0.494588 -0.534432 -0.232471  0.291512 -0.496996
RAD     -0.494588  1.000000  0.910228  0.464741 -0.444413  0.488676
TAX     -0.534432  0.910228  1.000000  0.460853 -0.441808  0.543993
PTRATIO -0.232471  0.464741  0.460853  1.000000 -0.177383  0.374044
B        0.291512 -0.444413 -0.441808 -0.177383  1.000000 -0.366087
LSTAT   -0.496996  0.488676  0.543993  0.374044 -0.366087  1.000000
```



$$y = \beta_0 + \sum \beta_i x_i$$

$$y = \left(\hat{\beta}_0 - \sum \frac{\hat{\beta}_i * \overline{x_i}}{\delta_i}\right) + \sum \left(\frac{\hat{\beta}_i}{\delta_i} * x_i\right)$$

```
   bias:  36.4911
   CRIM:  -0.1072
     ZN:   0.0464
  INDUS:   0.0209
   CHAS:   2.6886
    NOX: -17.7958
     RM:   3.8048
    AGE:   0.0008
    DIS:  -1.4758
    RAD:   0.3057
    TAX:  -0.0123
PTRATIO:  -0.9535
      B:   0.0094
  LSTAT:  -0.5255
```

```
17.796 NOX
 3.805 RM
 2.689 CHAS
 1.476 DIS
 0.953 PTRATIO
 0.525 LSTAT
 0.306 RAD
 0.107 CRIM
 0.046 ZN
 0.021 INDUS
 0.012 TAX
 0.009 B
 0.001 AGE
```

```
3.749 LSTAT
3.104 DIS
2.671 RM
2.659 RAD
2.076 TAX
2.062 PTRATIO
2.060 NOX
1.081 ZN
0.920 CRIM
0.857 B
0.682 CHAS
0.143 INDUS
0.021 AGE
```

```
0.057 LSTAT
0.044 RM
0.029 DIS
0.028 PTRATIO
0.011 NOX
0.011 RAD
0.006 B
0.006 ZN
0.006 TAX
0.006 CRIM
0.005 CHAS
0.000 INDUS
0.000 AGE
```

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$$

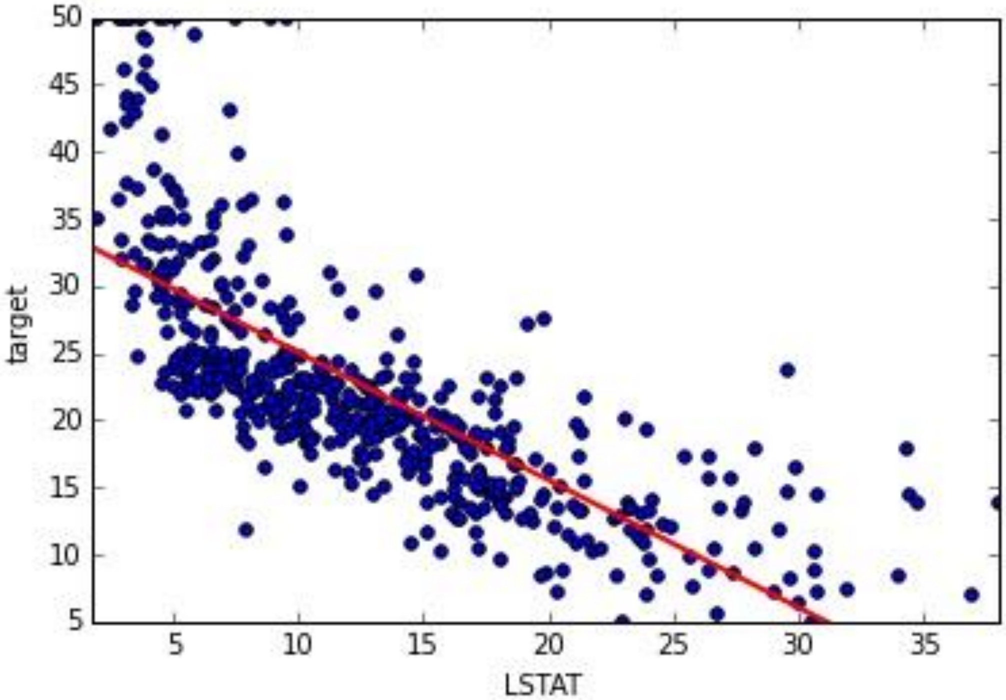
```
Adding interaction     CRIM *    CHAS R2: 0.011
Adding interaction     CRIM *      RM R2: 0.021
Adding interaction       ZN *      RM R2: 0.013
Adding interaction    INDUS *      RM R2: 0.038
Adding interaction    INDUS *     DIS R2: 0.013
Adding interaction      NOX *      RM R2: 0.027
Adding interaction       RM *     AGE R2: 0.024
Adding interaction       RM *     DIS R2: 0.018
Adding interaction       RM *     RAD R2: 0.049
Adding interaction       RM *     TAX R2: 0.054
Adding interaction       RM * PTRATIO R2: 0.041
Adding interaction       RM *       B R2: 0.020
Adding interaction       RM *   LSTAT R2: 0.064
```

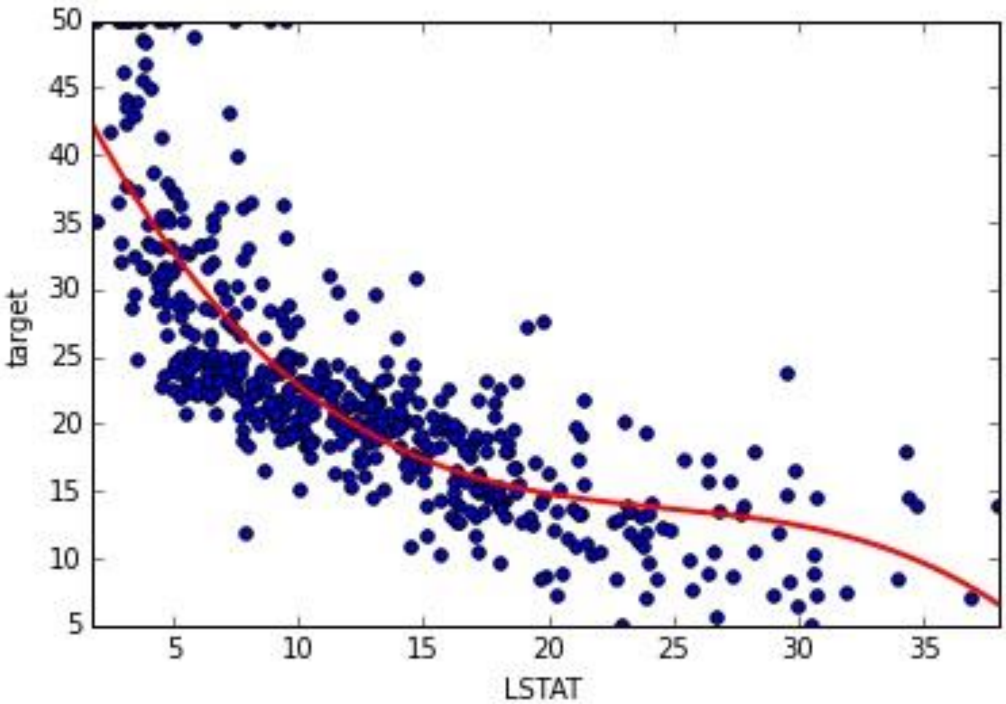
$$y = \beta_0 + \beta_1 x$$

$$y = \beta_0 + \beta_1 x + \beta_3 x^2$$

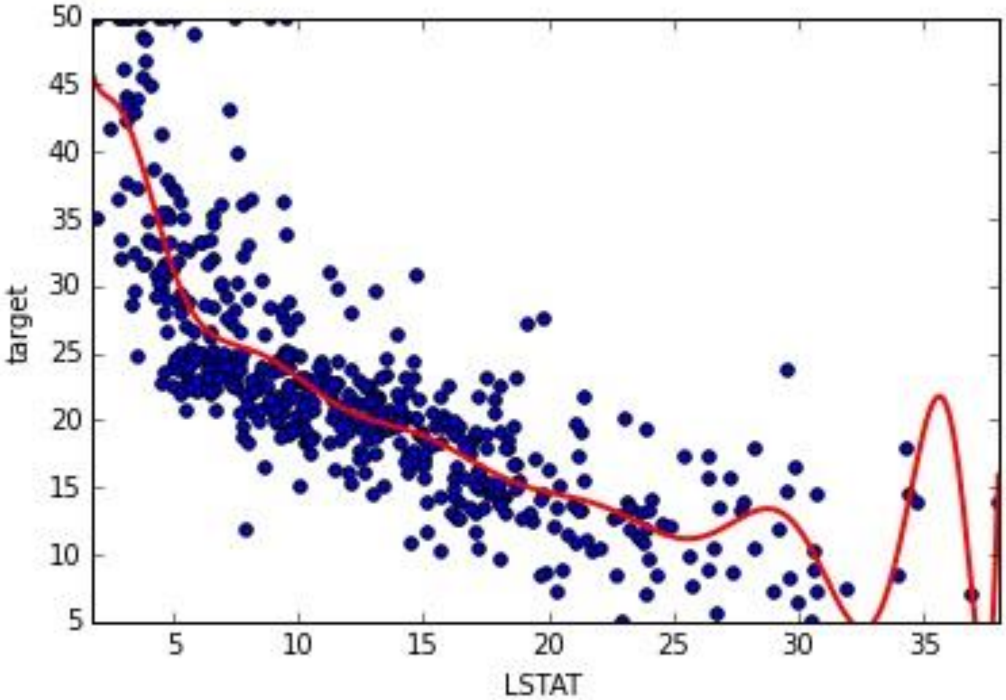
$$y = \beta_0 + \beta_1 x + \beta_3 x^2 + \beta_4 x^3$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2$$



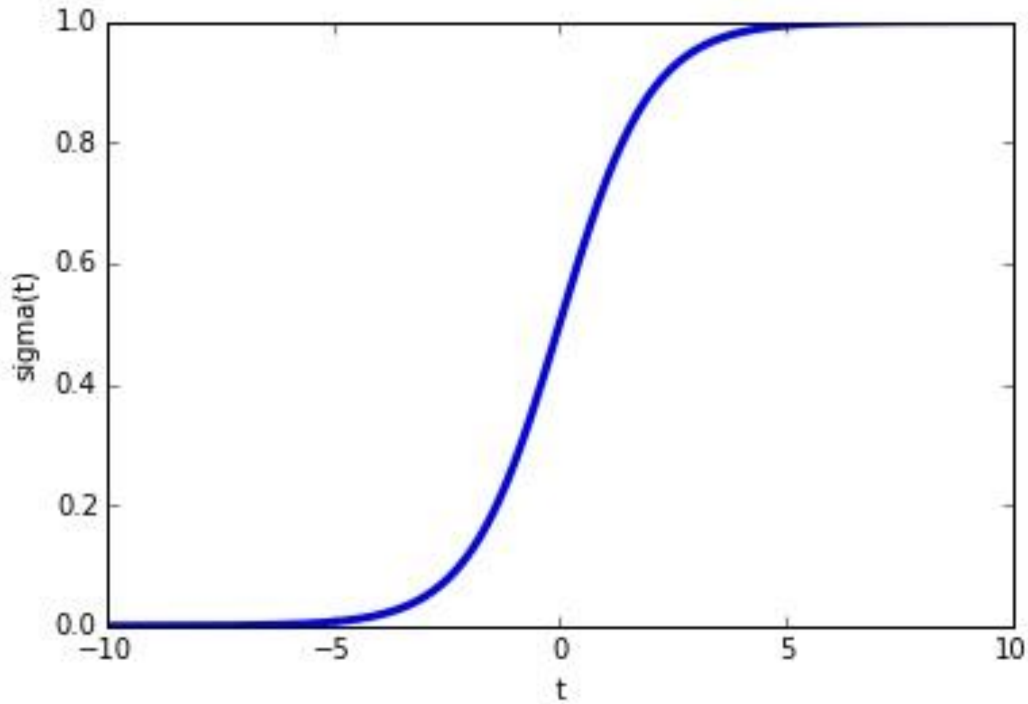


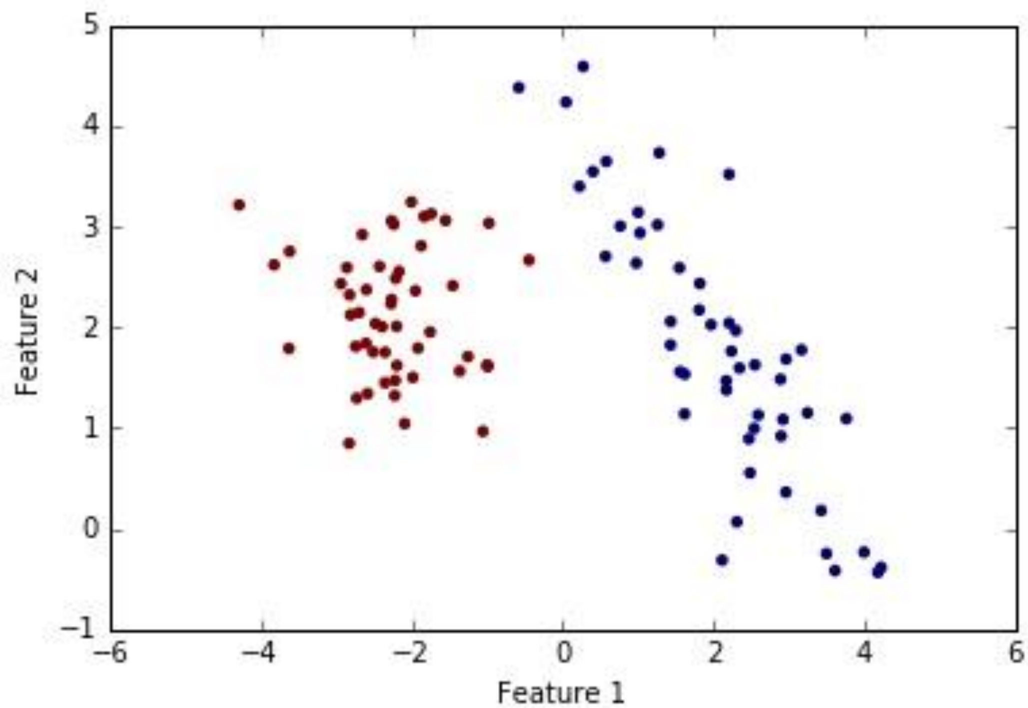

```
R2 degree -  1 polynomial :0.544
R2 degree -  2 polynomial :0.641
R2 degree -  3 polynomial :0.658
R2 degree -  5 polynomial :0.682
R2 degree - 15 polynomial :0.695
```

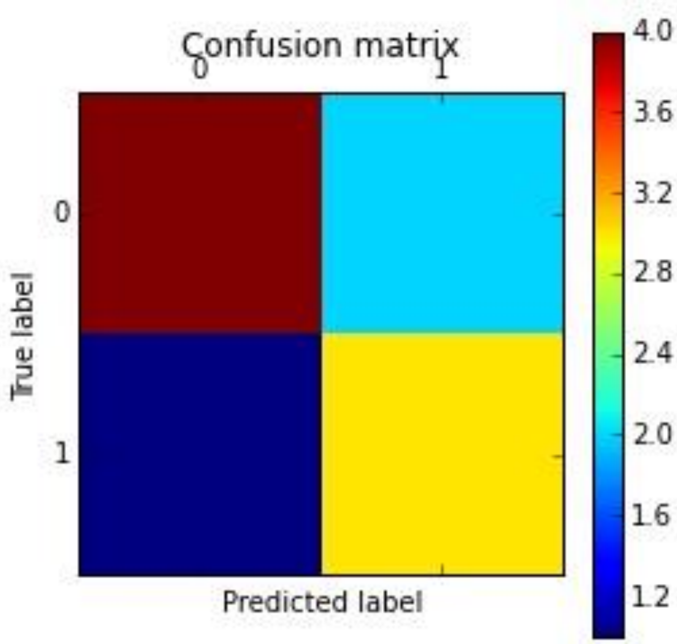


$\overline{x}$

Chapter 4: Logistic Regression





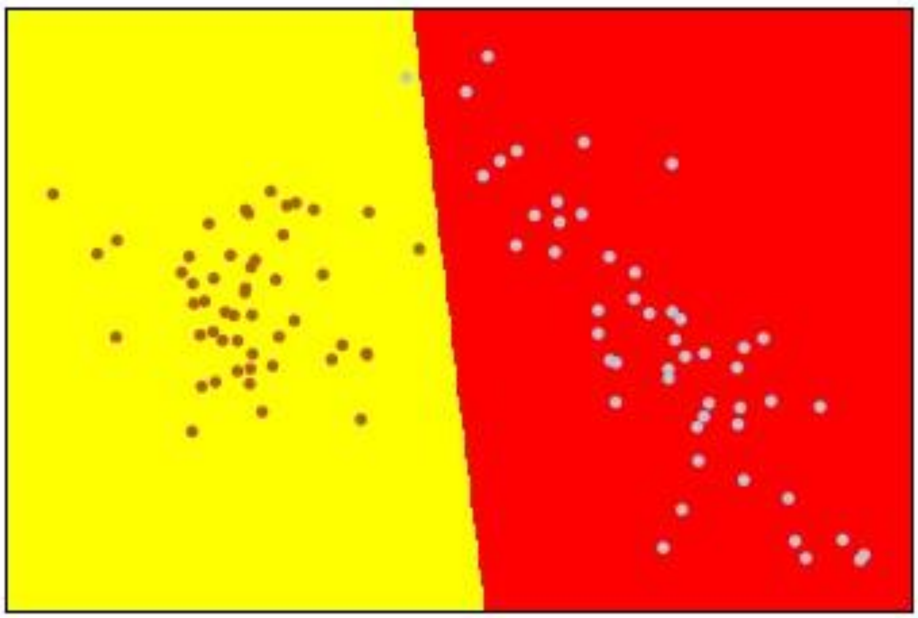


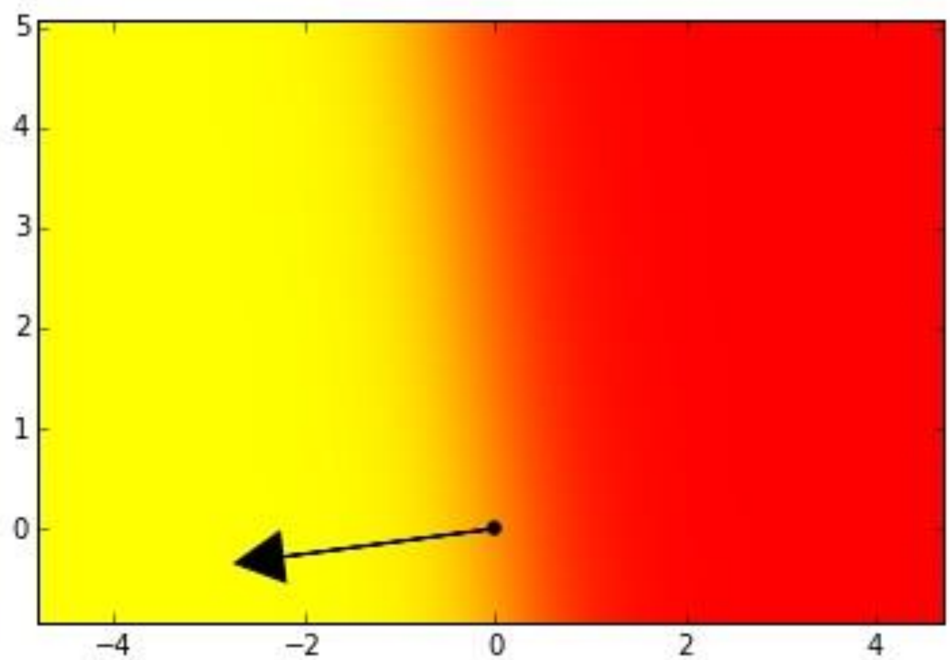

```
             precision    recall  f1-score   support

          0       0.80      0.67      0.73         6
          1       0.60      0.75      0.67         4

avg / total       0.72      0.70      0.70        10
```

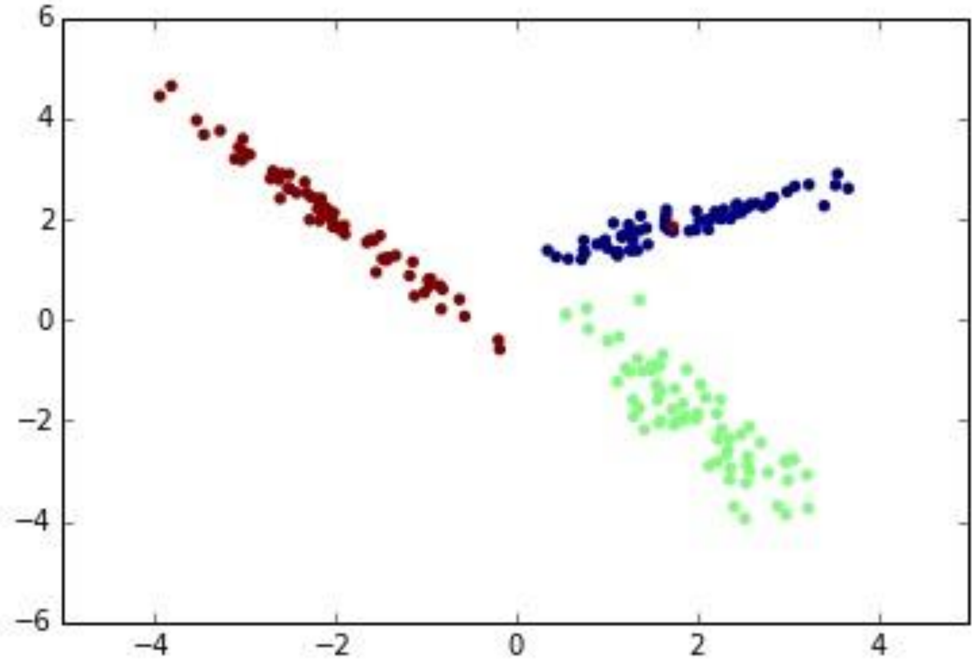
```
array([ 0.,  1.,  1.,  0.,  1.,  1.,  0.,  1.,  0.,  0.,  0.,  0.,  1.,
        0.,  1.,  0.,  0.,  1.,  1.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,
        0.,  0.,  0.,  1.,  0.,  1.,  0.])
```

```
array([ 0.06688448,  1.01981921,  1.08597427, -0.15225094,  1.05856628,
        0.8156161 ,  0.04837505,  0.7997539 ,  0.18942251, -0.03658995,
       -0.0462575 , -0.09640911,  1.0253004 , -0.17062754,  1.13642842,
        0.14052848, -0.00703683,  0.90903158,  1.26997191,  0.03606483,
       -0.19047191,  0.22476337, -0.05936491, -0.18559975,  0.28378888,
        0.01139188, -0.03559395,  0.22742328,  0.07485246,  1.24545626,
        0.13924533,  1.09388935,  0.35341582])
```

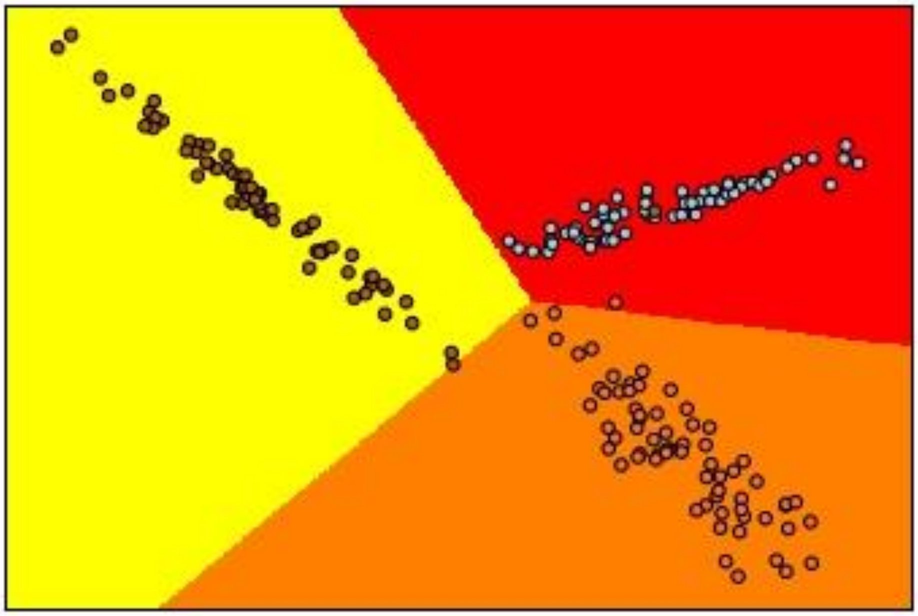
	precision	recall	f1-score	support
0.0	1.00	0.95	0.98	22
1.0	0.92	1.00	0.96	11
avg / total	0.97	0.97	0.97	33







	precision	recall	f1-score	support
0.0	1.00	1.00	1.00	24
1.0	1.00	1.00	1.00	22
2.0	1.00	1.00	1.00	20
avg / total	1.00	1.00	1.00	66



Logit Regression Results

Dep. Variable:	y	**No. Observations:**	10000
Model:	Logit	**Df Residuals:**	9989
Method:	MLE	**Df Model:**	10
Date:	Fri, 01 Jan 2016	**Pseudo R-squ.:**	0.3671
Time:	11:48:59	**Log-Likelihood:**	-4386.8
converged:	True	**LL-Null:**	-6931.5
		LLR p-value:	0.000

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
const	0.4299	0.039	11.023	0.000	0.353 0.506
x1	0.0671	0.015	4.410	0.000	0.037 0.097
x2	-0.7828	0.019	-41.947	0.000	-0.819 -0.746
x3	0.1221	0.016	7.815	0.000	0.091 0.153
x4	0.2841	0.016	18.150	0.000	0.253 0.315
x5	0.1469	0.014	10.283	0.000	0.119 0.175
x6	-0.3414	0.019	-17.636	0.000	-0.379 -0.303
x7	0.0503	0.014	3.481	0.000	0.022 0.079
x8	-0.1393	0.014	-9.642	0.000	-0.168 -0.111
x9	0.1127	0.014	7.931	0.000	0.085 0.141
x10	-0.4792	0.018	-27.340	0.000	-0.514 -0.445

```
[0.42991394845314063,
 0.067077096874709585,
 -0.7827957661488677,
 0.12208730826867409,
 0.28410283693190336,
 0.14689340914475549,
 -0.34143434245188609,
 0.050310756492560317,
 -0.1393205915231476,
 0.11267402173781312,
 -0.47916904027905627]
```

```
[0.42571117875899561,
 0.092754663986175351,
 -0.78381378869544127,
 0.093708745822509473,
 0.1675646650527122,
 0.10596527209458738,
 -0.41091578158018643,
 0.062219832489940362,
 -0.19435965629236054,
 0.2353120824478212,
 -0.48793778455042086]
```

$$(x_i, y_i) : x_i \in \mathbb{R}^n, y_i \in \{0,1\}$$

$$f : \mathbb{R}^n \to \{0,1\}$$

$$f1 = 2 \cdot \frac{precision \cdot recall}{precision \mid recall}$$

$$P\left(y_i = "1" \mid x_i\right)$$

$$y = X \cdot w$$

$$P\left(y = 1 \mid x\right) = \sigma\left(W^T \cdot x\right)$$

$$\sigma(t) = logit^{-1}(t) = \frac{1}{1+e^{-t}}$$

$$logit(p) = \log\left(\frac{p}{1-p}\right) = \log(p) - \log(1-p)$$

$$P(y = "1" \mid x) = \frac{1}{1+e^{-t}}$$
$$P(y = "0" \mid x) = 1 - \frac{1}{1+e^{-t}}$$

$$\log\left(\frac{P(y = "1" \mid x)}{P(y = "0" \mid x)}\right) = W^T \cdot x$$

$$logit\left(P(y = "1" \mid x)\right) = W^T \cdot x$$

$$\sigma(t) = \frac{1}{1+e^{-t}}$$

$$\sigma'(t) = \frac{\vartheta}{\vartheta z}\frac{1}{1+e^{-t}} = \frac{e^{-t}}{\left(1+e^{-t}\right)^2} = \left(\frac{1}{1+e^{-t}}\right)\cdot\left(1-\frac{1}{1+e^{-t}}\right) = \sigma(t)\cdot\left(1-\sigma(t)\right)$$

$$\begin{aligned} L(W) &= P(Y \mid X; W) \\ &= \prod_i P(y_i \mid x_i; W) \\ &= \prod_i \left(\sigma\left(W^T \cdot x_i\right)\right)^{y_i} \cdot \left(1-\sigma\left(W^T \cdot x_i\right)\right)^{1-y_i} \end{aligned}$$

$$\hat{L}(W) = \log\left(L(W)\right)$$
$$= \sum_i y_i \log\left(\sigma\left(W^T \cdot x_i\right)\right) + \left(1-y_i\right)\log\left(1-\sigma\left(W^T \cdot x_i\right)\right)$$

$$\frac{\vartheta}{\vartheta w_k}\hat{L}(W) = \cdots = \left(y - \sigma\left(W^T \cdot x\right)\right)\cdot x_k$$

$$W \leftarrow W + \alpha \nabla \hat{L}(W)$$

$$w_k \leftarrow w_k + \alpha \cdot \left(y - \frac{1}{1+e^{-W^T \cdot x}}\right)\cdot x_k$$

Chapter 5: Data Preparation

$$y = \beta_0 + \beta X$$

```
coefficients: [ -0.10717   0.0464    0.02086   2.68856 -17.79576   3.80475   0.00075
  -1.47576   0.30566  -0.01233  -0.95346   0.00939  -0.52547]
intercept: 36.491
```

```
CRIM          0.00632
ZN            0.00000
INDUS         0.46000
CHAS          0.00000
NOX           0.38500
RM            3.56100
AGE           2.90000
DIS           1.12960
RAD           1.00000
TAX         187.00000
PTRATIO      12.60000
B             0.32000
LSTAT         1.73000
target        5.00000
dtype: float64
```

```
coefficients: [ -0.10717   0.0464    0.02086   2.68856 -17.79576   3.80475   0.00075
  -1.47576   0.30566  -0.01233  -0.95346   0.00939  -0.52547]
intercept: 22.533
```

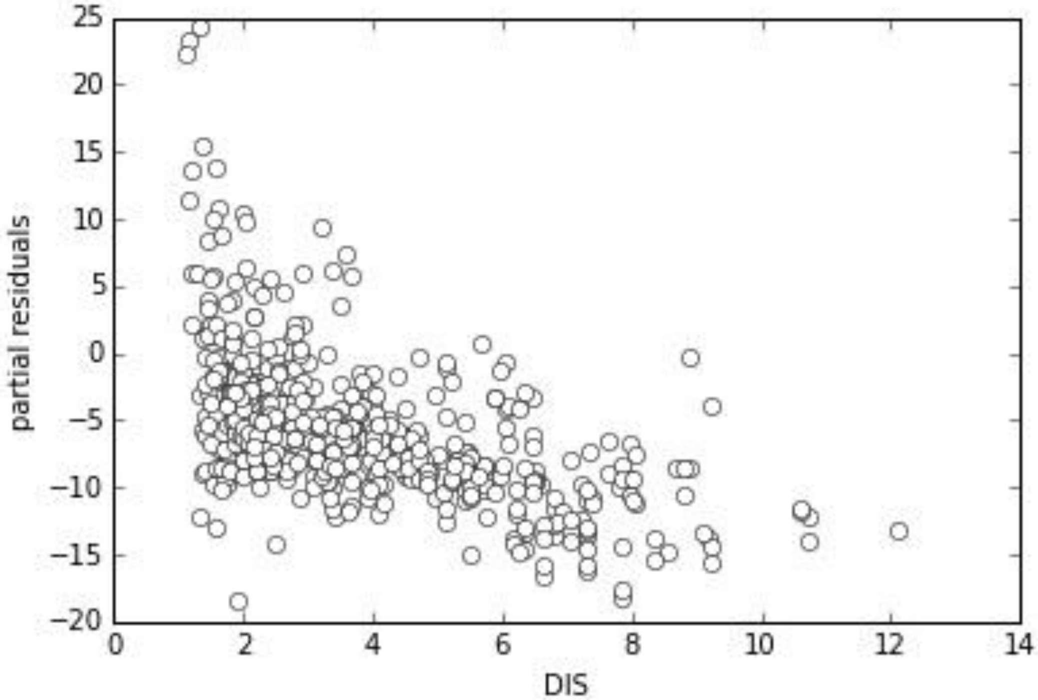
```
coefficients: [-0.92041  1.08098  0.14297  0.6822   -2.06009   2.67064   0.02112 -3.10445
  2.65879 -2.0759   -2.06216   0.85664 -3.74868]
intercept: 22.533
```

```
coefficients: [ -9.53495   4.63952   0.56907   2.68856  -8.64874  19.857     0.07293
 -16.22877   7.03007  -6.46058  -8.96256   3.72488 -19.04291]
intercept: 26.613
```

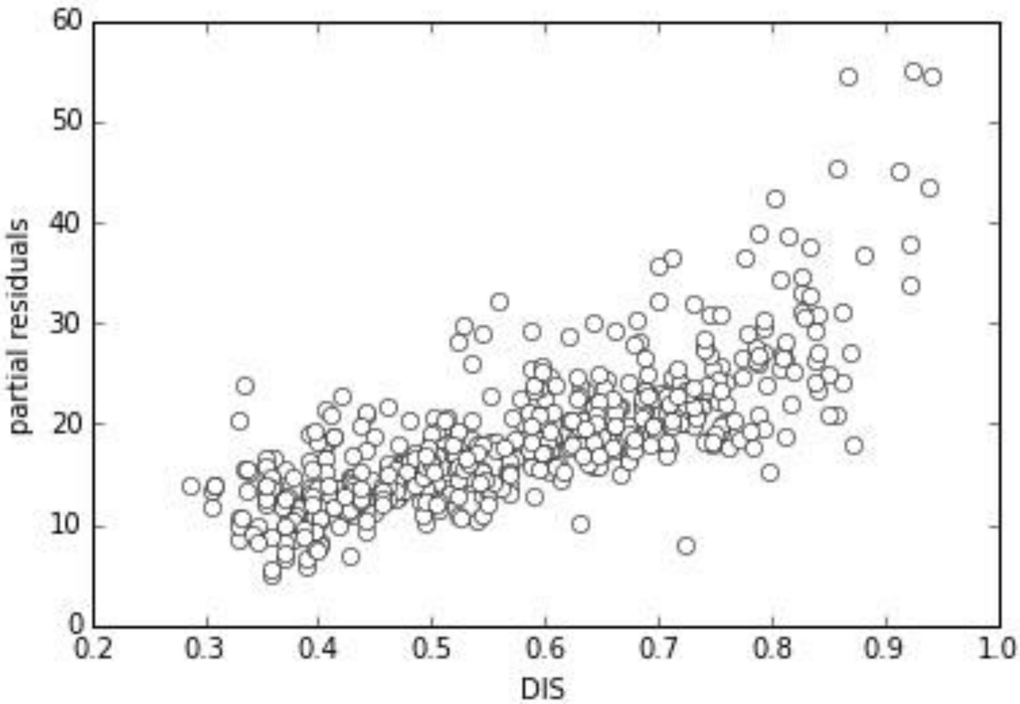
```
Optimization terminated successfully.
         Current function value: 0.206632
         Iterations 9
                           Logit Regression Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  506
Model:                          Logit   Df Residuals:                      492
Method:                           MLE   Df Model:                           13
Date:                Tue, 20 Oct 2015   Pseudo R-squ.:                  0.6289
Time:                        16:33:29   Log-Likelihood:                -104.56
converged:                       True   LL-Null:                       -281.76
                                        LLR p-value:                 9.147e-68
==============================================================================
                 coef    std err          z      P>|z|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
const         -3.0541      0.356     -8.572      0.000        -3.752    -2.356
x1            -0.0949      0.389     -0.244      0.807        -0.857     0.667
x2             0.2543      0.252      1.008      0.314        -0.240     0.749
x3            -0.7570      0.403     -1.880      0.060        -1.546     0.032
x4             0.2452      0.205      1.195      0.232        -0.157     0.648
x5            -0.7924      0.519     -1.527      0.127        -1.810     0.225
x6             1.3244      0.318      4.168      0.000         0.702     1.947
x7             0.0982      0.313      0.314      0.754        -0.515     0.712
x8            -1.2390      0.345     -3.591      0.000        -1.915    -0.563
x9             2.7664      0.719      3.849      0.000         1.358     4.175
x10           -1.8228      0.680     -2.682      0.007        -3.155    -0.491
x11           -0.7635      0.264     -2.888      0.004        -1.282    -0.245
x12           -0.2062      0.349     -0.591      0.554        -0.890     0.477
x13           -2.6208      0.521     -5.031      0.000        -3.642    -1.600
==============================================================================

Optimization terminated successfully.
         Current function value: 0.556842
         Iterations 5
                           Logit Regression Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  506
Model:                          Logit   Df Residuals:                      505
Method:                           MLE   Df Model:                            0
Date:                Tue, 20 Oct 2015   Pseudo R-squ.:                   0.000
Time:                        16:33:29   Log-Likelihood:                -281.76
converged:                       True   LL-Null:                       -281.76
                                        LLR p-value:                       nan
==============================================================================
                 coef    std err          z      P>|z|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
const         -1.1251      0.103    -10.886      0.000        -1.328    -0.923
==============================================================================

probability of value above 25 using just a constant: 0.245
```

```
Optimization terminated successfully.
         Current function value: 0.292346
         Iterations 32
                           Logit Regression Results
==============================================================================
Dep. Variable:                      y   No. Observations:                   36
Model:                          Logit   Df Residuals:                       29
Method:                           MLE   Df Model:                            6
Date:                Tue, 20 Oct 2015   Pseudo R-squ.:                  0.5744
Time:                        16:33:30   Log-Likelihood:                -10.524
converged:                       True   LL-Null:                       -24.731
                                        LLR p-value:                 7.856e-05
=====================================================================================
                       coef    std err          z      P>|z|      [95.0% Conf. Int.]
-------------------------------------------------------------------------------------
const                0.2393   1.76e+07   1.36e-08      1.000     -3.44e+07  3.44e+07
outlook_overcast     2.9833   6.69e+07   4.46e-08      1.000     -1.31e+08  1.31e+08
outlook_rainy       -2.1746   6.69e+07  -3.25e-08      1.000     -1.31e+08  1.31e+08
outlook_sunny       -0.5695   6.69e+07  -8.51e-09      1.000     -1.31e+08  1.31e+08
temperature_cool    -2.1996      6e+07  -3.66e-08      1.000     -1.18e+08  1.18e+08
temperature_hot      0.3045      6e+07   5.07e-09      1.000     -1.18e+08  1.18e+08
temperature_mild     2.1344      6e+07   3.55e-08      1.000     -1.18e+08  1.18e+08
humidity_high       -2.0459   2.24e+07  -9.15e-08      1.000     -4.38e+07  4.38e+07
humidity_normal      2.2851   2.24e+07   1.02e-07      1.000     -4.38e+07  4.38e+07
windy_FALSE          1.3162   4.47e+07   2.94e-08      1.000     -8.77e+07  8.77e+07
windy_TRUE          -1.0770   4.47e+07  -2.41e-08      1.000     -8.77e+07  8.77e+07
=====================================================================================
```

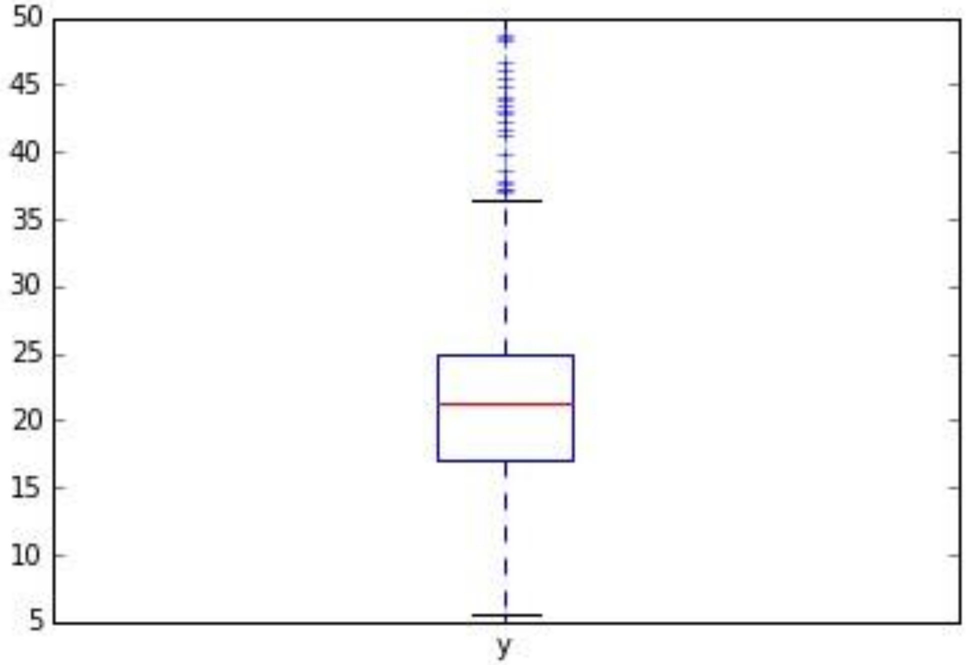
```
Optimization terminated successfully.
         Current function value: 0.292346
         Iterations 8
                           Logit Regression Results
==============================================================================
Dep. Variable:                      y   No. Observations:                   36
Model:                          Logit   Df Residuals:                       29
Method:                           MLE   Df Model:                            6
Date:                Tue, 20 Oct 2015   Pseudo R-squ.:                  0.5744
Time:                        16:33:30   Log-Likelihood:                -10.524
converged:                       True   LL-Null:                       -24.731
                                        LLR p-value:                 7.856e-05
=====================================================================================
                       coef    std err          z      P>|z|      [95.0% Conf. Int.]
-------------------------------------------------------------------------------------
const                5.4055      2.196      2.462      0.014         1.102     9.709
outlook_overcast     3.5528      1.721      2.064      0.039         0.179     6.927
outlook_rainy       -1.6051      1.357     -1.183      0.237        -4.265     1.055
temperature_cool    -4.3340      1.867     -2.322      0.020        -7.993    -0.675
temperature_hot     -1.8299      1.478     -1.238      0.216        -4.727     1.067
humidity_high       -4.3310      1.645     -2.633      0.008        -7.555    -1.107
windy_TRUE          -2.3932      1.325     -1.807      0.071        -4.989     0.203
=====================================================================================
```

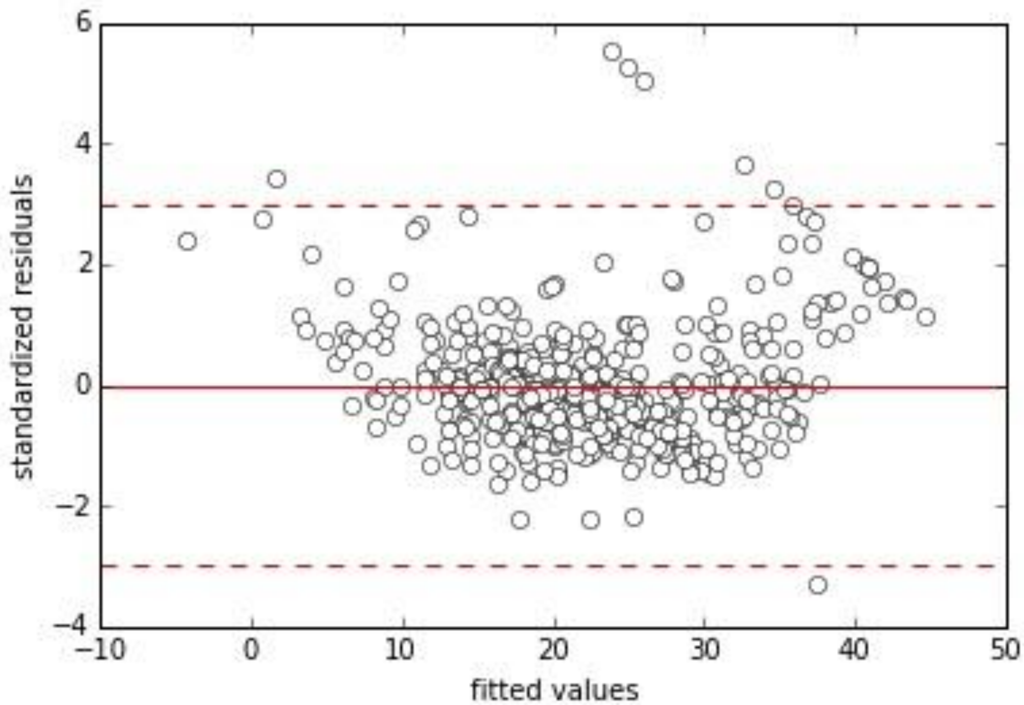


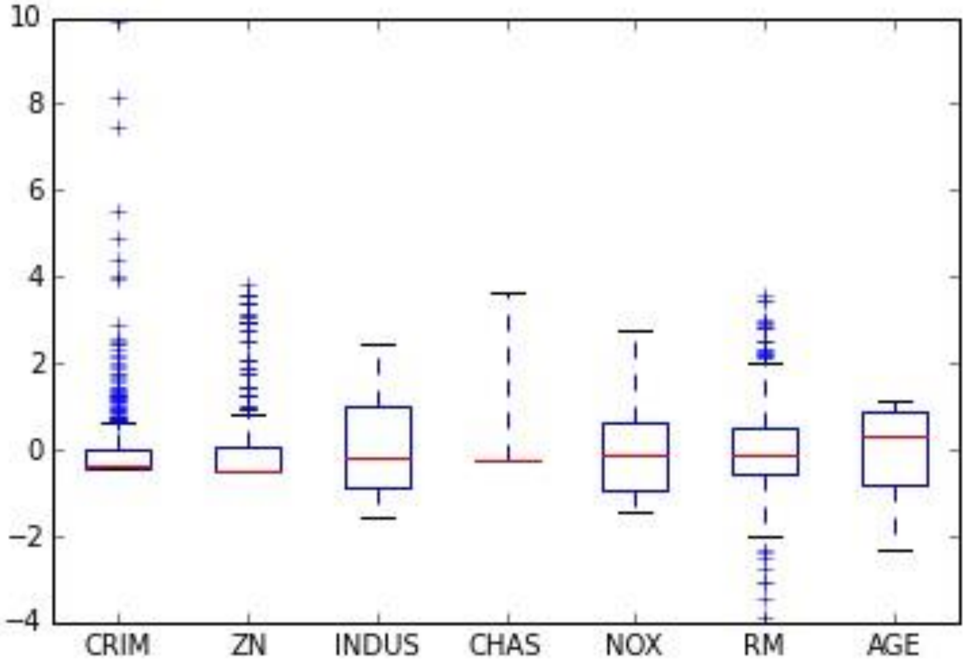
partial residuals
DIS
25
20
15
10
5
0
−5
−10
−15
−20
0
2
4
6
8
10
12
14

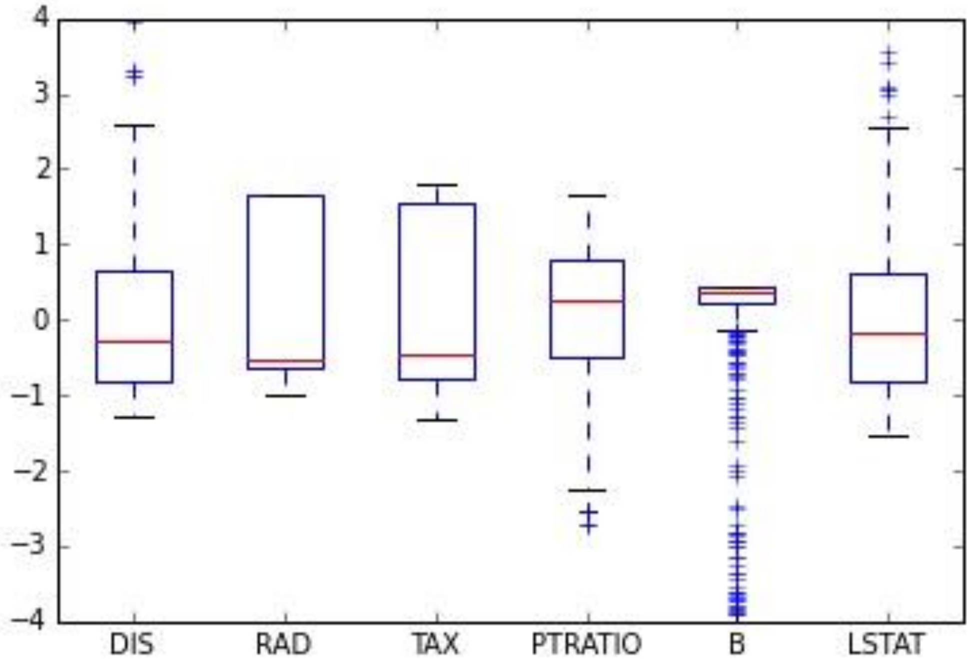


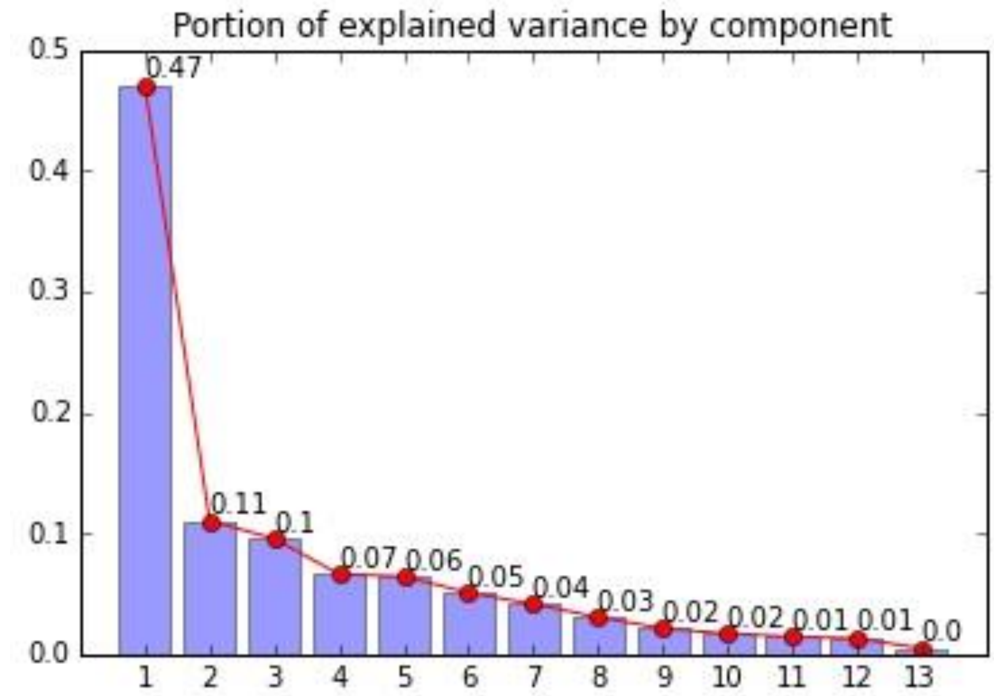
partial residuals
DIS
60
50
40
30
20
10
0
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

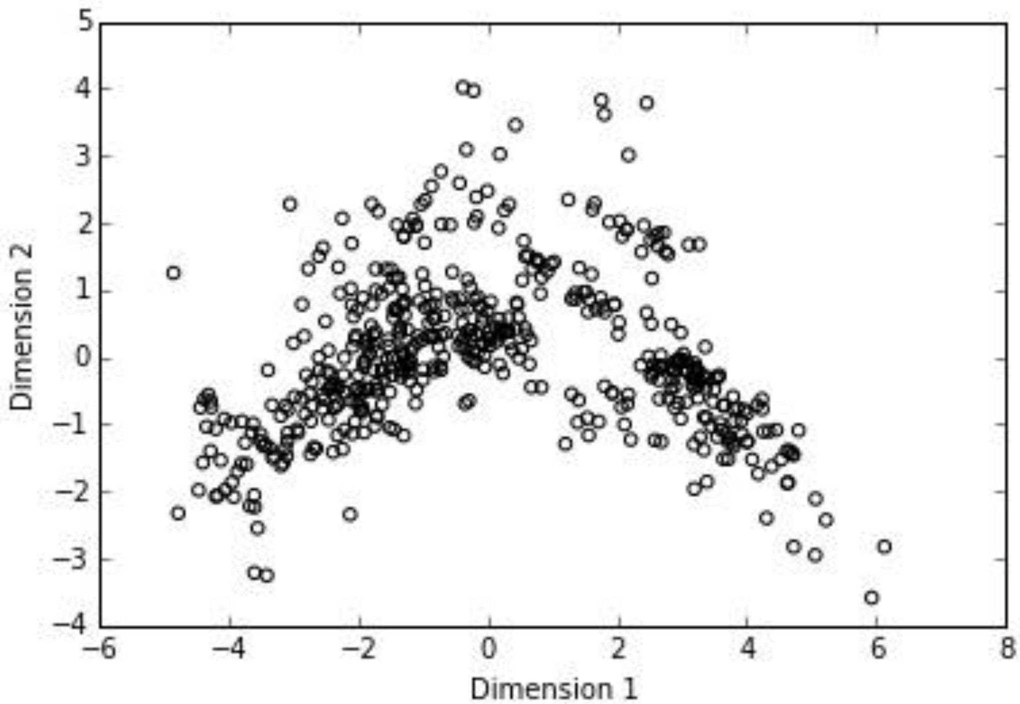

50
45
40
35
30
25
20
15
10
5
y

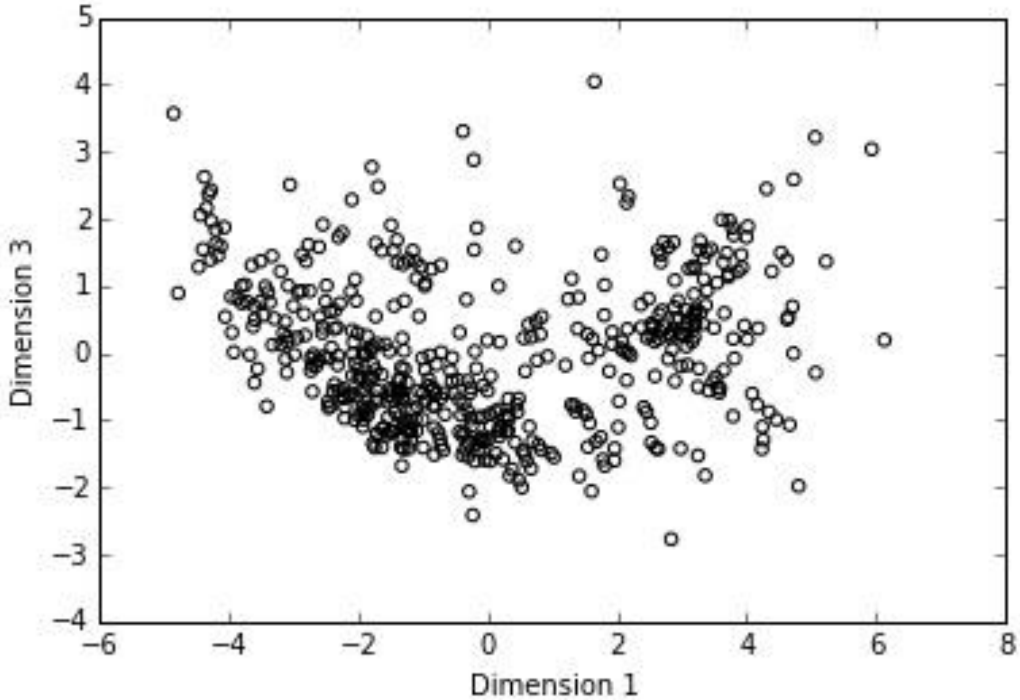

6
4
2
0
−2
−4
standardized residuals
−10
0
10
20
30
40
50
fitted values


10
8
6
4
2
0
−2
−4
CRIM
ZN
INDUS
CHAS
NOX
RM
AGE

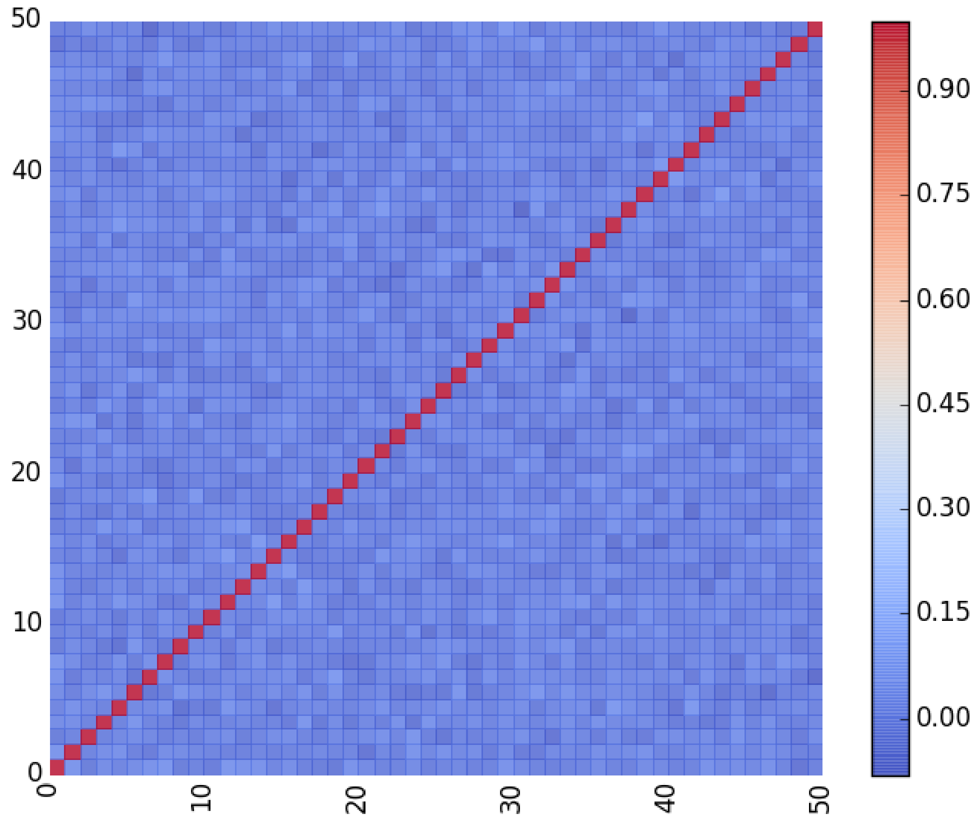

4
3
2
1
0
−1
−2
−3
−4
DIS
RAD
TAX
PTRATIO
B
LSTAT

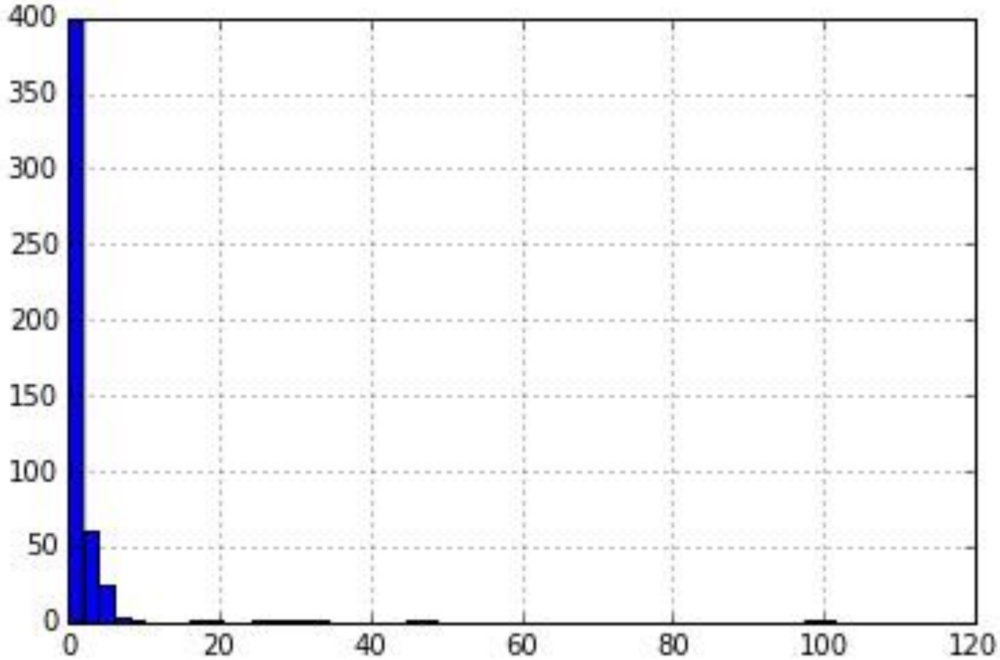

Portion of explained variance by component
0.47
0.11
0.1
0.07
0.06
0.05
0.04
0.03
0.02
0.02
0.01
0.01
0.0


Dimension 2
Dimension 1


Dimension 3
Dimension 1
5
4
3
2
1
0
−1
−2
−3
−4
−6
−4
−2
0
2
4
6
8

```
[9, 3, 8, 5, 7, 0, 8, 3, 9, 3] [1, 2, 4, 6]
[4, 7, 3, 5, 7, 1, 4, 3, 2, 1] [0, 8, 9, 6]
[7, 8, 5, 3, 7, 5, 3, 6, 6, 3] [0, 1, 2, 9, 4]
[1, 6, 7, 4, 3, 1, 9, 5, 4, 6] [0, 8, 2]
[6, 3, 6, 1, 6, 6, 0, 7, 3, 8] [9, 2, 4, 5]
```

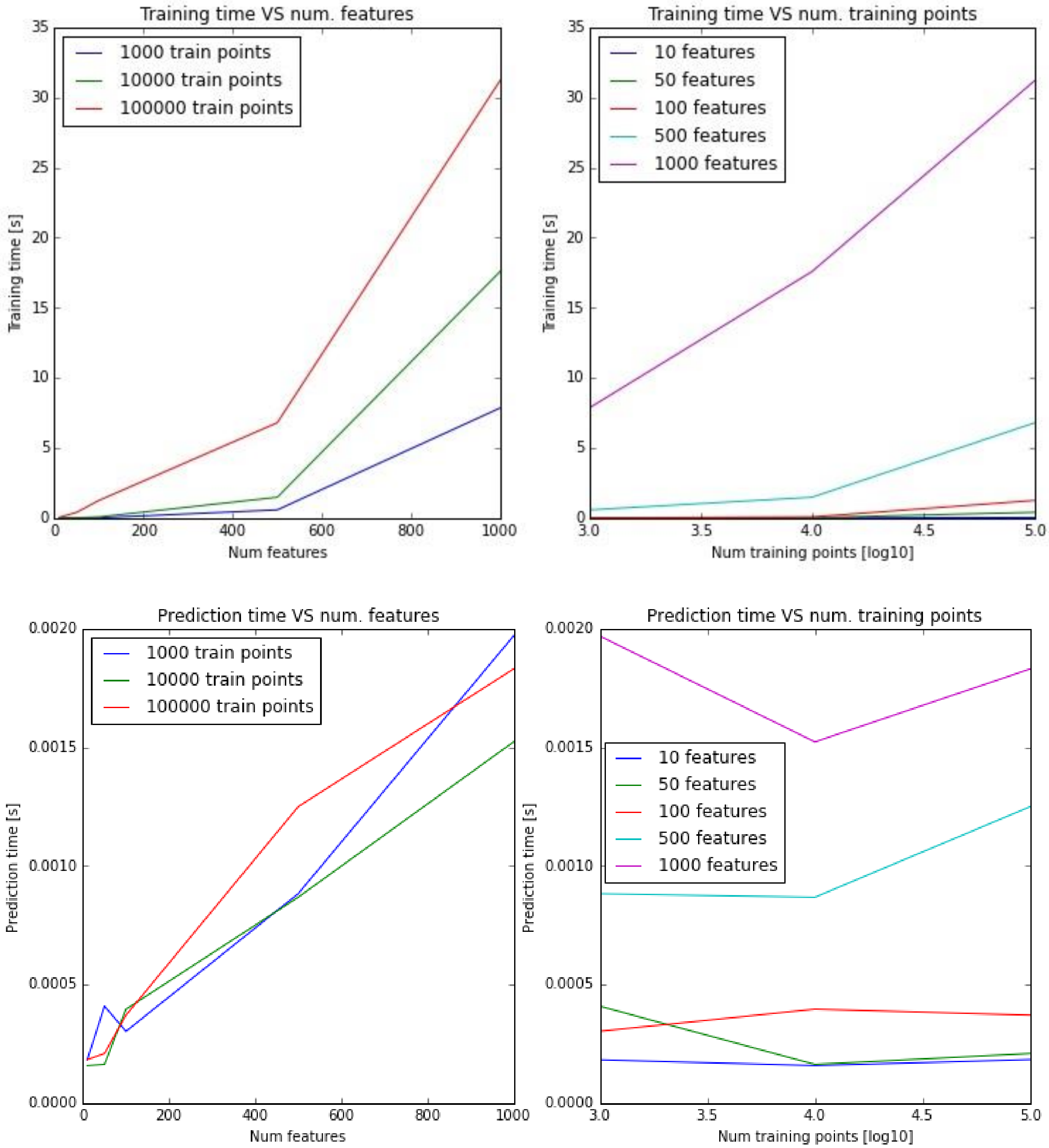


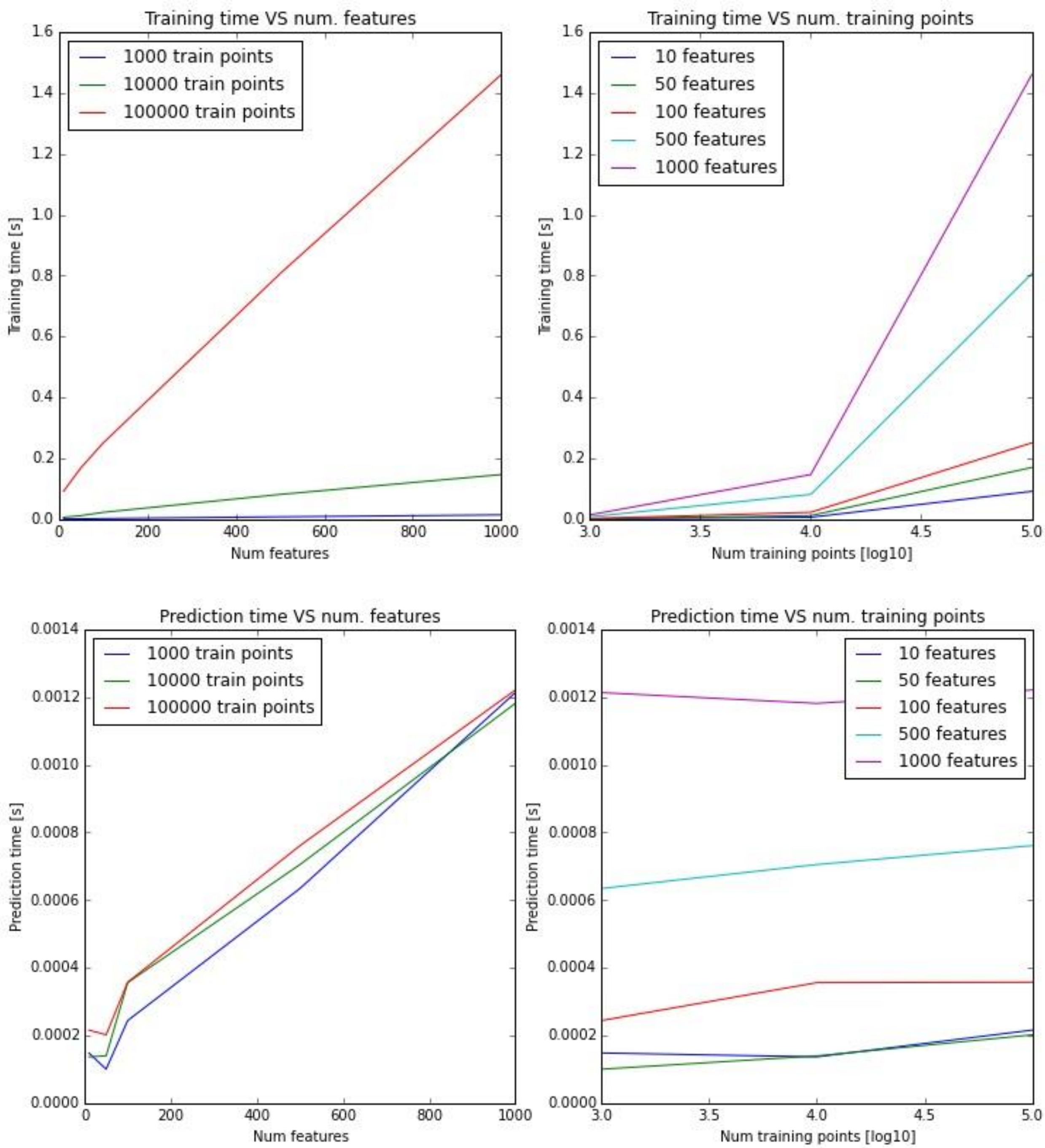


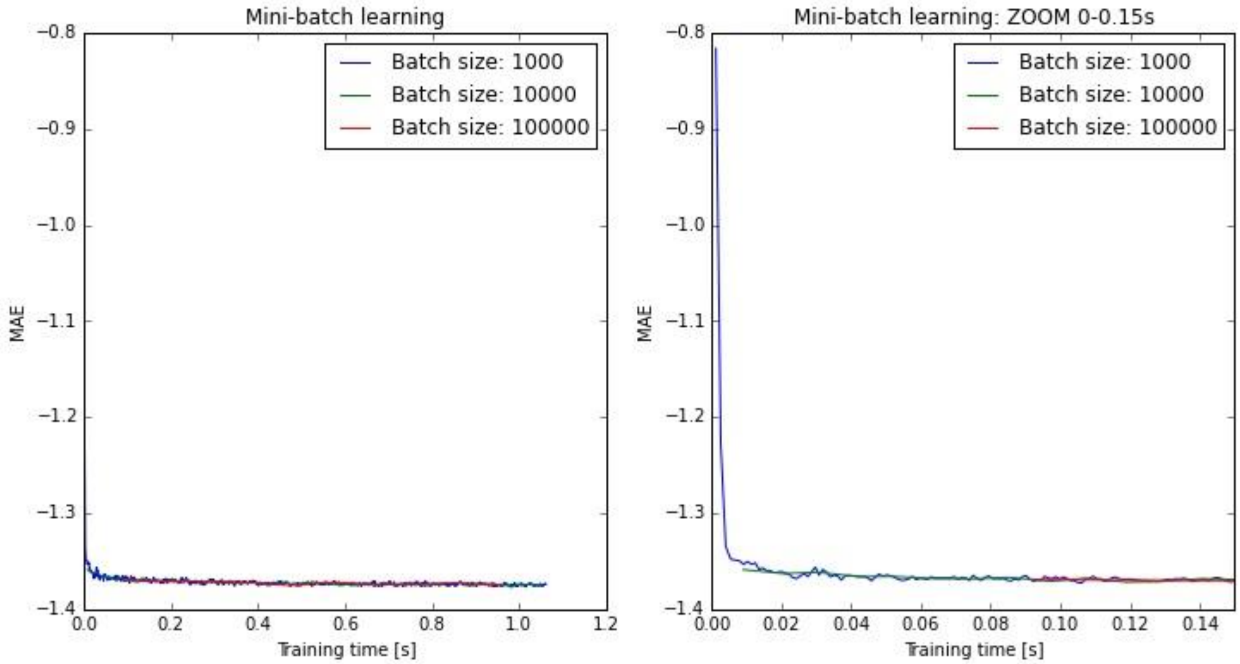
$$w_j = w_j - \frac{\alpha}{n} * \left(\sum (Xw - y) * x_j - \lambda * {w_j}^2 \right)$$

$$w_j = w_j - \frac{\alpha}{n} * \left(\sum (Xw - y) * x_j + \lambda * |w_j| \right)$$

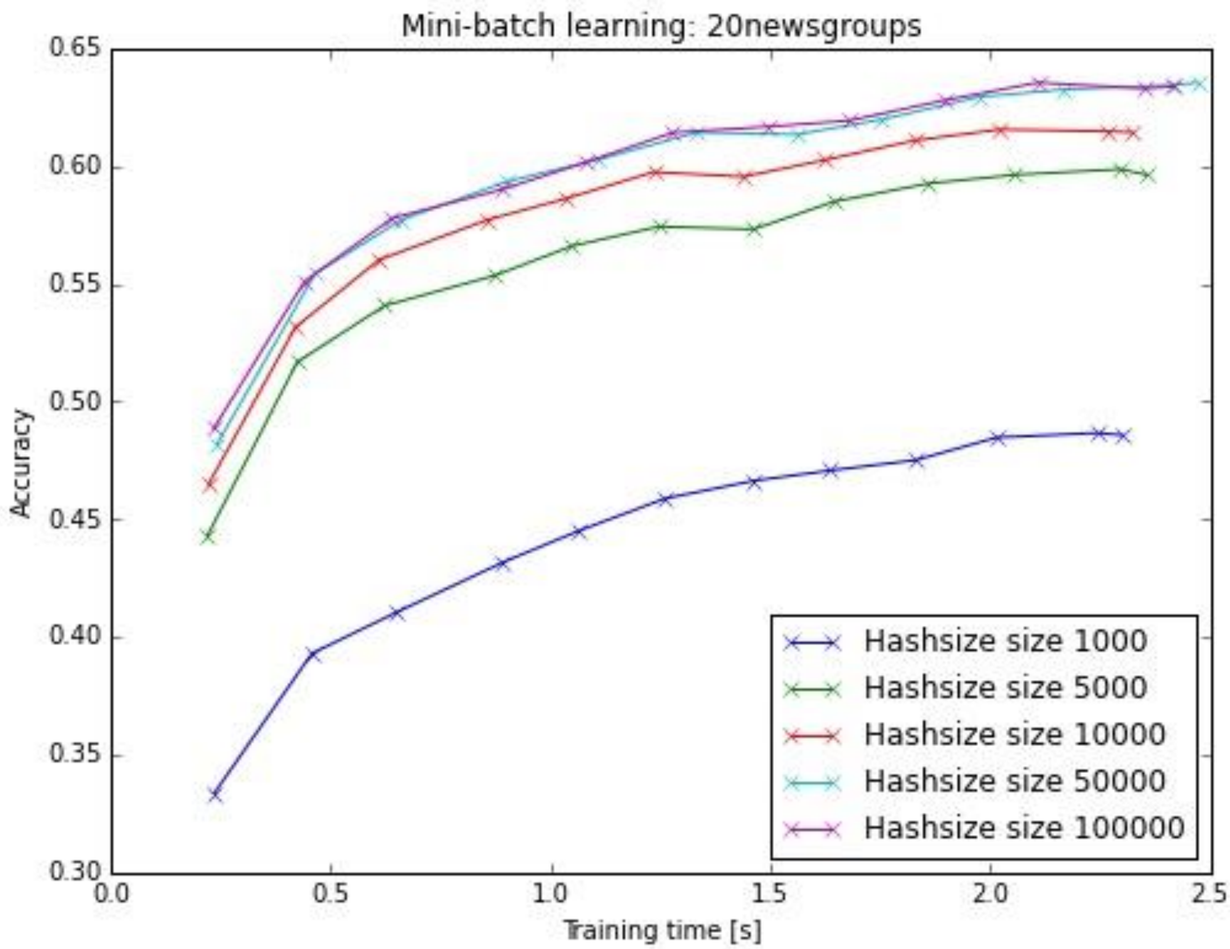
Chapter 7: Online and Batch Learning




Training time VS num. features
1000 train points
10000 train points
100000 train points
Training time [s]
Num features
Training time VS num. training points
10 features
50 features
100 features
500 features
1000 features
Training time [s]
Num training points [log10]
Prediction time VS num. features
1000 train points
10000 train points
100000 train points
Prediction time [s]
Num features
Prediction time VS num. training points
10 features
50 features
100 features
500 features
1000 features
Prediction time [s]
Num training points [log10]



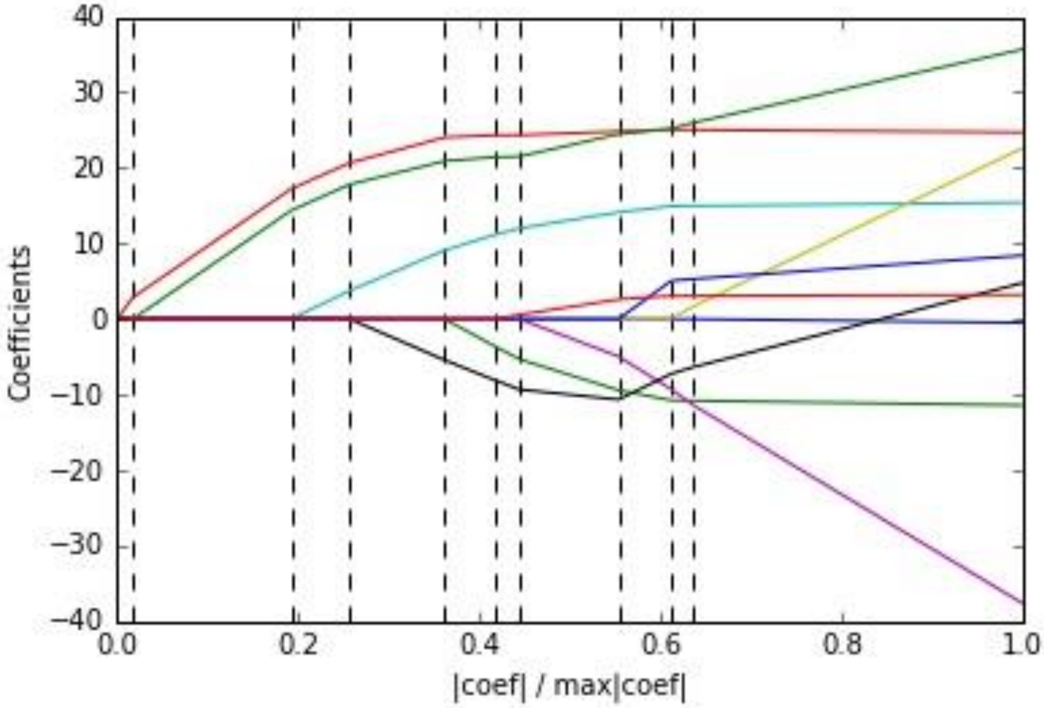
Mini-batch learning
Mini-batch learning: ZOOM 0-0.15s
Batch size: 1000
Batch size: 10000
Batch size: 100000
MAE
Training time [s]

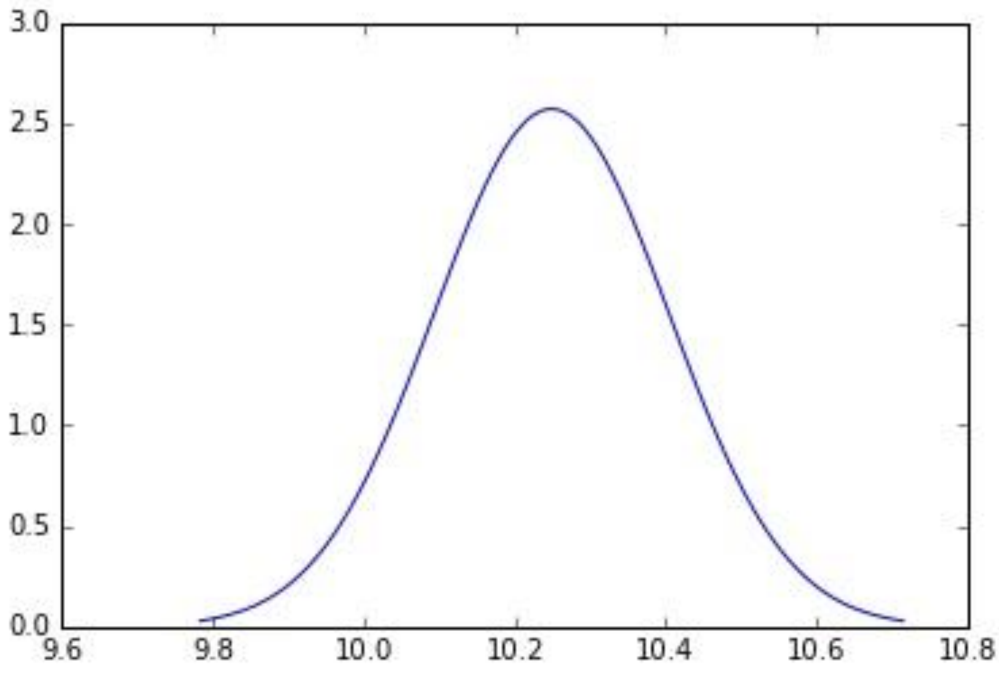


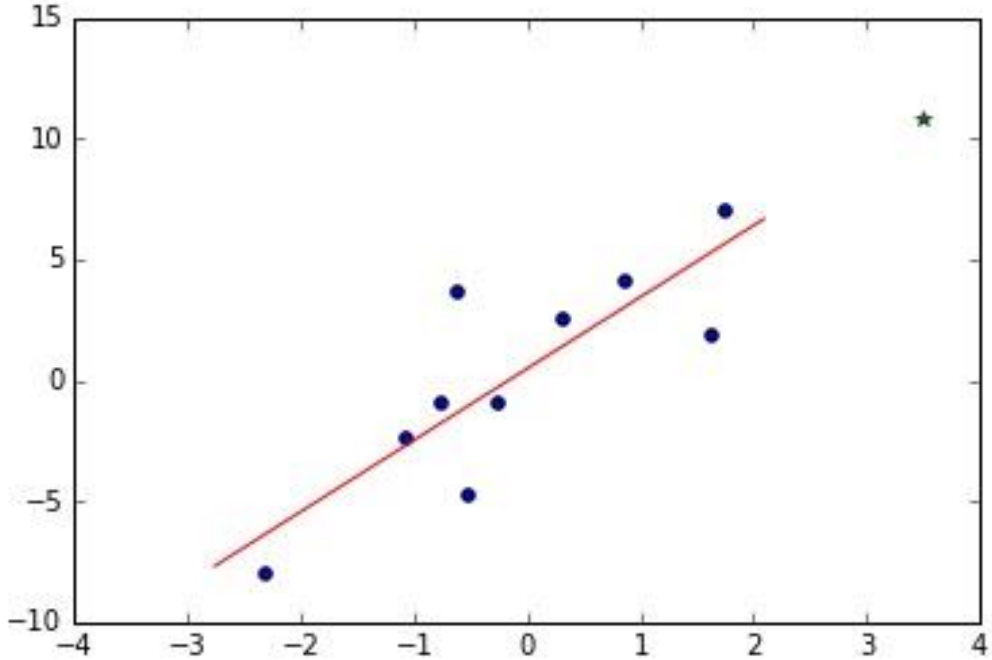
Mini-batch learning: 20newsgroups
Accuracy
Training time [s]
Hashsize size 1000
Hashsize size 5000
Hashsize size 10000
Hashsize size 50000
Hashsize size 100000

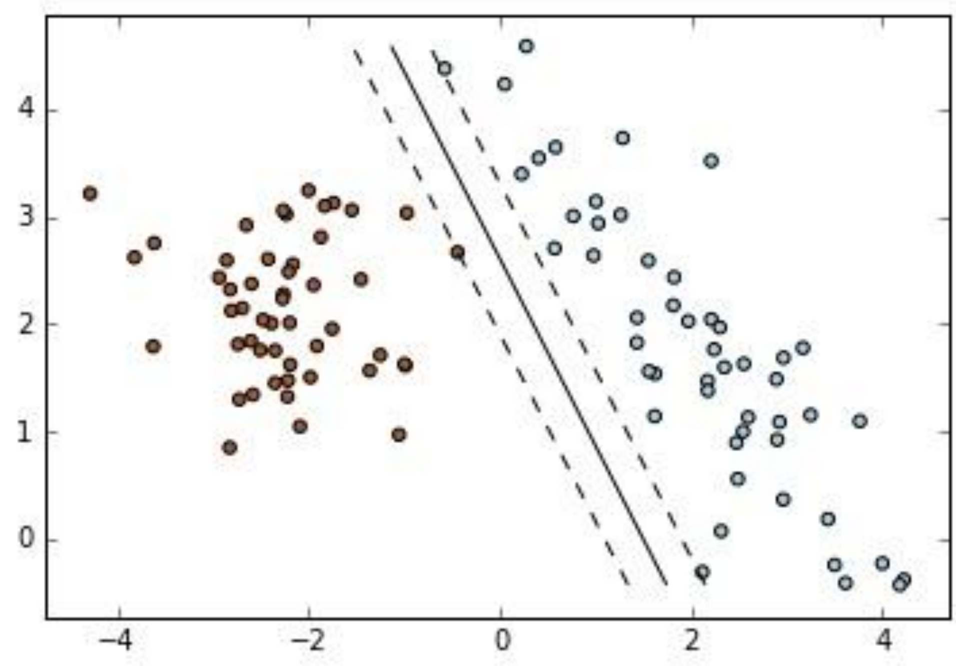
```
Report: Mini-batch size 1000
First output after [s]: 0.0007998943328857422
First model MAE [log10]: -0.942320304943
Total training time [s]: 1.3718714714050293
Final MAE [log10]:  -1.24036819201

Report: Mini-batch size 10000
First output after [s]: 0.007853984832763672
First model MAE [log10]: -1.23171862851
Total training time [s]: 1.308701992034912
Final MAE [log10]:  -1.24038903474

Report: Mini-batch size 100000
First output after [s]: 0.05989503860473633
First model MAE [log10]: -1.24053929732
Total training time [s]: 1.1995868682861328
Final MAE [log10]:  -1.24053790326
```

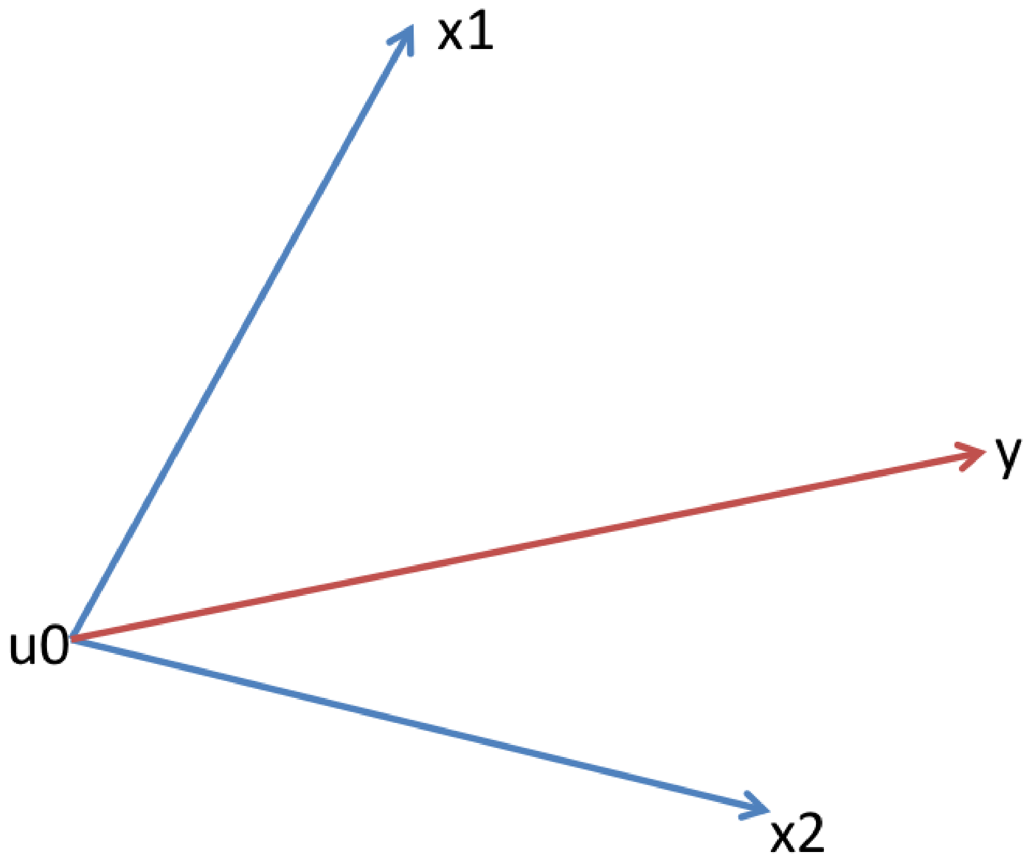
Chapter 8: Advanced Regression Methods

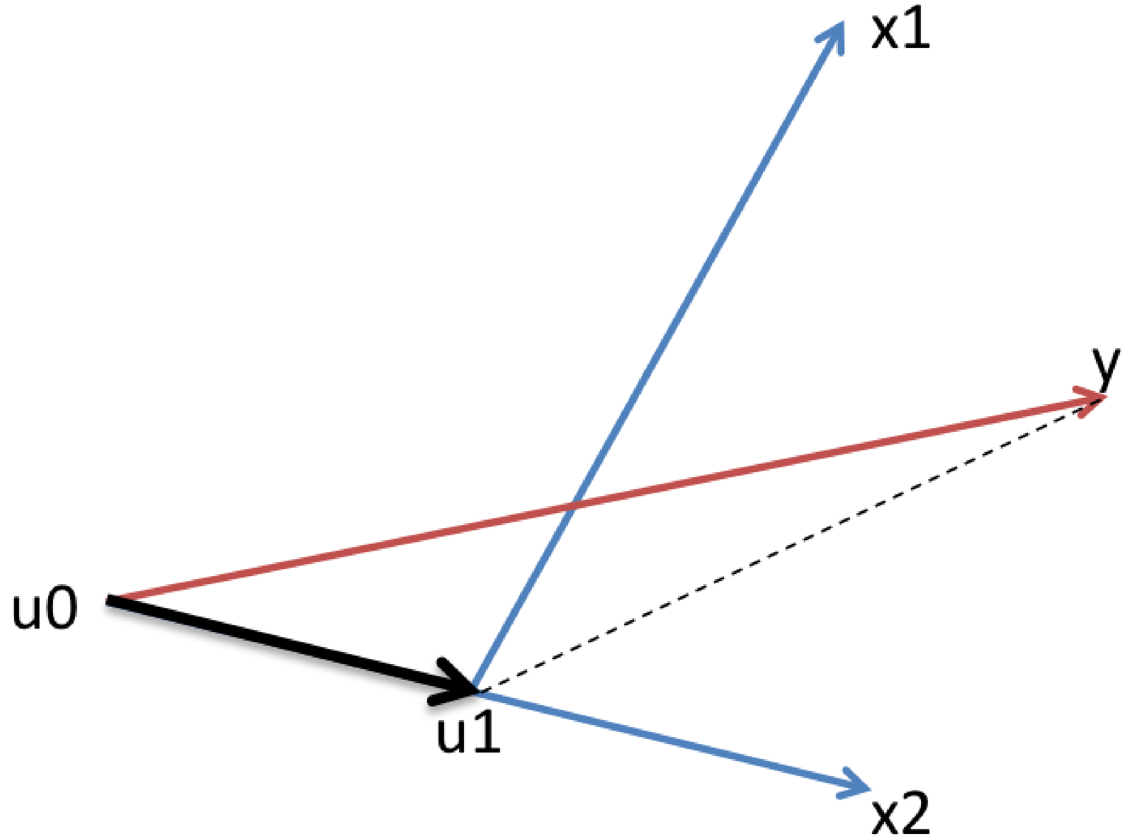


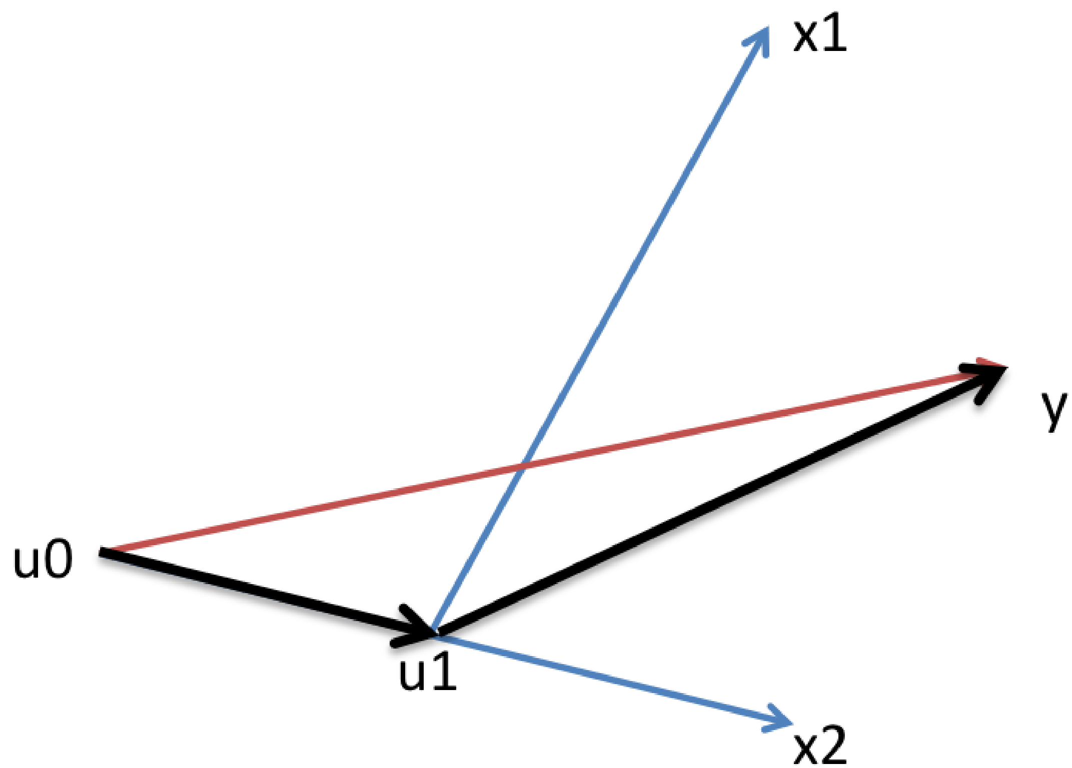



15
10
5
0
−5
−10
−4
−3
−2
−1
0
1
2
3
4


4
3
2
1
0
−4
−2
0
2
4


x1
y
u0
x2


x1
y
u0
u1
x2

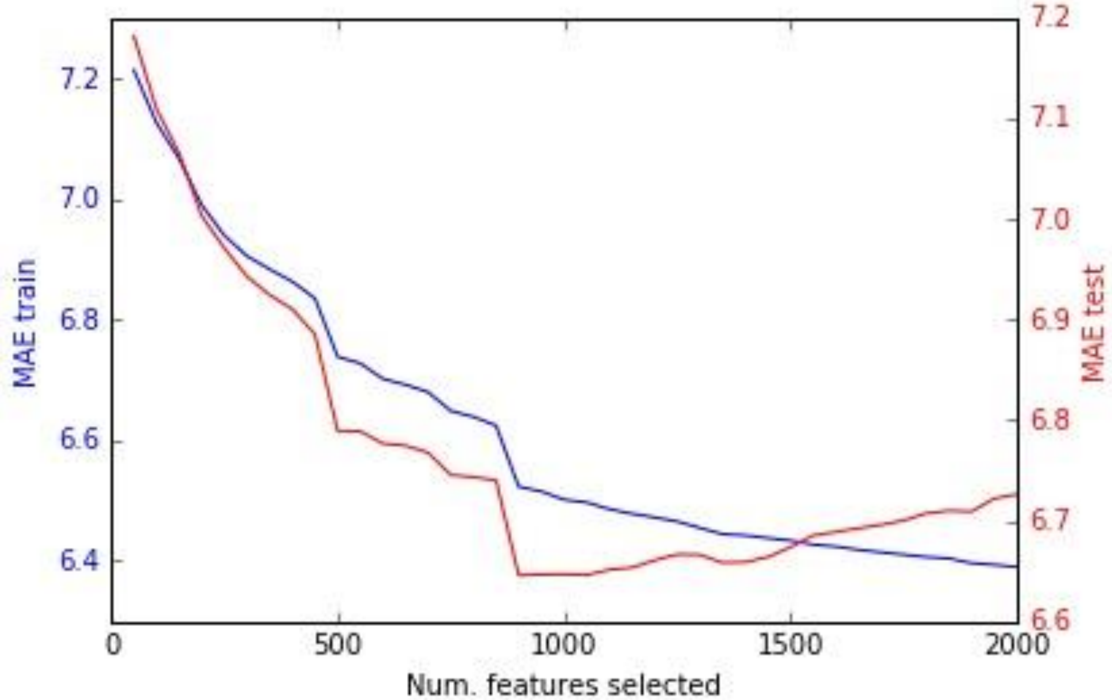


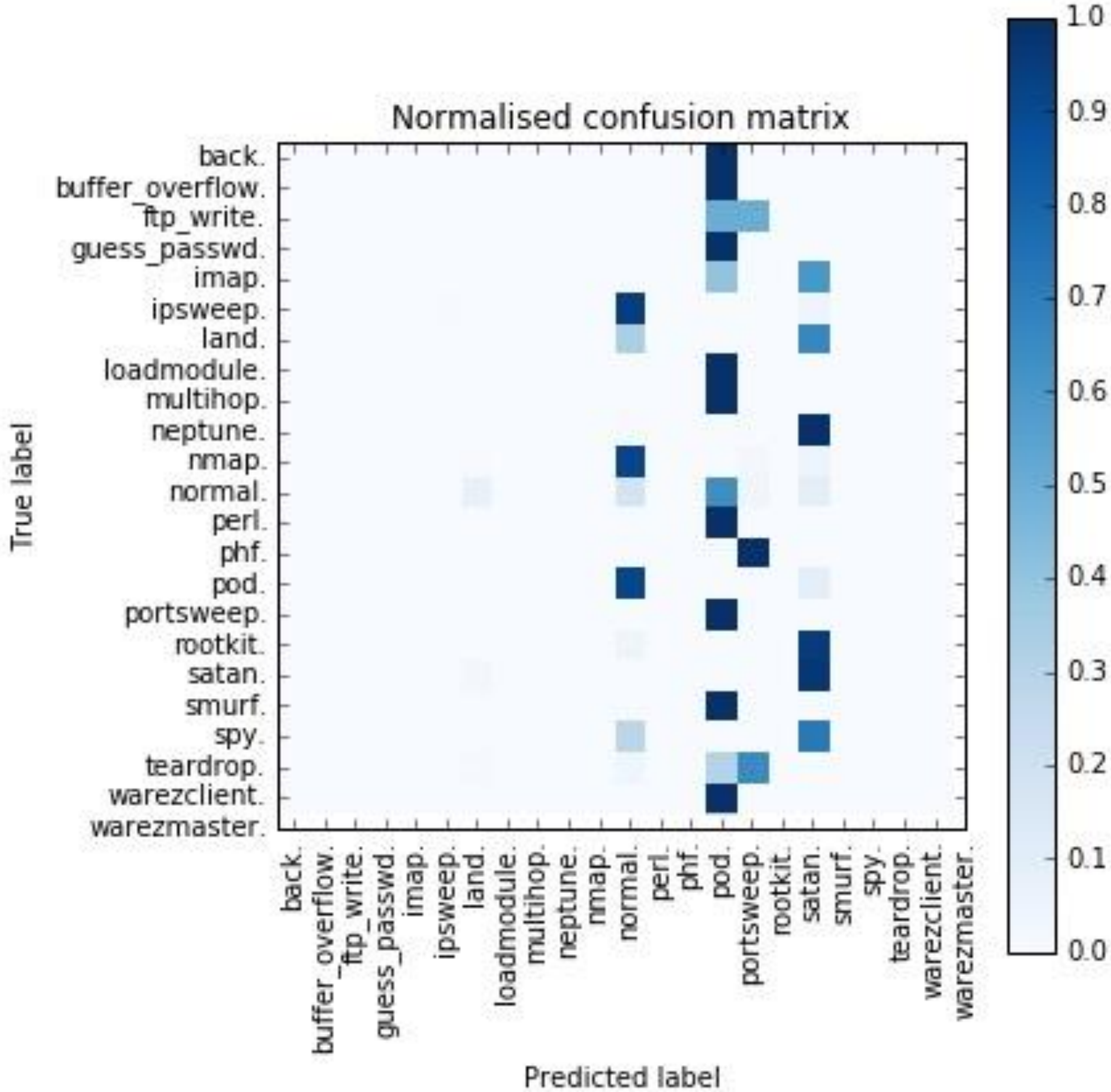
Step	Added	Dropped	Active set size	C
0	2		1	19960.733269
1	8		2	18696.7980058
2	3		3	9521.69759738
3	6		4	6645.07641798
4	1		5	2735.84447649
5	9		6	1866.54369652
6	4		7	1449.91074453
7	7		8	420.081823008
8	5		9	115.157274041
9	0		10	106.993857228

```
[(0.52639646470399315, 'LSTAT'),
 (0.27921428015177541, 'RM'),
 (0.054353831310065687, 'DIS'),
 (0.031820451224154722, 'CRIM'),
 (0.029793467094947356, 'NOX'),
 (0.021350472586185009, 'PTRATIO'),
 (0.015375071104791901, 'AGE'),
 (0.015233565046354791, 'TAX'),
 (0.01095820296701624, 'B'),
 (0.0075592385798185944, 'INDUS'),
 (0.0055375893522671962, 'RAD'),
 (0.0013486340199339781, 'ZN'),
 (0.0010587318586900362, 'CHAS')]
```

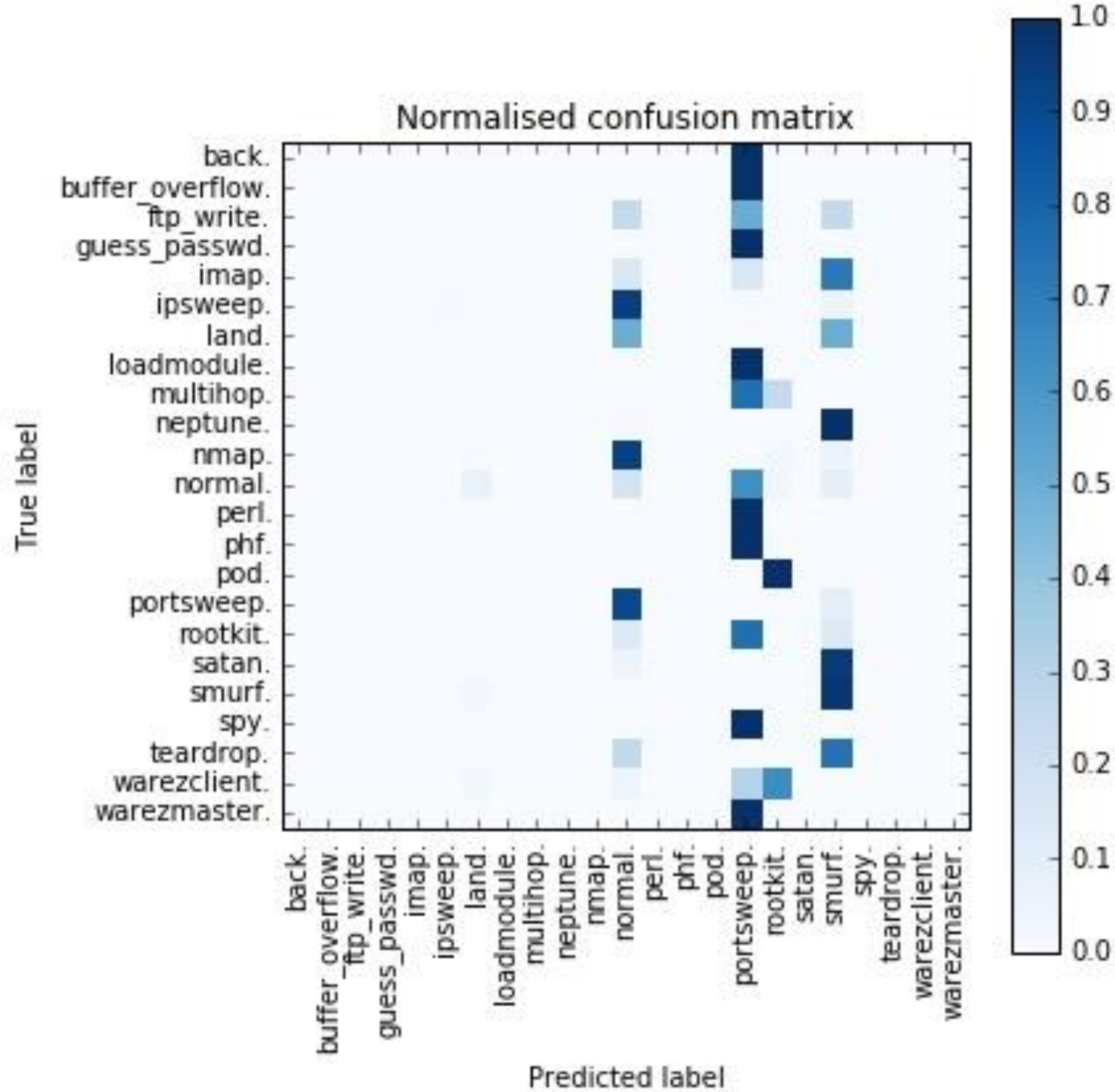
```
[(0.26442820639779868, 'LSTAT'),
 (0.21170609523931225, 'RM'),
 (0.11520512234965929, 'DIS'),
 (0.078532434845484278, 'TAX'),
 (0.075850985431776763, 'PTRATIO'),
 (0.075660468754102 9, 'NOX'),
 (0.052097327327291075, 'B'),
 (0.041177393920216847, 'CRIM'),
 (0.034255068725583829, 'AGE'),
 (0.023541808250096587, 'INDUS'),
 (0.012189199051061582, 'CHAS'),
 (0.011705380397086919, 'RAD'),
 (0.0036505093105288107, 'ZN')]
```

$$loss(x) = max(0.1 - l \cdot w \cdot x)$$

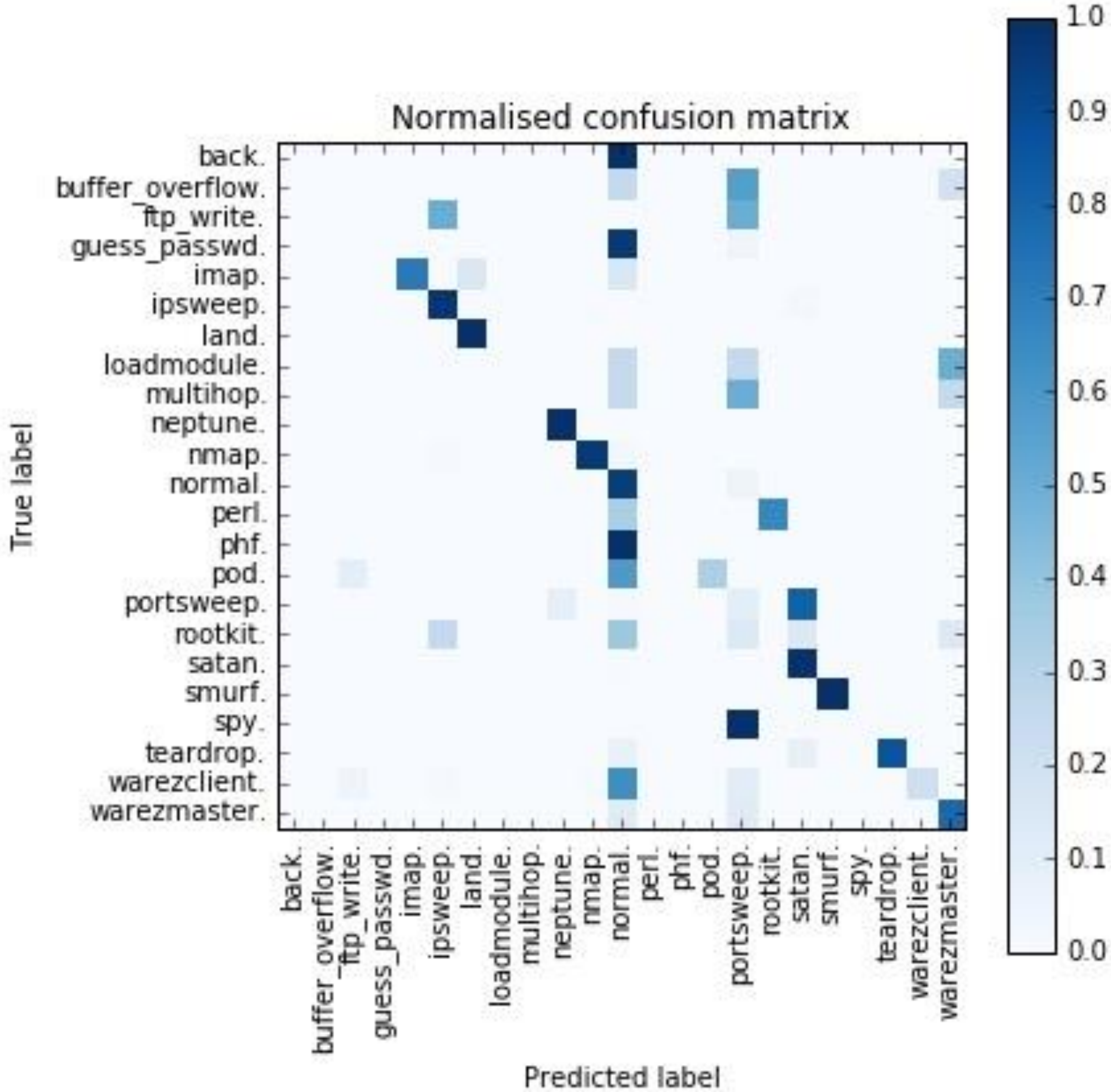

MAE train
MAE test
Num. features selected
7.2
7.0
6.8
6.6
6.4
7.2
7.1
7.0
6.9
6.8
6.7
6.6
0
500
1000
1500
2000



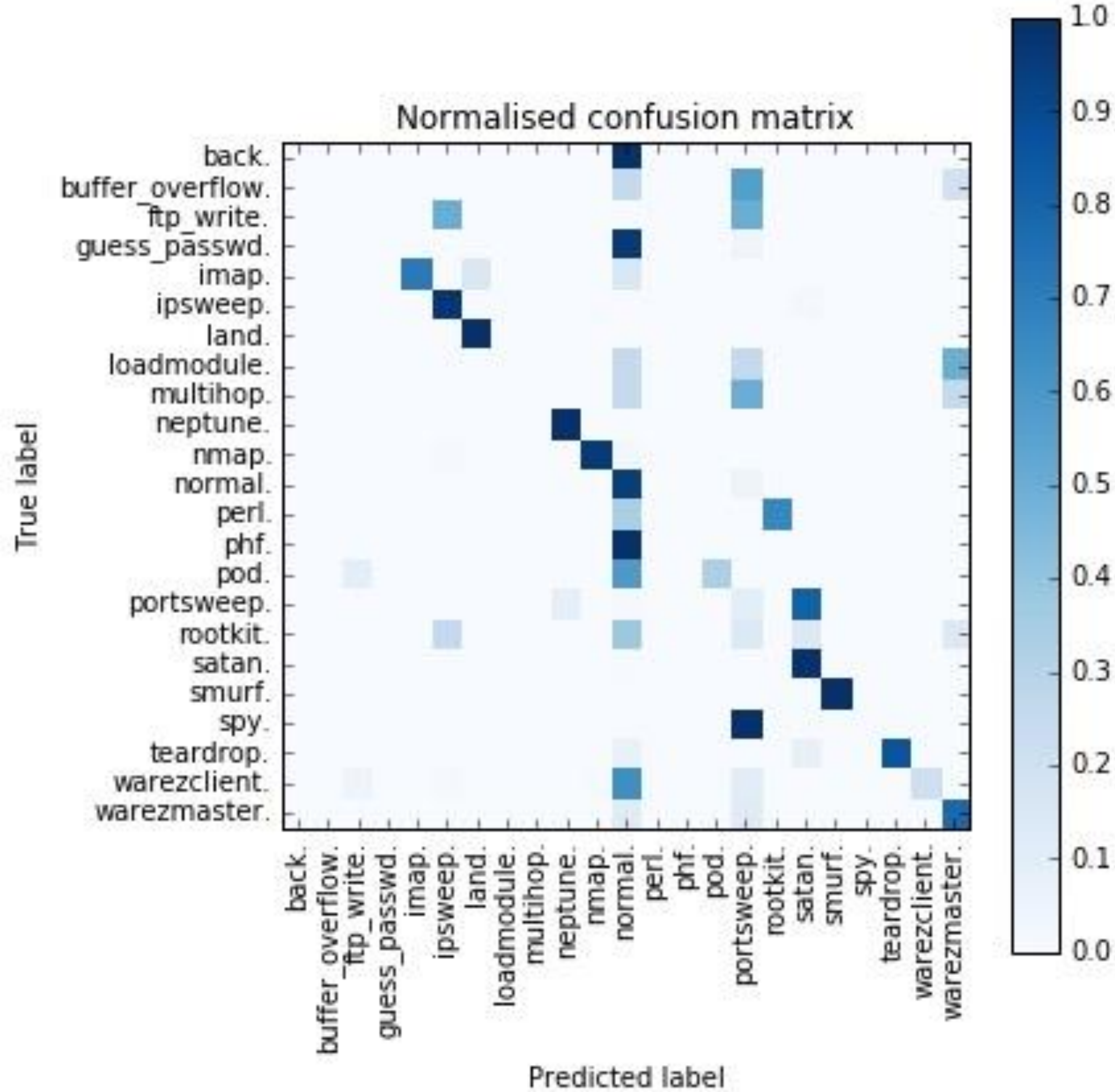
Normalised confusion matrix
True label
Predicted label
back.
buffer_overflow.
ftp_write.
guess_passwd.
imap.
ipsweep.
land.
loadmodule.
multihop.
neptune.
nmap.
normal.
perl.
phf.
pod.
portsweep.
rootkit.
satan.
smurf.
spy.
teardrop.
warezclient.
warezmaster.
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0



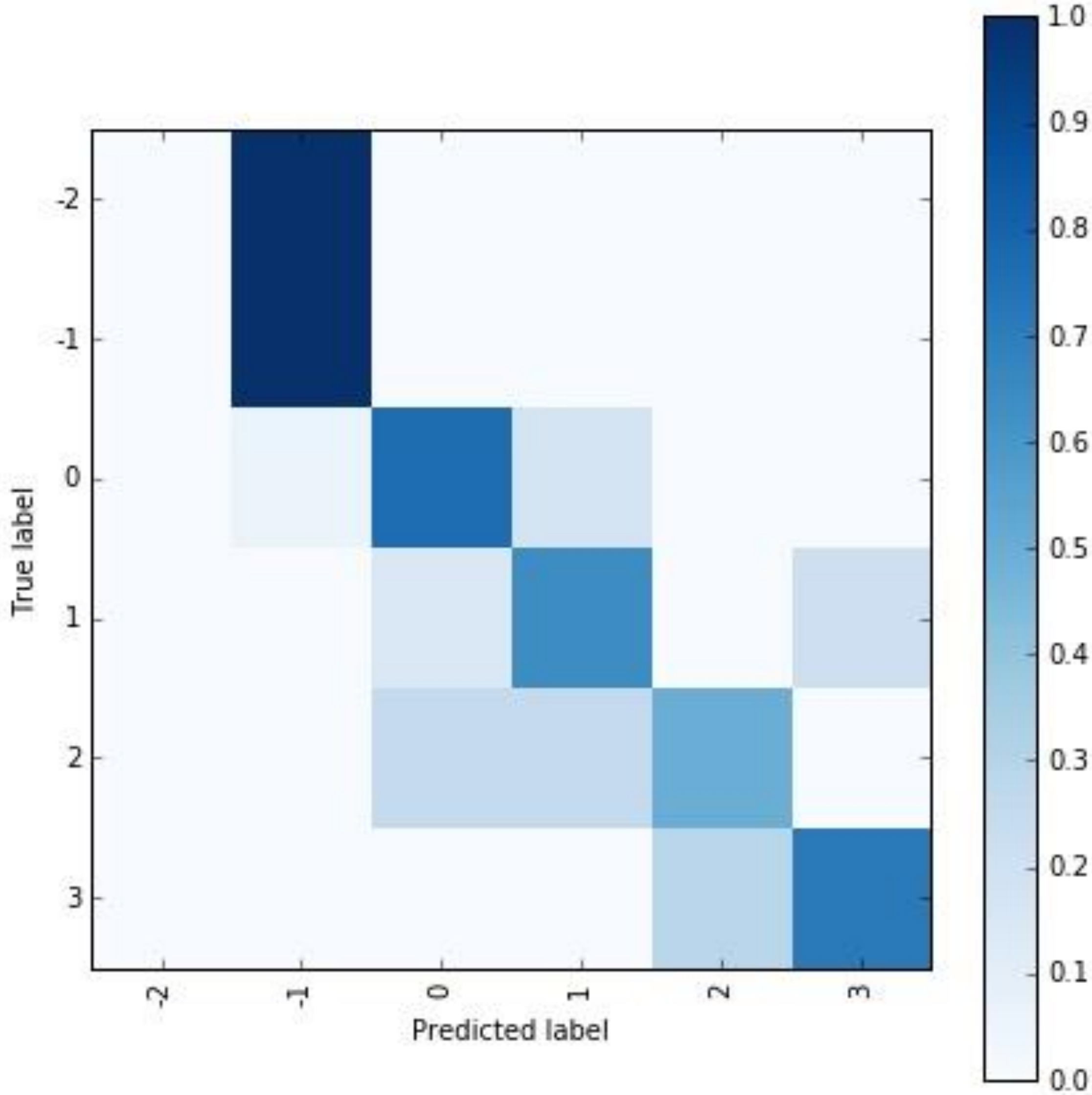
Normalised confusion matrix
True label
Predicted label
back.
buffer_overflow.
ftp_write.
guess_passwd.
imap.
ipsweep.
land.
loadmodule.
multihop.
neptune.
nmap.
normal.
perl.
phf.
pod.
portsweep.
rootkit.
satan.
smurf.
spy.
teardrop.
warezclient.
warezmaster.
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0



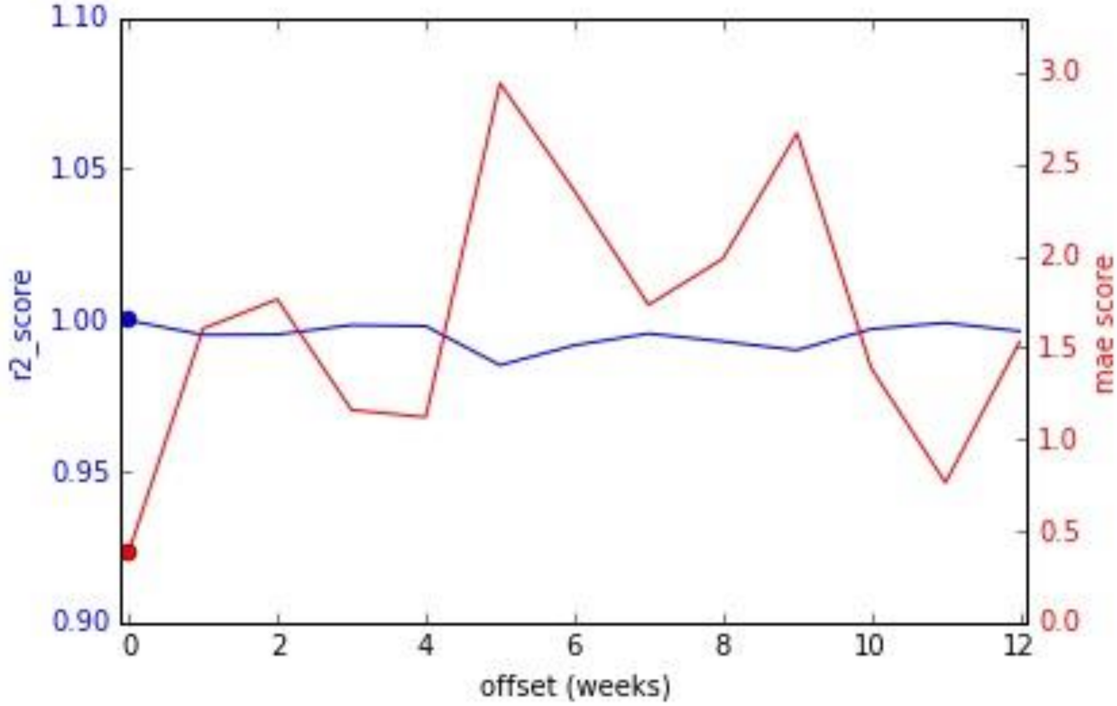
Normalised confusion matrix
True label
Predicted label
back.
buffer_overflow.
ftp_write.
guess_passwd.
imap.
ipsweep.
land.
loadmodule.
multihop.
neptune.
nmap.
normal.
perl.
phf.
pod.
portsweep.
rootkit.
satan.
smurf.
spy.
teardrop.
warezclient.
warezmaster.
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

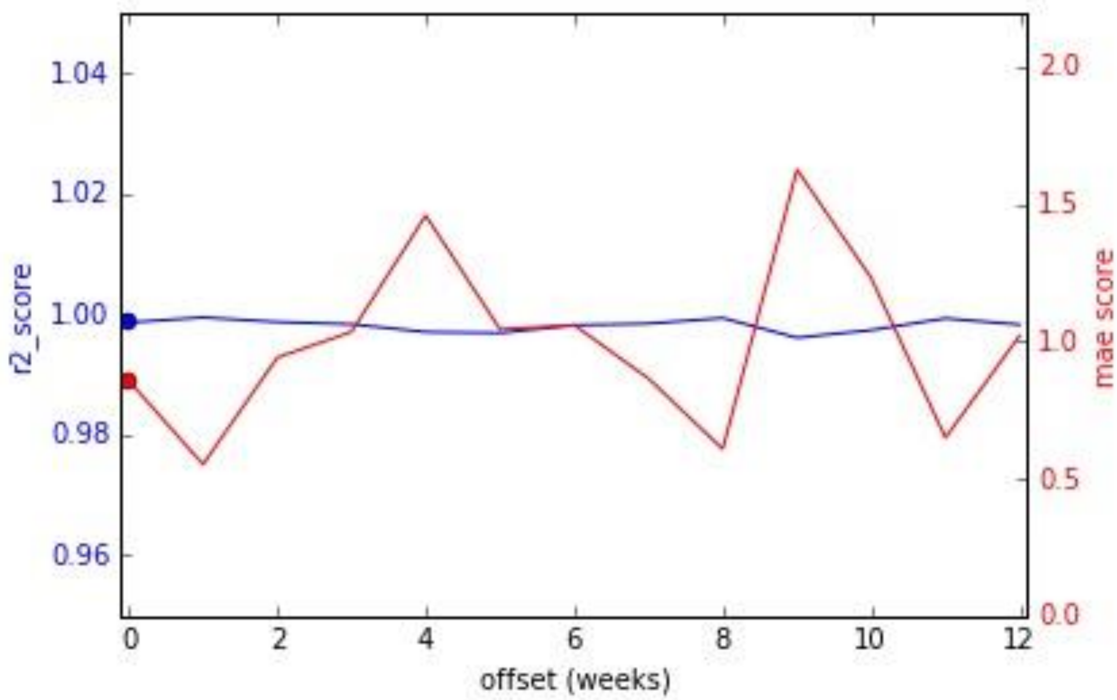


Normalised confusion matrix
True label
Predicted label
back.
buffer_overflow.
ftp_write.
guess_passwd.
imap.
ipsweep.
land.
loadmodule.
multihop.
neptune.
nmap.
normal.
perl.
phf.
pod.
portsweep.
rootkit.
satan.
smurf.
spy.
teardrop.
warezclient.
warezmaster.
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0



True label
-2
-1
0
1
2
3
Predicted label
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0


r2_score
mae score
offset (weeks)
1.10
1.05
1.00
0.95
0.90
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
10
12

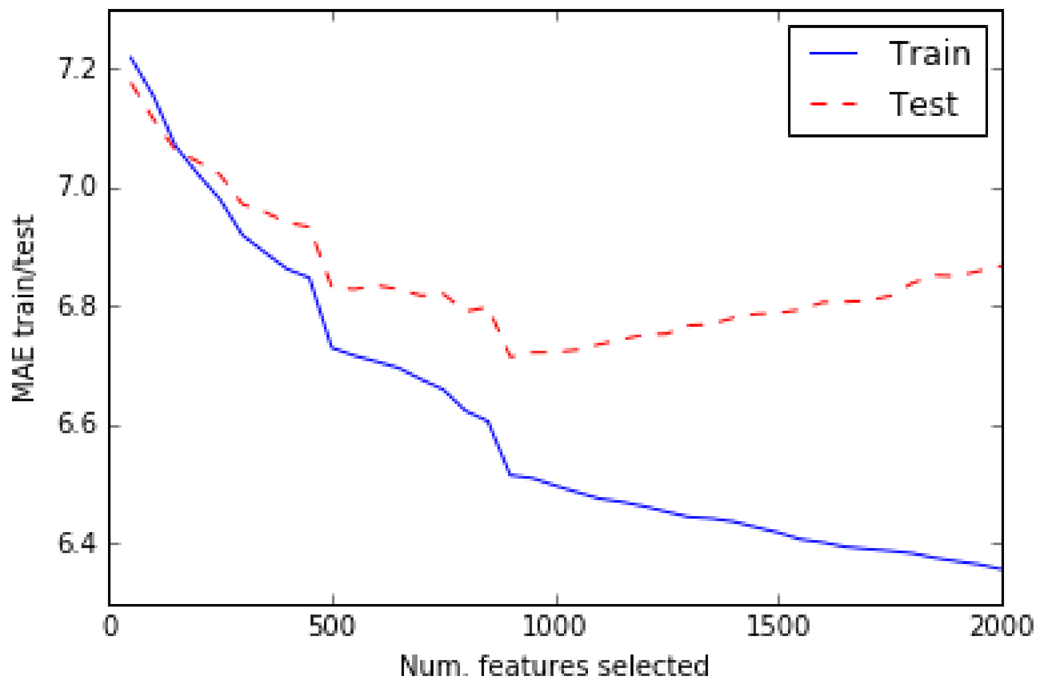

r2_score
mae score
offset (weeks)
1.04
1.02
1.00
0.98
0.96
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
10
12

```
quarter,stock,date,open,high,low,close,volume,percent_change_price,p
ercent_change_volume_over_last_wk,previous_weeks_volume,next_weeks_o
pen,next_weeks_close,percent_change_next_weeks_price,days_to_next_di
vidend,percent_return_next_dividend
1,AA,1/7/2011,$15.82,$16.72,$15.78,$16.42,239655616,3.79267,,,$16.71
,$15.97,-4.42849,26,0.182704
```

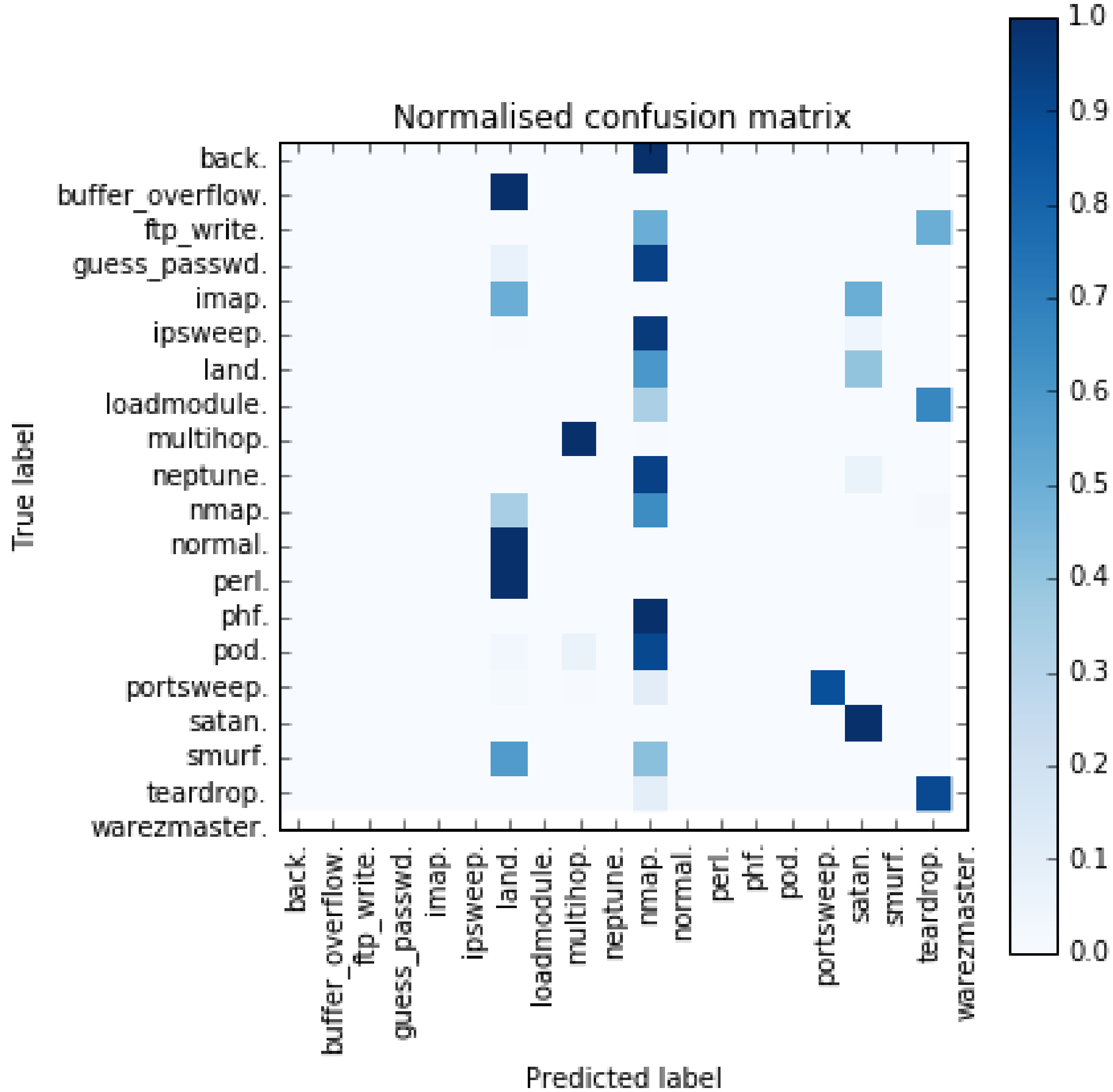
```
2001,49.94357,21.47114,73.07750,8.74861,-17.40628,-13.09905,-25.0120
2,-12.23257,7.83089,-2.46783,3.32136,-2.31521,10.20556,611.10913,951
.08960,698.11428,408.98485,383.70912,326.51512,238.11327,251.42414,1
87.17351,100.42652,179.19498,-8.41558,-317.87038,95.86266,48.10259,-
95.66303,-18.06215,1.96984,34.42438,11.72670,1.36790,7.79444,-0.3699
4,-133.67852,-83.26165,-37.29765,73.04667,-37.36684,-3.13853,-24.215
31,-13.23066,15.93809,-18.60478,82.15479,240.57980,-10.29407,31.5843
1,-25.38187,-3.90772,13.29258,41.55060,-7.26272,-21.00863,105.50848,
64.29856,26.08481,-44.59110,-8.30657,7.93706,-10.73660,-95.44766,-82
.03307,-35.59194,4.69525,70.95626,28.09139,6.02015,-37.13767,-41.124
50,-8.40816,7.19877,-8.60176,-5.90857,-12.32437,14.68734,-54.32125,4
0.14786,13.01620,-54.40548,58.99367,15.37344,1.11144,-23.08793,68.40
795,-1.82223,-27.46348,2.26327
2001,48.73215,18.42930,70.32679,12.94636,-10.32437,-24.83777,8.76630
,-0.92019,18.76548,4.59210,2.21920,0.34006,44.38997,2056.93836,605.4
0696,457.41175,777.15347,415.64880,746.47775,366.45320,317.82946,273
.07917,141.75921,317.35269,19.48271,-65.25496,162.75145,135.00765,-9
6.28436,-86.87955,17.38087,45.90742,32.49908,-32.85429,45.10830,26.8
4939,-302.57328,-41.71932,-138.85034,202.18689,-33.44277,195.04749,-
16.93235,-1.09168,-25.38061,-12.19034,-125.94783,121.74212,136.67075
,41.18157,28.55107,1.52298,70.99515,-43.63073,-42.55014,129.82848,79
.95420,-87.14554,-45.75446,-65.82100,-43.90031,-19.45705,12.59163,-4
07.64130,42.91189,12.15850,-88.37882,42.25246,46.49209,-30.17747,45.
98495,130.47892,13.88281,-4.00055,17.85965,-18.32138,-87.99109,14.37
524,-22.70119,-58.81266,5.66812,-19.68073,33.04964,42.87836,-9.90378
,-32.22788,70.49388,12.04941,58.43453,26.92061
```

```
0,tcp,http,SF,215,45076,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,0.00,0.0
0,0.00,0.00,1.00,0.00,0.00,0,0,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.
00,normal.
0,tcp,http,SF,162,4528,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,2,2,0.00,0.00
,0.00,0.00,1.00,0.00,0.00,1,1,1.00,0.00,1.00,0.00,0.00,0.00,0.00,0.0
0,normal.
```

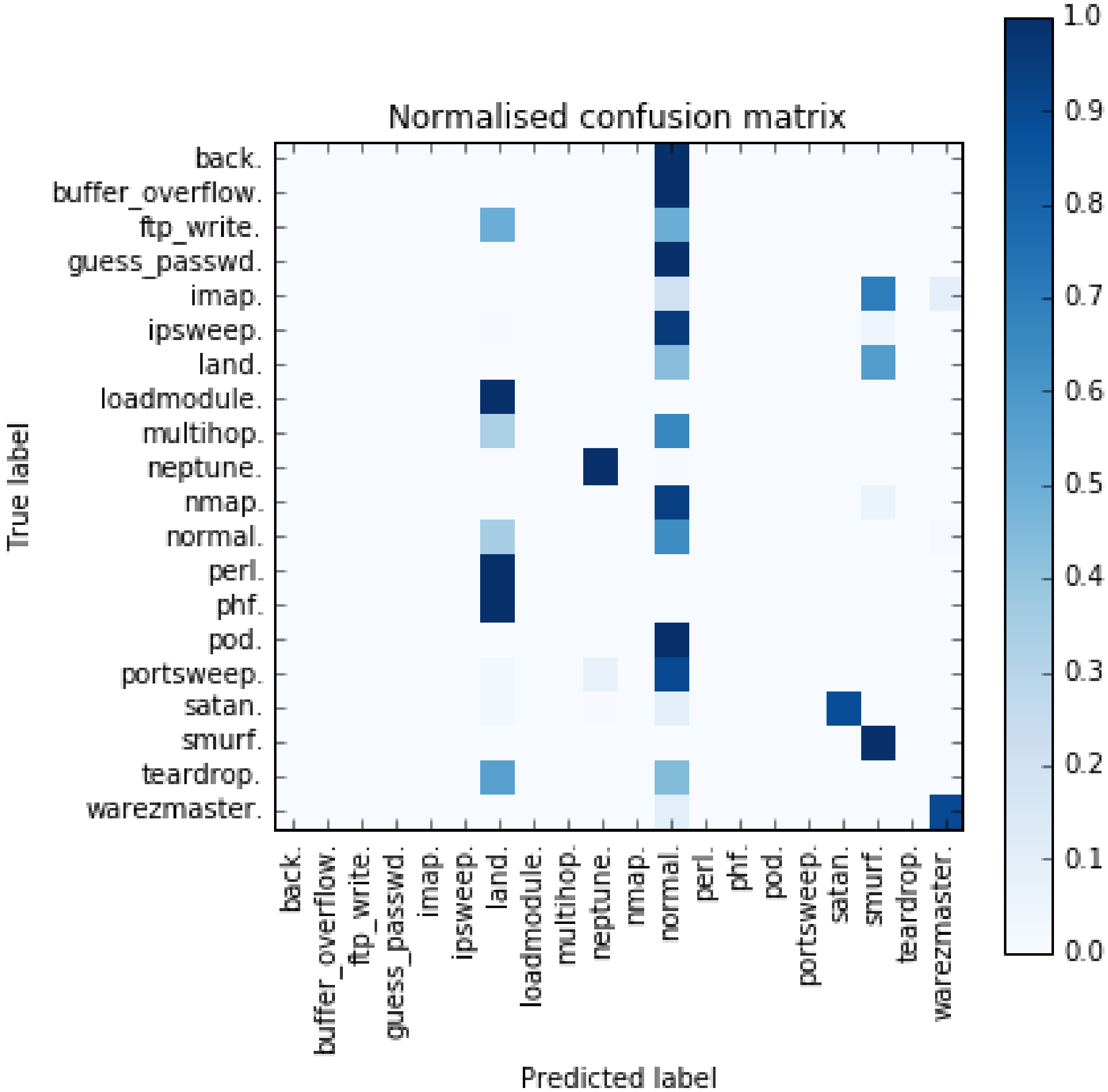
```
3,?,alfa-romero,gas,std,two,convertible,rwd,front,88.60,168.80,64.10
,48.80,2548,dohc,four,130,mpfi,3.47,2.68,9.00,111,5000,21,27,13495
3,?,alfa-romero,gas,std,two,convertible,rwd,front,88.60,168.80,64.10
,48.80,2548,dohc,four,130,mpfi,3.47,2.68,9.00,111,5000,21,27,16500
```

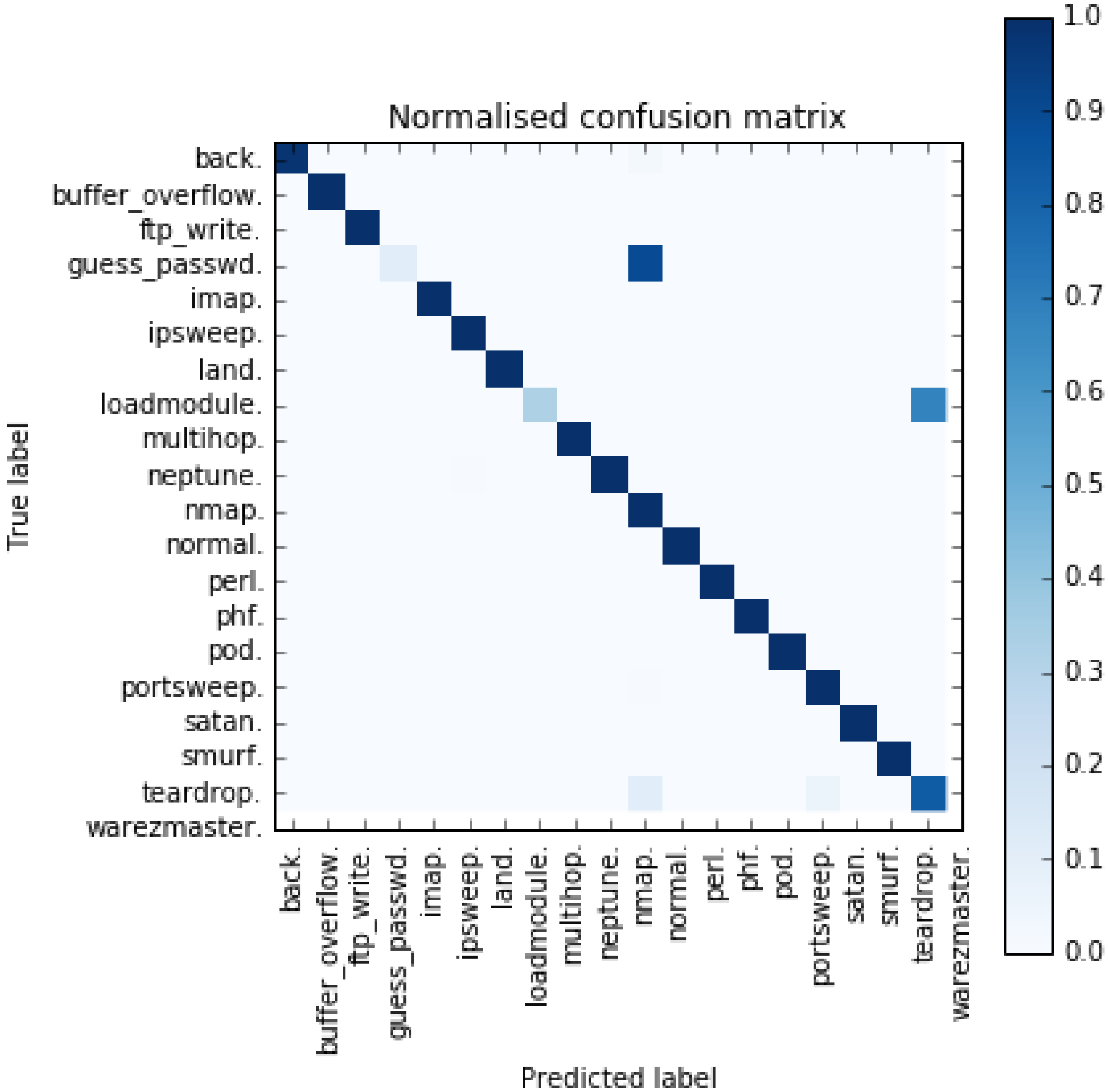


Train
Test
7.2
7.0
6.8
6.6
6.4
MAE train/test
0
500
1000
1500
2000
Num. features selected

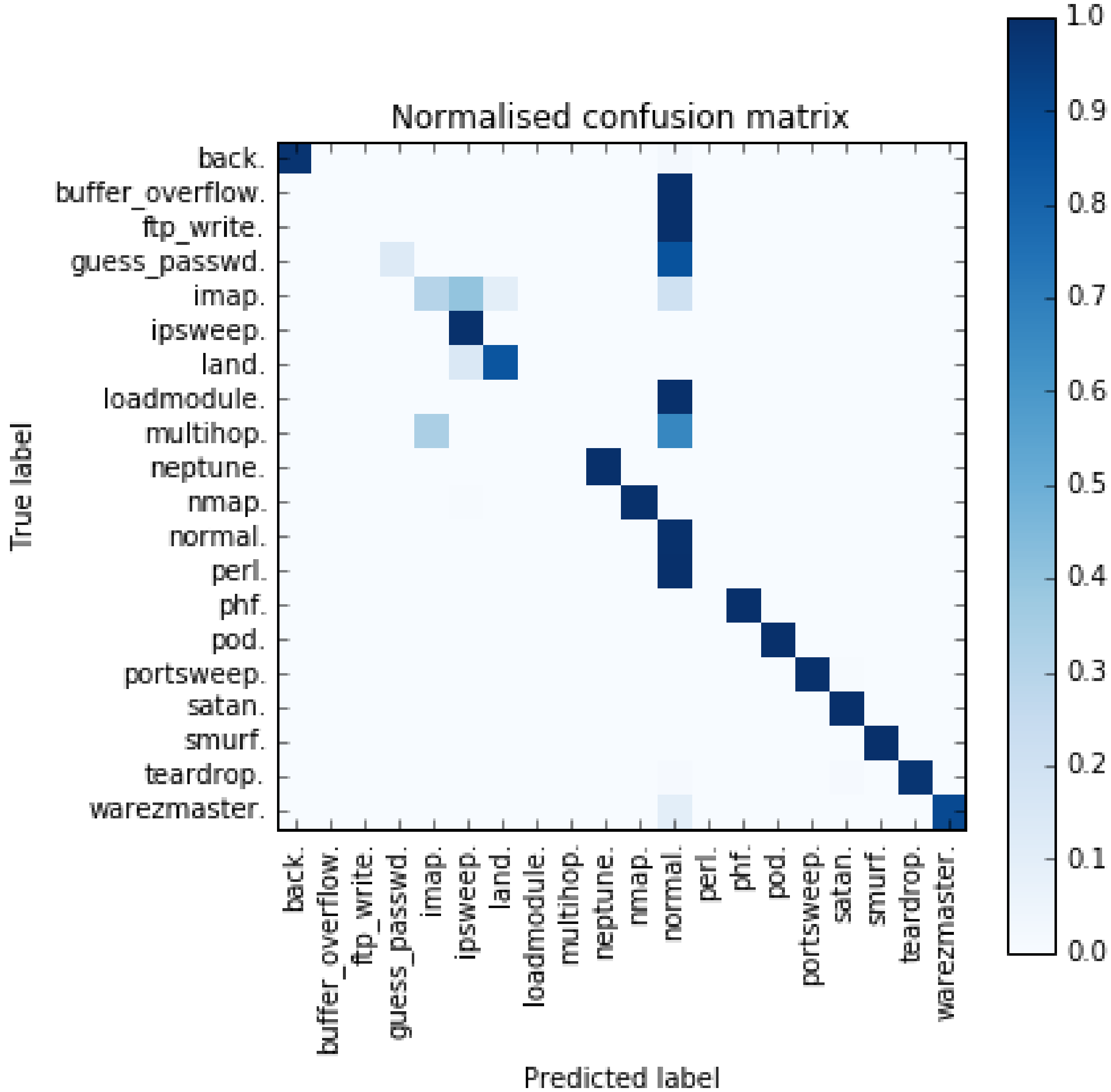


Normalised confusion matrix
True label
Predicted label
back.
buffer_overflow.
ftp_write.
guess_passwd.
imap.
ipsweep.
land.
loadmodule.
multihop.
neptune.
nmap.
normal.
perl.
phf.
pod.
portsweep.
satan.
smurf.
teardrop.
warezmaster.
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0


Normalised confusion matrix
True label
Predicted label
back.
buffer_overflow.
ftp_write.
guess_passwd.
imap.
ipsweep.
land.
loadmodule.
multihop.
neptune.
nmap.
normal.
perl.
phf.
pod.
portsweep.
satan.
smurf.
teardrop.
warezmaster.
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0



Normalised confusion matrix
True label
Predicted label
back.
buffer_overflow.
ftp_write.
guess_passwd.
imap.
ipsweep.
land.
loadmodule.
multihop.
neptune.
nmap.
normal.
perl.
phf.
pod.
portsweep.
satan.
smurf.
teardrop.
warezmaster.
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0



Normalised confusion matrix
True label
Predicted label
back.
buffer_overflow.
ftp_write.
guess_passwd.
imap.
ipsweep.
land.
loadmodule.
multihop.
neptune.
nmap.
normal.
perl.
phf.
pod.
portsweep.
satan.
smurf.
teardrop.
warezmaster.
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0