Errata for Machine Learning in Action by Peter Harrington
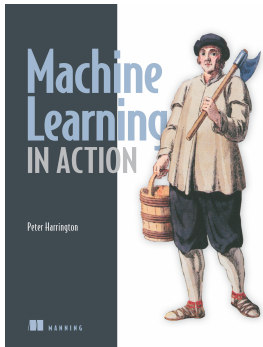
Updated December, 2013

Hi, and thanks for taking a look at **Machine Learning in Action**. If you find any errors that aren't published below, please submit them to the book's Author Online Forum.

On page 17: (not in a code listing)

The line:
myEye = randMat*invRandMat
should appear above the line:
>>> myEye – eye(4)

Listing 2.2, bottom of page 25 thru page26

A better version of the function file2matrix() is given below. The code in the book will work.

```
def file2matrix(filename):
    fr = open(filename)
    arrayOLines = fr.readlines()
    numberOfLines = len(arrayOLines)
    returnMat = zeros((numberOfLines,3))
    classLabelVector = []
    index = 0
    for line in arrayOLines:
        line = line.strip()
        listFromLine = line.split('\t')
        returnMat[index,:] = listFromLine[0:3]
        classLabelVector.append(int(listFromLine[-1]))
        index += 1
    return returnMat,classLabelVector
```

Page 26:

datingDataMat, datingLabels = kNN.file2matrix('datingTestSet.txt')
should be:
datingDataMat, datingLabels = kNN.file2matrix('datingTestSet2.txt')

>>> datingLabels[0:20]
['didntLike', 'smallDoses',……
should be:
>>> datingLabels[0:20]
[3, 2, 1, 1, 1, 1, 3, 3, 1, 3, 1, 1, 2, 1, 1, 1, 1, 1, 2, 3]

Listing 2.5, page 32

datingDataMat, datingLabels = file2matrix('datingTestSet.txt')
should be:
datingDataMat, datingLabels = file2matrix('datingTestSet2.txt')

Page 41 (not a code listing)

$l(x_i) = \log_2 p(x_i)$
Should be:
$l(x_i) = -\log_2 p(x_i)$

Page 42: Listing 3.1

The line:
labelCounts[currentLabel] = 0
should be indented from the lines above and below it. The code in the repo is correct.

Listing 3.3, page 45

bestFeature = I
should be:
bestFeature = i

Page 52 (not a code listing)

>>> treePlotter.retrieveTree(1)
should return:
{'no surfacing': {0: 'no', 1: {'flippers': {0: {'head': {0: 'no', 1: 'yes'}}, 1: 'no'}}}}

Page 92

Before the line:
>>> reload(logRegres)
add the line:
>>> from numpy import *
At the beginning of the book I mention that this should be added to every interactive session, but it will be helpful for people to see this line here if they don't remember that.

Page 104 (not a code listing)

|wTx+b|/ ||w||
should be:
|wTA+b|/||w||

Listing 8.4, page 168

The line:
returnMat = zeros((numIt,n))
Should be added before the line:
ws = zeros((n,1)); wsTest = ws.copy(); wsMax = ws.copy()

Listing 9.5, page 195

yHat = mat((m,1))
Should be:
yHat = mat(zeros((m,1)))

Page 230, second paragraph:

You can't create a set of just one integer in Python. It needs to be a list (try it out).

should be

The frozenset constructor requires something iterable and won't take a single integer. To accommodate this, we package each single integer in a list.