1

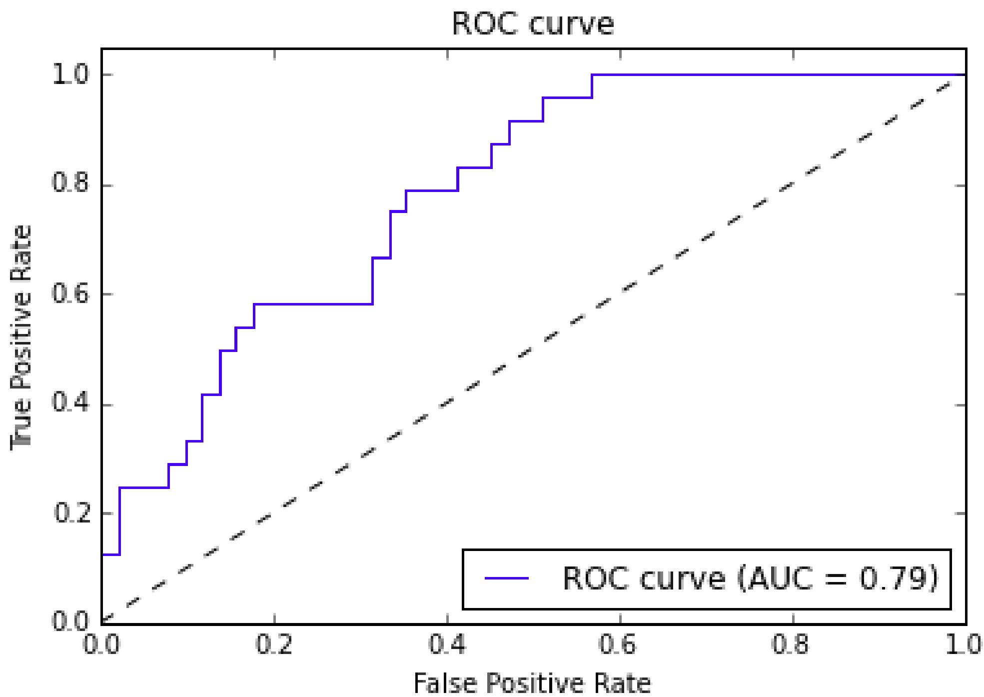
Introducing Test-driven Machine Learning

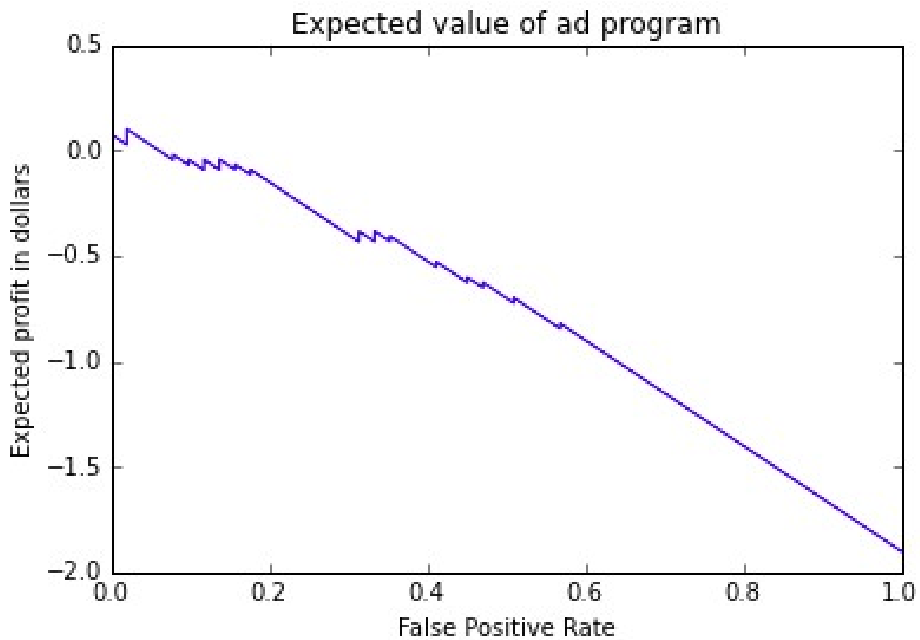
Our first test

```
E
======================================================================
ERROR: number_guesser_tests.given_no_information_when_asked_to_guess_t
est
----------------------------------------------------------------------
Traceback (most recent call last):
  File "/Users/justin/Envs/default/lib/python2.7/site-packages/nose/ca
se.py", line 197, in runTest
    self.test(*self.arg)
  File "/Users/justin/Documents/Code/Machine-Learning-Test-by-Test/Cha
pter 1/number_guesser_tests.py", line 2, in given_no_information_when_
asked_to_guess_test
    number_guesser = NumberGuesser()
NameError: global name 'NumberGuesser' is not defined

----------------------------------------------------------------------
Ran 1 test in 0.002s

FAILED (errors=1)
```

```
.
----------------------------------------------------------------------
Ran 1 test in 0.002s

OK
```

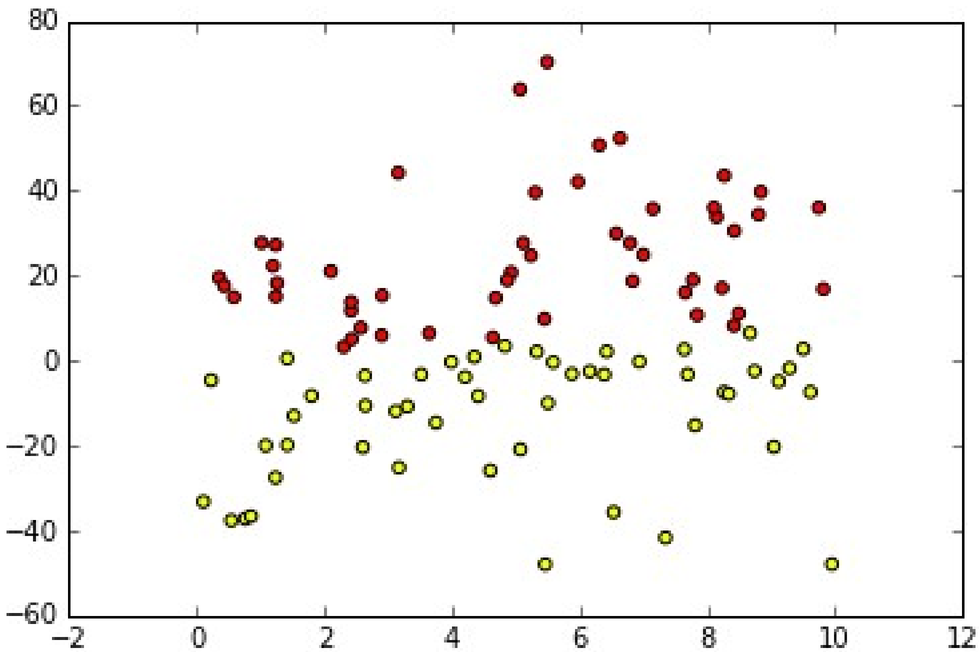
Quantifying the classification models




Expected value of ad program
Expected profit in dollars
False Positive Rate
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.2
0.4
0.6
0.8
1.0

2
Perceptively Testing a Perceptron

Getting started



	Training rate		0.25					
Iteration	Training 1	Training 2	Training label	Weights 1	Weights 2	Weight 1 update	Weight 2 update	Predicted value
1	1	1	1	0.96897982	0.97171753	0.968979818	0.971717531	1
1	1	0	1	0.96897982	0.97171753	0.968979818	0.971717531	1
1	0	1	1	0.96897982	0.97171753	0.968979818	0.971717531	1
1	0	0	0	0.96897982	0.97171753	0.968979818	0.971717531	0

$$w_{i+1,j} = w_{i,j} + \eta * w_{i,j} * (t_j - p_j)$$

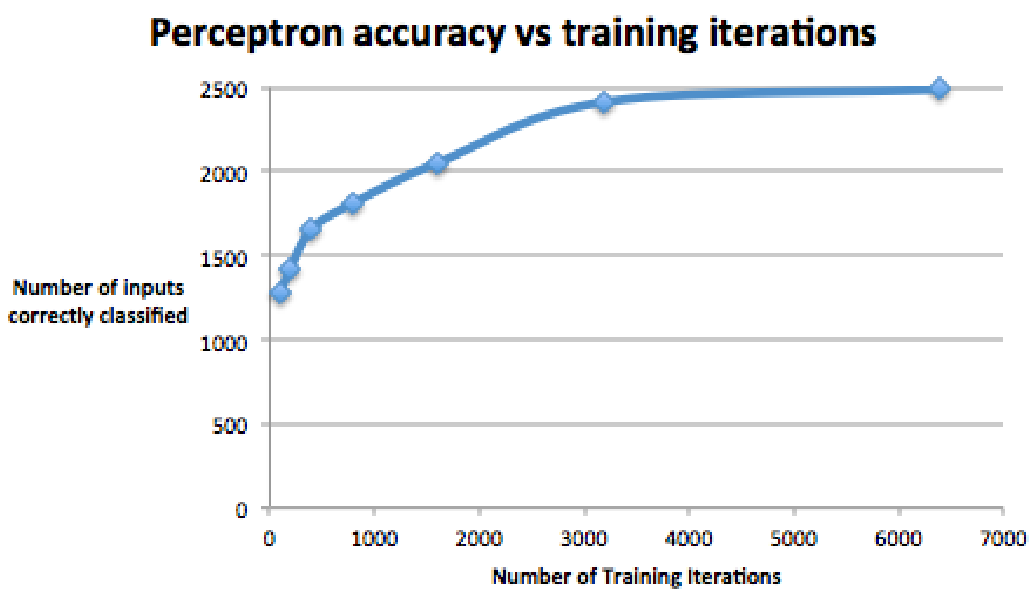
	A	B	C	D	E	F	G	H	I
1		Training rate		0.25					
2									
3	Iteration	Training 1	Training 2	Training label	Weights 1	Weights 2	Weight 1 update	Weight 2 update	Predicted value
4	1	1	1	1	0.96897982	0.97171753	=E4+D1*($D4-$I4)*B4		1

$$p_j = \sum w_i * x_i > 0$$

	A	B	C	D	E	F	G	H	I	J
1		Training rate		0.25						
2										
3	Iteration	Training 1	Training 2	Training label	Weights 1	Weights 2	Weight 1 update	Weight 2 update	Predicted value	
4	1	1	1	1	0.96897982	0.97171753	0.968979818	0.971717531	=IF((B4*E4+C4*F4)>0,1,0)	
5	1	1	0	1	0.96897982	0.97171753	0.968979818	0.971717531	1	
6	1	0	1	1	0.96897982	0.97171753	0.968979818	0.971717531	1	
7	1	0	0	0	0.96897982	0.97171753	0.968979818	0.971717531	0	

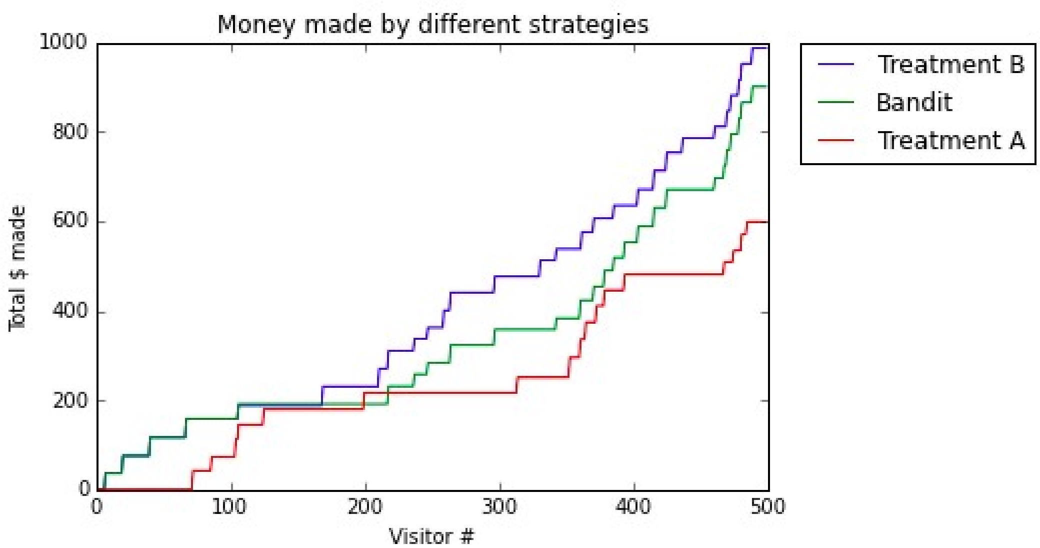
	A	B	C	D	E	F	G	H	I
1		Training rate		0.1					
2									
3	Iteration	Training 1	Training 2	Training label	Weights 1	Weights 2	Weight 1 update	Weight 2 update	Predicted value
4	1	5	-1	1	0.431	0.02	0.431	0.02	1
5	1	2	-1	0	0.431	0.02	0.231	0.12	1
6	1	0	-1	0	0.231	0.12	0.231	0.12	0
7	1	-2	-1	0	0.231	0.12	0.231	0.12	0
8	2	5	-1	1	0.231	0.12	0.231	0.12	1
9	2	2	-1	0	0.231	0.12	0.031	0.22	1
10	2	0	-1	0	0.031	0.22	0.031	0.22	0
11	2	-2	-1	0	0.031	0.22	0.031	0.22	0
12	3	5	-1	1	0.031	0.22	0.531	0.12	0
13	3	2	-1	0	0.531	0.12	0.331	0.22	1
14	3	0	-1	0	0.331	0.22	0.331	0.22	0
15	3	-2	-1	0	0.331	0.22	0.331	0.22	0
16	4	5	-1	1	0.331	0.22	0.331	0.22	1
17	4	2	-1	0	0.331	0.22	0.131	0.32	1
18	4	0	-1	0	0.131	0.32	0.131	0.32	0
19	4	-2	-1	0	0.131	0.32	0.131	0.32	0
20	5	5	-1	1	0.131	0.32	0.131	0.32	1
21	5	2	-1	0	0.131	0.32	0.131	0.32	0
22	5	0	-1	0	0.131	0.32	0.131	0.32	0
23	5	-2	-1	0	0.131	0.32	0.131	0.32	0

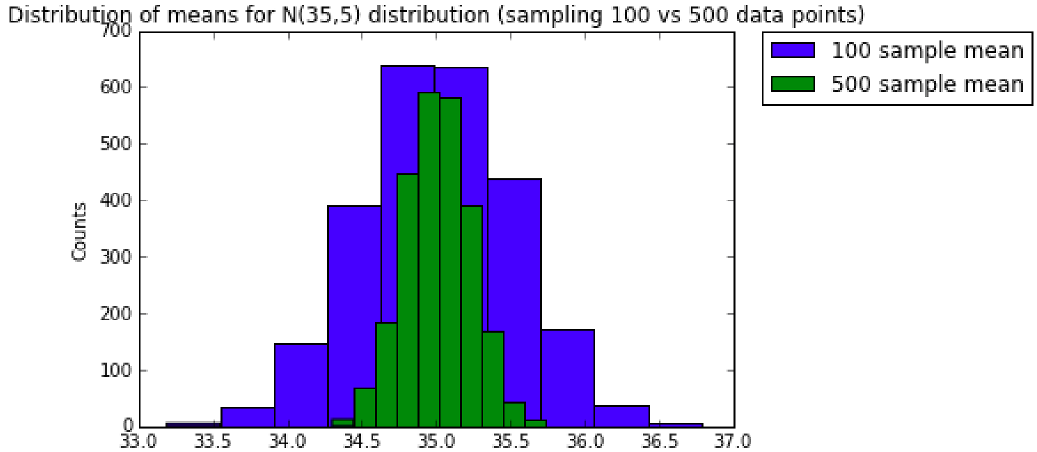
```
======================================================================
FAIL: tests.detect_a_complicated_example_test
----------------------------------------------------------------------
Traceback (most recent call last):
  File "/Library/Python/2.7/site-packages/nose-1.3.0-py2.7.egg/nose/ca
se.py", line 197, in runTest
    self.test(*self.arg)
  File "/Users/justin/Documents/Code/test-driven-machine-learning/Chap
ter 2 Redux/tests.py", line 75, in detect_a_complicated_example_test
    "Perceptron should be much better than random. {0} correct".format
(correctly_classified)
AssertionError: Perceptron should be much better than random. 1367 cor
rect

----------------------------------------------------------------------
Ran 5 tests in 0.082s

FAILED (failures=1)
```

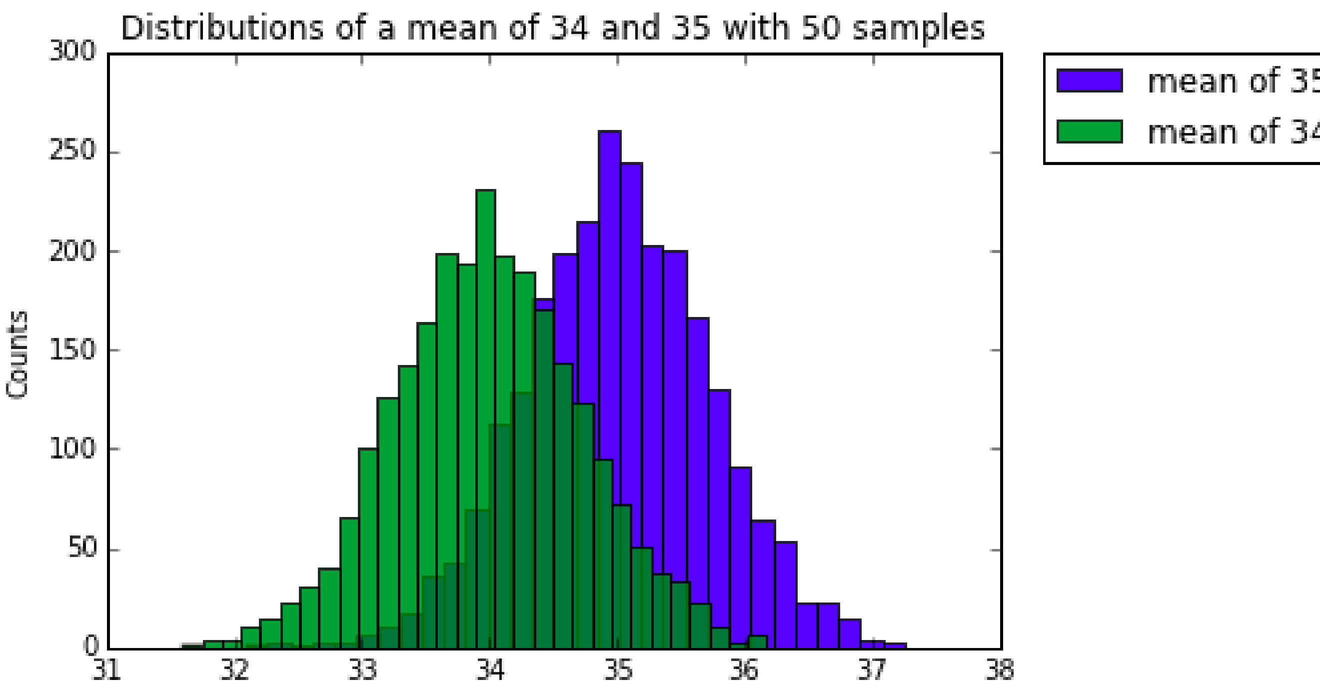


3
Exploring the Unknown with Multi-armed Bandits

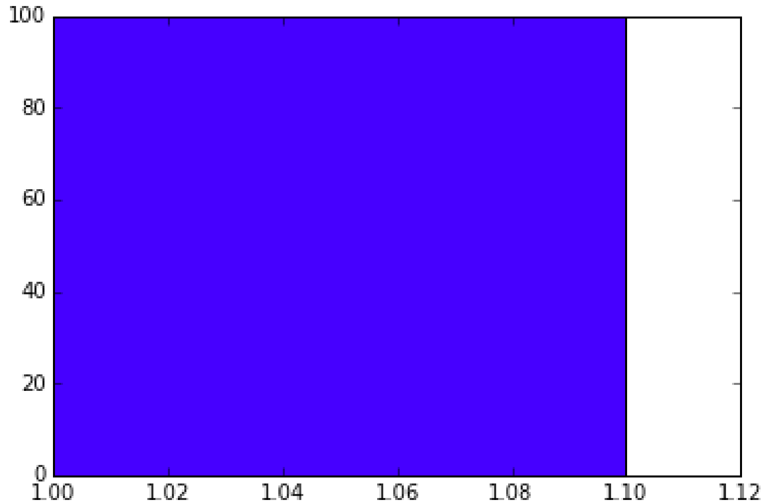
Simulating real world situations

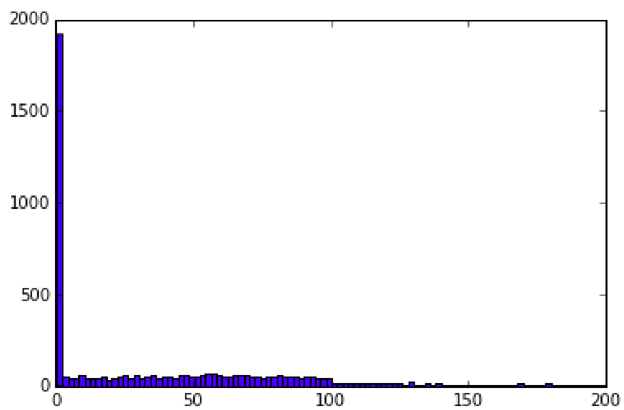



Distribution of means for N(35,5) distribution (sampling 100 vs 500 data points)
100 sample mean
500 sample mean
Counts
0
100
200
300
400
500
600
700
33.0
33.5
34.0
34.5
35.0
35.5
36.0
36.5
37.0


Distributions of a mean of 34 and 35 with 50 samples
mean of 35
mean of 34
Counts
0
50
100
150
200
250
300
31
32
33
34
35
36
37
38

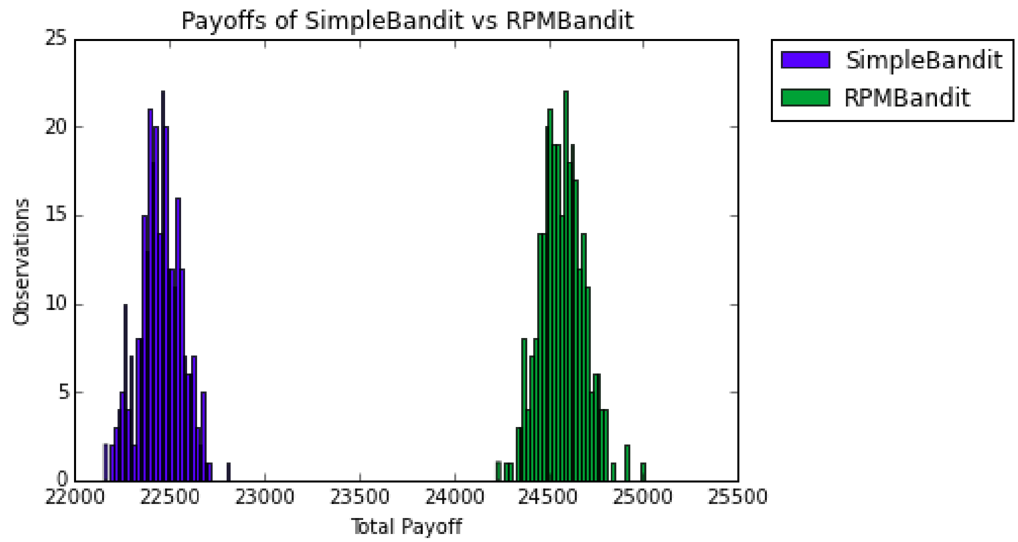
The problem with straight bootstrapping





Multi-armed armed bandit throw down

```
SimpleBandit: 22443.9230864
RPMBandit: 24567.9045708
''
```



4

Predicting Values with

Building the foundations of our model

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          dependent_var   R-squared:                       0.005
Model:                            OLS   Adj. R-squared:                 -0.030
Method:                 Least Squares   F-statistic:                    0.1499
Date:                Thu, 19 Feb 2015   Prob (F-statistic):              0.702
Time:                        20:57:08   Log-Likelihood:                -222.90
No. Observations:                  30   AIC:                             449.8
Df Residuals:                      28   BIC:                             452.6
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept    224.5840    520.609      0.431      0.669      -841.836  1291.004
ind_var_d     -1.1946      3.085     -0.387      0.702        -7.515     5.125
==============================================================================
Omnibus:                        0.043   Durbin-Watson:                   2.057
Prob(Omnibus):                  0.979   Jarque-Bera (JB):                0.058
Skew:                          -0.017   Prob(JB):                        0.971
Kurtosis:                       2.787   Cond. No.                     1.14e+03
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          dependent_var   R-squared:                       0.182
Model:                            OLS   Adj. R-squared:                  0.152
Method:                 Least Squares   F-statistic:                     6.215
Date:                Thu, 19 Feb 2015   Prob (F-statistic):             0.0188
Time:                        21:16:15   Log-Likelihood:                -219.98
No. Observations:                  30   AIC:                             444.0
Df Residuals:                      28   BIC:                             446.8
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept     33.4182     70.006      0.477      0.637      -109.982   176.819
ind_var_a      3.0475      1.222      2.493      0.019         0.544     5.551
==============================================================================
Omnibus:                        0.175   Durbin-Watson:                   1.811
Prob(Omnibus):                  0.916   Jarque-Bera (JB):                0.023
Skew:                          -0.053   Prob(JB):                        0.988
Kurtosis:                       2.915   Cond. No.                         57.3
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          dependent_var   R-squared:                       0.818
Model:                            OLS   Adj. R-squared:                  0.804
Method:                 Least Squares   F-statistic:                     60.62
Date:                Thu, 19 Feb 2015   Prob (F-statistic):           1.04e-10
Time:                        21:19:56   Log-Likelihood:                -197.44
No. Observations:                  30   AIC:                             400.9
Df Residuals:                      27   BIC:                             405.1
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept     94.5490     34.216      2.763      0.010        24.344   164.754
ind_var_a      2.7750      0.588      4.720      0.000         1.569     3.981
ind_var_b    115.1101     11.853      9.712      0.000        90.791   139.430
==============================================================================
Omnibus:                        0.248   Durbin-Watson:                   2.031
Prob(Omnibus):                  0.883   Jarque-Bera (JB):                0.442
Skew:                          -0.101   Prob(JB):                        0.802
Kurtosis:                       2.441   Cond. No.                         58.5
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          dependent_var   R-squared:                       0.845
Model:                            OLS   Adj. R-squared:                  0.820
Method:                 Least Squares   F-statistic:                     34.12
Date:                Thu, 19 Feb 2015   Prob (F-statistic):           8.64e-10
Time:                        21:33:11   Log-Likelihood:                -195.00
No. Observations:                  30   AIC:                             400.0
Df Residuals:                      25   BIC:                             407.0
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept    286.7307    225.631      1.271      0.216      -177.965   751.427
ind_var_a      2.5552      0.574      4.453      0.000         1.373     3.737
ind_var_b    112.2556     11.438      9.814      0.000        88.698   135.813
ind_var_c     -6.4966      3.134     -2.073      0.049       -12.951    -0.042
ind_var_d     -0.4223      1.293     -0.327      0.747        -3.085     2.240
==============================================================================
Omnibus:                        0.100   Durbin-Watson:                   2.208
Prob(Omnibus):                  0.951   Jarque-Bera (JB):                0.030
Skew:                           0.028   Prob(JB):                        0.985
Kurtosis:                       2.856   Cond. No.                     1.19e+03
==============================================================================

.
----------------------------------------------------------------------
Ran 1 test in 0.414s

OK
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          dependent_var   R-squared:                       0.987
Model:                            OLS   Adj. R-squared:                  0.984
Method:                 Least Squares   F-statistic:                     356.7
Date:                Thu, 19 Feb 2015   Prob (F-statistic):           1.07e-21
Time:                        22:02:03   Log-Likelihood:                -158.16
No. Observations:                  30   AIC:                             328.3
Df Residuals:                      24   BIC:                             336.7
Df Model:                           5
Covariance Type:            nonrobust
=======================================================================================
                          coef    std err          t      P>|t|      [95.0% Conf. Int.]
---------------------------------------------------------------------------------------
Intercept              25.6266     24.999      1.025      0.316      -25.968    77.221
ind_var_a               2.7083      0.171     15.820      0.000        2.355     3.062
ind_var_b              -1.5527      8.798     -0.176      0.861      -19.712    16.606
ind_var_c              -0.3917      1.036     -0.378      0.709       -2.529     1.746
ind_var_e              -0.2006      0.032     -6.231      0.000       -0.267    -0.134
ind_var_b:ind_var_c     5.6450      0.371     15.225      0.000        4.880     6.410
==============================================================================
Omnibus:                        0.697   Durbin-Watson:                   2.070
Prob(Omnibus):                  0.706   Jarque-Bera (JB):                0.584
Skew:                          -0.318   Prob(JB):                        0.747
Kurtosis:                       2.750   Cond. No.                     1.48e+03
==============================================================================
```

5

Making Decisions Black and White with Logistic Regression

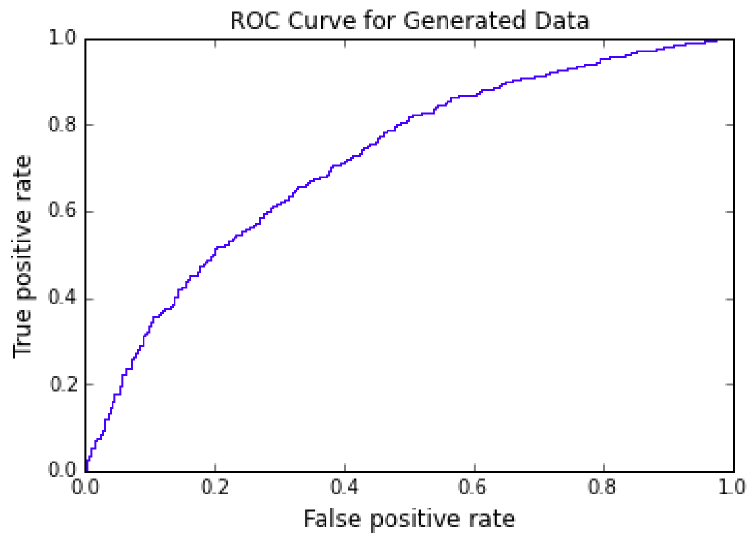
Generating logistic data

Logit Regression Results

Dep. Variable:	y	**No. Observations:**	1000
Model:	Logit	**Df Residuals:**	998
Method:	MLE	**Df Model:**	1
Date:	Sun, 01 Mar 2015	**Pseudo R-squ.:**	0.1476
Time:	16:02:11	**Log-Likelihood:**	-589.61
converged:	True	**LL-Null:**	-691.69
		LLR p-value:	2.598e-46

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
Intercept	-1.8773	0.156	-12.055	0.000	-2.182 -1.572
x	0.0349	0.003	12.897	0.000	0.030 0.040

Measuring model accuracy



Generating a more complex example

Test driving our model

```
F
======================================================================
FAIL: logistic_regression_tests.logistic_regression_test
----------------------------------------------------------------------
Traceback (most recent call last):
  File "/Library/Python/2.7/site-packages/nose-1.3.0-py2.7.egg/nose/ca
se.py", line 197, in runTest
    self.test(*self.arg)
  File "/Users/justin/Documents/Code/Machine-Learning-Test-by-Test/Cha
pter 5/logistic_regression_tests.py", line 13, in logistic_regression_
test
    assert auc > .6, 'AUC should be significantly above random'
AssertionError: AUC should be significantly above random
-------------------- >> begin captured stdout << ---------------------
Optimization terminated successfully.
         Current function value: 0.426086
         Iterations 6
AUC score: 0.510791645978

--------------------- >> end captured stdout << ----------------------

----------------------------------------------------------------------
Ran 1 test in 0.527s

FAILED (failures=1)
```

```
                           Logit Regression Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 1000
Model:                          Logit   Df Residuals:                      998
Method:                           MLE   Df Model:                            1
Date:                Sun, 01 Mar 2015   Pseudo R-squ.:               0.0001789
Time:                        21:16:05   Log-Likelihood:                -426.09
converged:                       True   LL-Null:                       -426.16
                                        LLR p-value:                    0.6962
==============================================================================
                 coef    std err          z      P>|z|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept      1.8273      0.292      6.258      0.000         1.255     2.400
variable_d    -0.0167      0.043     -0.390      0.696        -0.101     0.067
==============================================================================
```

```
                           Logit Regression Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 1000
Model:                          Logit   Df Residuals:                      997
Method:                           MLE   Df Model:                            2
Date:                Sun, 01 Mar 2015   Pseudo R-squ.:                 0.06133
Time:                        21:36:58   Log-Likelihood:                -400.03
converged:                       True   LL-Null:                       -426.16
                                        LLR p-value:                 4.464e-12
==============================================================================
                 coef    std err          z      P>|z|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept     -2.7924      0.852     -3.278      0.001        -4.462    -1.123
variable_b     0.0878      0.014      6.445      0.000         0.061     0.114
variable_c    -0.1242      0.045     -2.783      0.005        -0.212    -0.037
==============================================================================
AUC score: 0.678741000497
```

```
                           Logit Regression Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 1000
Model:                          Logit   Df Residuals:                      996
Method:                           MLE   Df Model:                            3
Date:                Sun, 01 Mar 2015   Pseudo R-squ.:                  0.2088
Time:                        21:44:45   Log-Likelihood:                -337.19
converged:                       True   LL-Null:                       -426.16
                                        LLR p-value:                 2.458e-38
==============================================================================
                 coef    std err          z      P>|z|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept     -0.9958      0.932     -1.068      0.286        -2.823     0.832
variable_a    -0.0392      0.004     -9.828      0.000        -0.047    -0.031
variable_b     0.0996      0.015      6.677      0.000         0.070     0.129
variable_c    -0.1487      0.049     -3.048      0.002        -0.244    -0.053
==============================================================================
AUC score: 0.813842167329
```

7

Optimizing by Choosing a New Algorithm

Upgrading the classifier

```
......E
======================================================================
ERROR: naive_bayes_tests.given_two_classes_with_two_dimension_inputs_t
est
----------------------------------------------------------------------
Traceback (most recent call last):
  File "/Library/Python/2.7/site-packages/nose-1.3.0-py2.7.egg/nose/ca
se.py", line 197, in runTest
    self.test(*self.arg)
  File "/Users/justin/Documents/Code/Machine-Learning-Test-by-Test/Cha
pter 7/naive_bayes_tests.py", line 76, in given_two_classes_with_two_d
imension_inputs_test
    assert results['class b'] > results['class a'], "Should classify a
s class b because of dimension 2."
ValueError: The truth value of an array with more than one element is
ambiguous. Use a.any() or a.all()
-------------------- >> begin captured stdout << ---------------------
{'class b': array([ 0.379657  ,  0.99954911]), 'class a': array([  6.2
0342995e-01,   4.50893925e-04])}

--------------------- >> end captured stdout << ----------------------

----------------------------------------------------------------------
Ran 7 tests in 0.157s

FAILED (errors=1)
```

```
......F
======================================================================
FAIL: naive_bayes_tests.given_two_classes_with_two_dimension_inputs_te
st
----------------------------------------------------------------------
Traceback (most recent call last):
  File "/Library/Python/2.7/site-packages/nose-1.3.0-py2.7.egg/nose/ca
se.py", line 197, in runTest
    self.test(*self.arg)
  File "/Users/justin/Documents/Code/Machine-Learning-Test-by-Test/Cha
pter 7/naive_bayes_tests.py", line 76, in given_two_classes_with_two_d
imension_inputs_test
    assert results['class b'] > results['class a'], "Should classify a
s class b because of dimension 2."
AssertionError: Should classify as class b because of dimension 2.
-------------------- >> begin captured stdout << ---------------------
{'class b': 0.5, 'class a': 0.5}

--------------------- >> end captured stdout << ----------------------

----------------------------------------------------------------------
Ran 7 tests in 0.151s

FAILED (failures=1)
```

Applying our classifier

```
-------------------- >> begin captured stdout << ---------------------
{'F': [{'mean': 63.696218000000002, 'variance': 12.162890218475999},
       {'mean': 164.53548800000002, 'variance': 1653.804317381856},
       {'mean': 28.529872000000001, 'variance': 46.772100127615992}],
 'M': [{'mean': 69.163201999999984, 'variance': 13.695821797196},
       {'mean': 198.34125400000002, 'variance': 1589.8334884054839},
       {'mean': 29.133306000000001, 'variance': 28.673036156363999}]}
Men
M
M
M
Women
F
F
F

--------------------- >> end captured stdout << ----------------------

----------------------------------------------------------------------
Ran 10 tests in 1.290s

FAILED (failures=1)
```

```
    assert False
AssertionError:
-------------------- >> begin captured stdout << ---------------------
Correct rate: 0.7936, Total: 5000
{'F': [{'mean': 63.595020512820511, 'variance': 16.240506356844183},
       {'mean': 166.52739487179485, 'variance': 1994.8113539380934},
       {'mean': 36.486458119658124, 'variance': 30504.141355082436}],
 'M': [{'mean': 69.002157831325292, 'variance': 24.52815720761939},
       {'mean': 198.35300722891566, 'variance': 1702.1144923903094},
       {'mean': 37.555443373493979, 'variance': 26226.362753873291}]}

--------------------- >> end captured stdout << ----------------------

----------------------------------------------------------------------
Ran 11 tests in 20.608s

FAILED (failures=1)
```

Upgrading to Random Forest

```
-------------------- >> begin captured stdout << ---------------------
Correct rate: 0.0, Total: 5000

--------------------- >> end captured stdout << ----------------------

----------------------------------------------------------------------
Ran 12 tests in 21.813s

FAILED (errors=1)
```

8

Exploring Scikit-learn Test First

Getting choosey

```
-------------------- >> begin captured stdout << ---------------------
Classifier: <choosey.CopyCatClassifier instance at 0x107433f38>; Number right: 0
Classifier: <libs.NaiveBayes.Classifier instance at 0x107433f80>; Number right: 811
Classifier: <libs.RandomForest.Classifier instance at 0x107433fc8>; Number right: 809
<libs.RandomForest.Classifier instance at 0x107433fc8>
```

Developing testable documentation

Decision trees

```
-------------------- >> begin captured stdout << ---------------------
Classifier: <choosey.CopyCatClassifier instance at 0x10481b0e0>; Number right: 0
Classifier: <libs.NaiveBayes.Classifier instance at 0x10481b128>; Number right: 827
Classifier: <libs.RandomForest.Classifier instance at 0x10481b170>; Number right: 808
Classifier: <libs.DecisionTree.Classifier instance at 0x10481b1b8>; Number right: 780
<libs.DecisionTree.Classifier instance at 0x10481b1b8>

--------------------- >> end captured stdout << ----------------------
```

*Chapter 6 & 9 do not have images