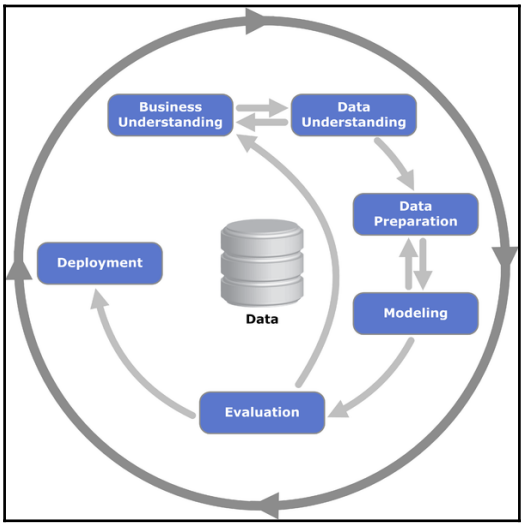
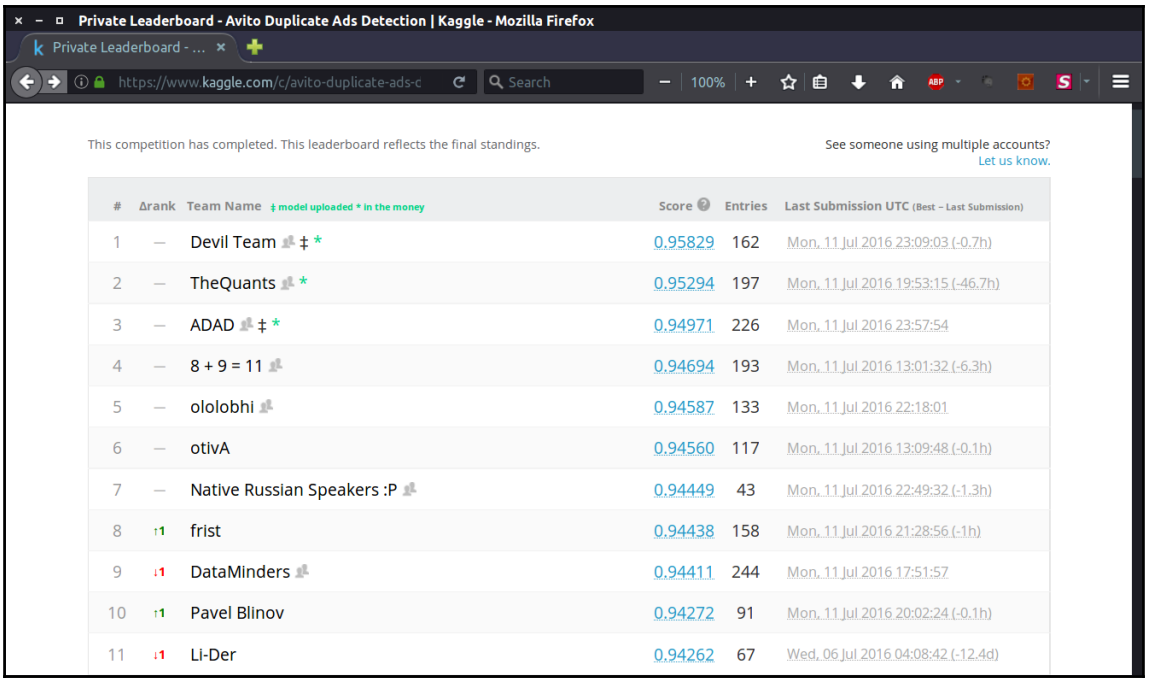
Chapter 1: Data Science Using Java



Chapter 2: Data Processing Toolbox

Order	Column Title	Data Type
1	word	Fixed Number of Words
2	pos	Custom List

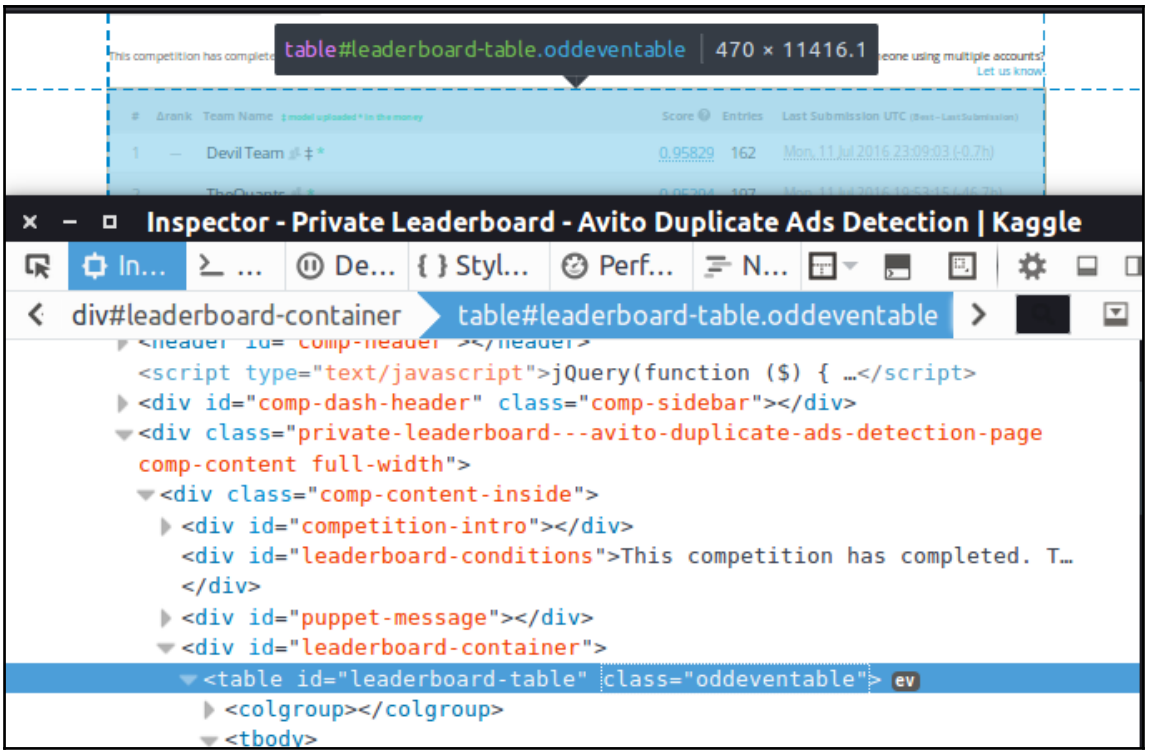
Order	Column Title	Data Type	Examples	Options
1	name	Names	John Smith	Name Surname
2	email	Email	No examples available.	No options available.
3	country	Country	No examples available.	Limit countries to those selected above
4	salary	Currency	Please Select	Format: XXXXXX Range 25000 To 150000 Currency Symbol $ prefix
5	experience	Number Range	No examples available.	Between 0 and 50



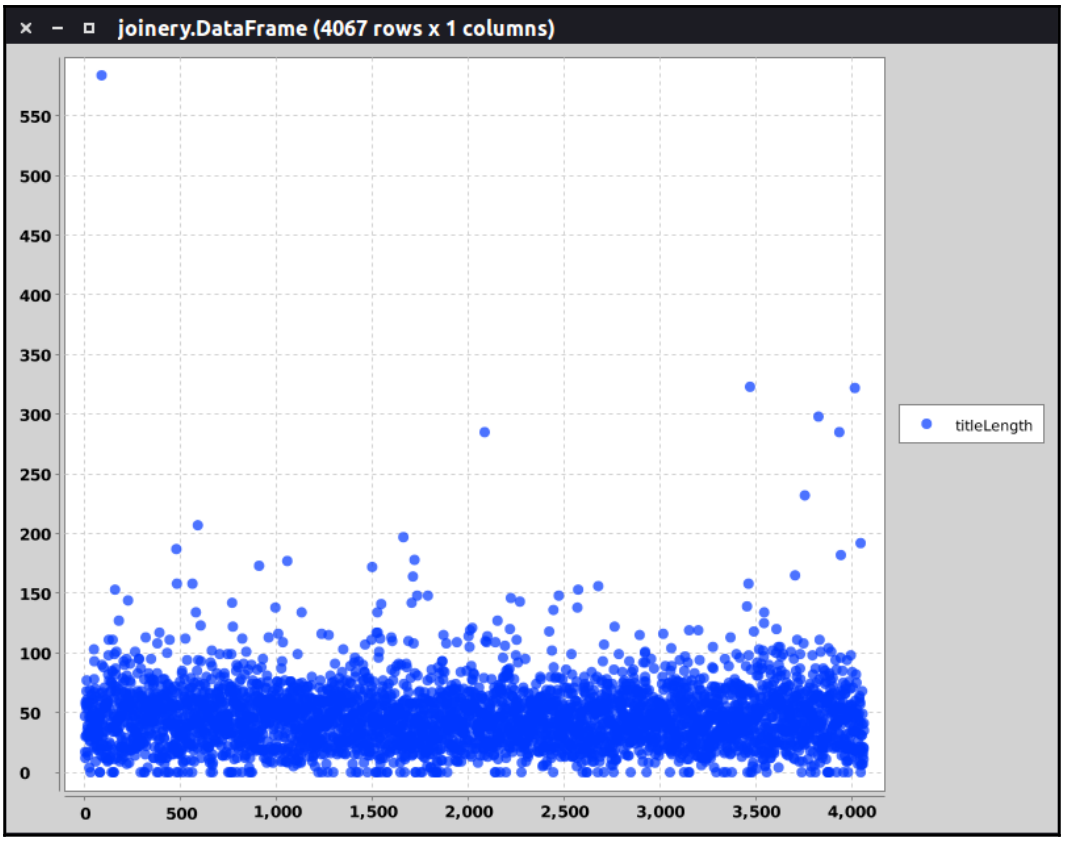
This competition has completed. This leaderboard reflects the final standings.

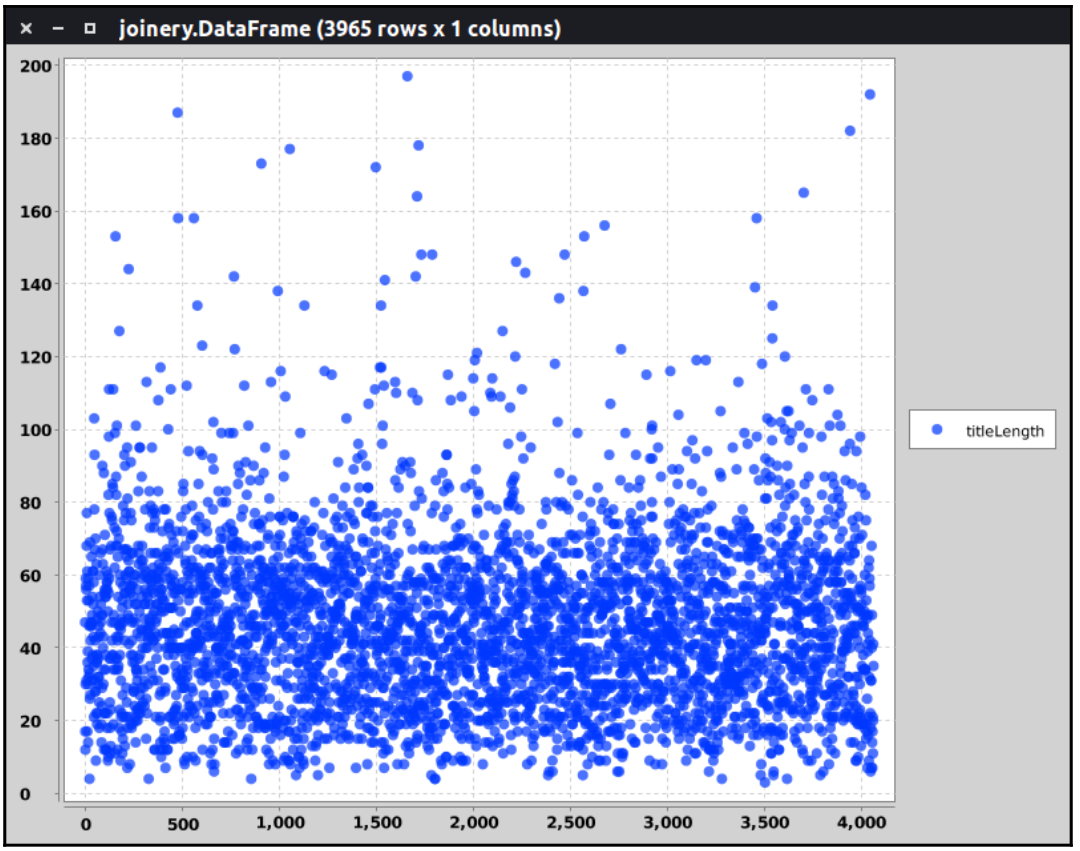
See someone using multiple accounts?
Let us know.

#	Δrank	Team Name ‡ model uploaded * in the money	Score	Entries	Last Submission UTC (Best – Last Submission)
1	–	Devil Team ‡ *	0.95829	162	Mon, 11 Jul 2016 23:09:03 (-0.7h)
2	–	TheQuants *	0.95294	197	Mon, 11 Jul 2016 19:53:15 (-46.7h)
3	–	ADAD ‡ *	0.94971	226	Mon, 11 Jul 2016 23:57:54
4	–	8 + 9 = 11	0.94694	193	Mon, 11 Jul 2016 13:01:32 (-6.3h)
5	–	ololobhi	0.94587	133	Mon, 11 Jul 2016 22:18:01
6	–	otivA	0.94560	117	Mon, 11 Jul 2016 13:09:48 (-0.1h)
7	–	Native Russian Speakers :P	0.94449	43	Mon, 11 Jul 2016 22:49:32 (-1.3h)
8	↑1	frist	0.94438	158	Mon, 11 Jul 2016 21:28:56 (-1h)
9	↓1	DataMinders	0.94411	244	Mon, 11 Jul 2016 17:51:57
10	↑1	Pavel Blinov	0.94272	91	Mon, 11 Jul 2016 20:02:24 (-0.1h)
11	↓1	Li-Der	0.94262	67	Wed, 06 Jul 2016 04:08:42 (-12.4d)



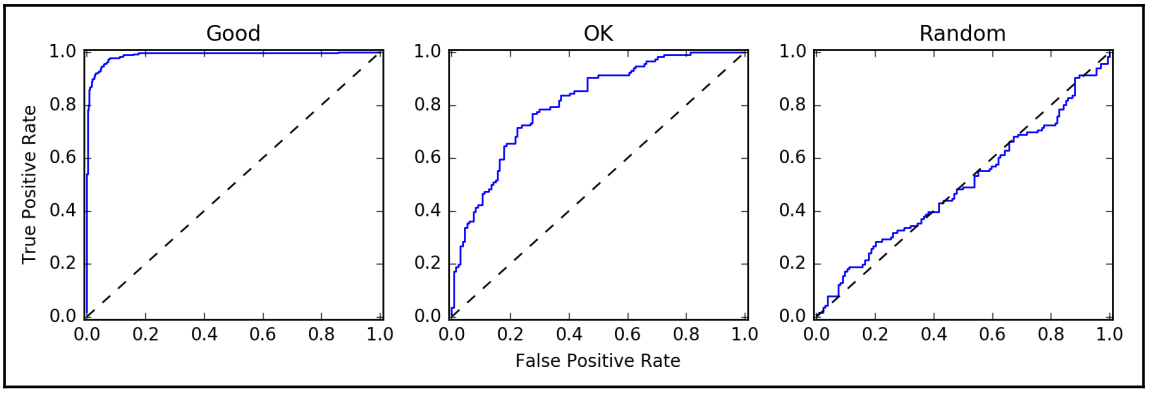
Chapter 3: Exploratory Data Analysis

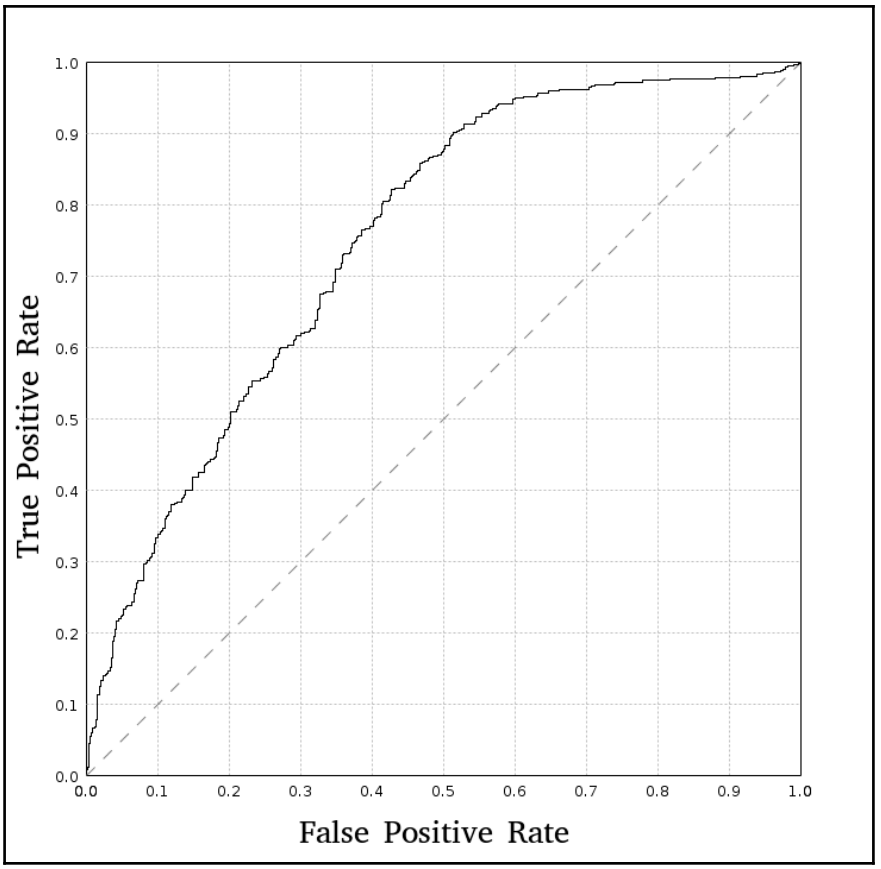

joinery.DataFrame (4067 rows x 1 columns)
550
500
450
400
350
300
250
200
150
100
50
0
0
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
titleLength

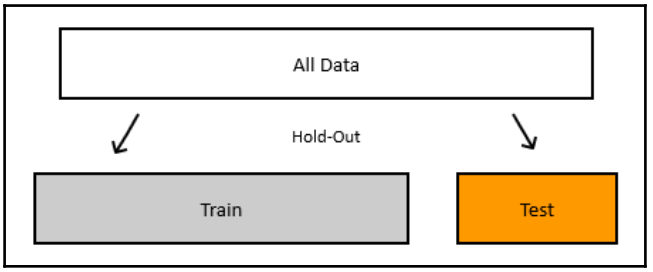


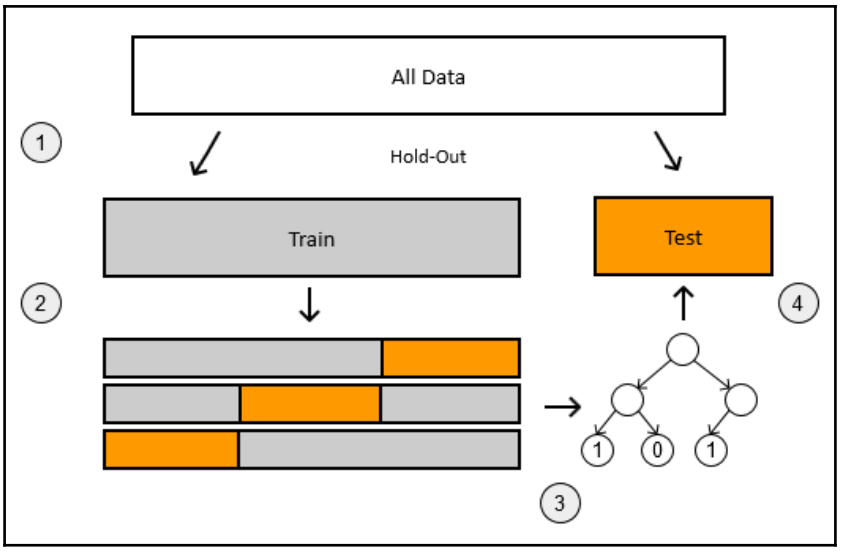
Chapter 4: Supervised Learning - Classification and Regression

		Actual Class y	
		Positive	Negative
$h_\theta(x)$ **Predicted outcome**	Predicted positive outcome	**True positive** (TP)	**False positive** (FP)
	Predicted negative outcome	**False negative** (FN)	**True negative** (TN)

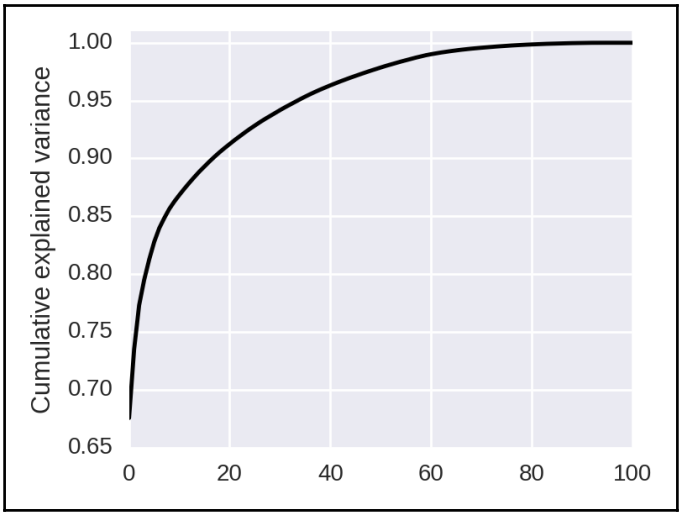


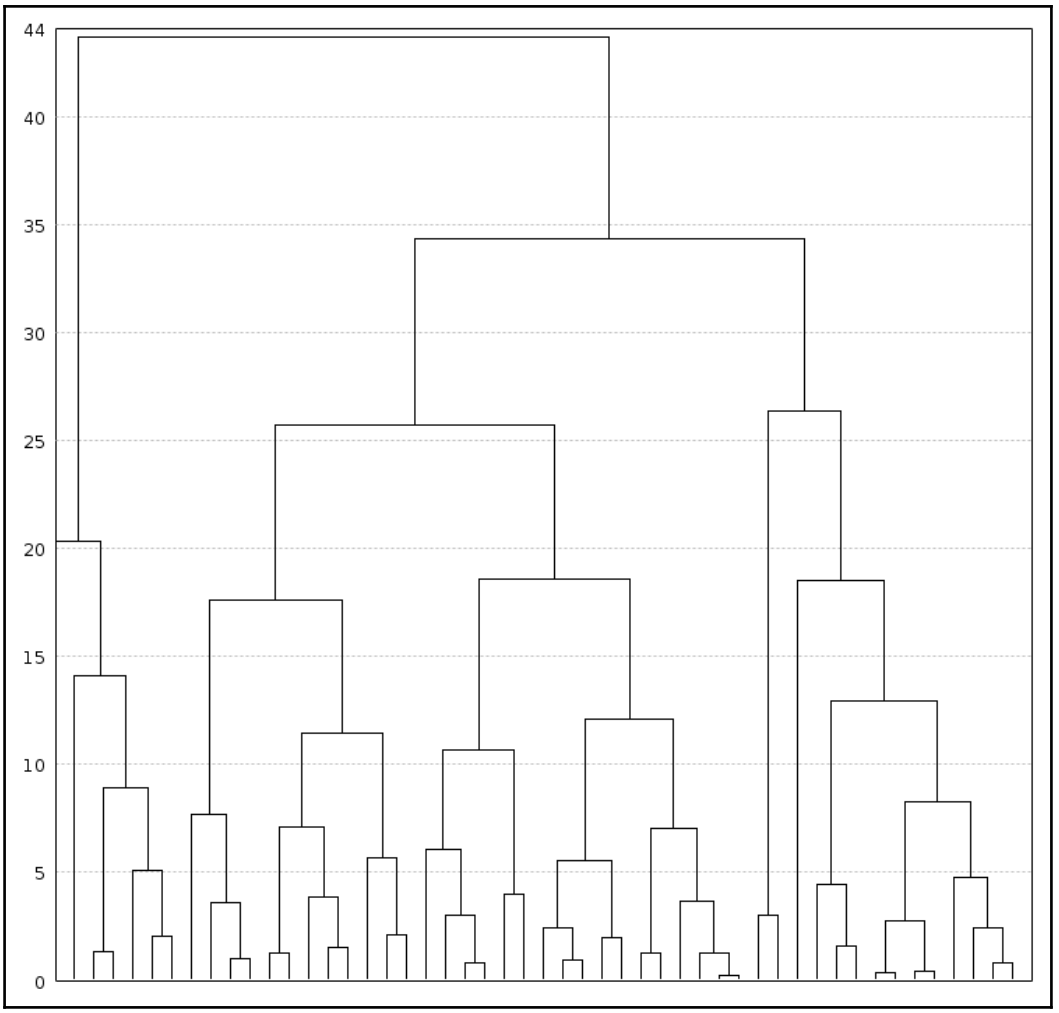

True Positive Rate
False Positive Rate
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

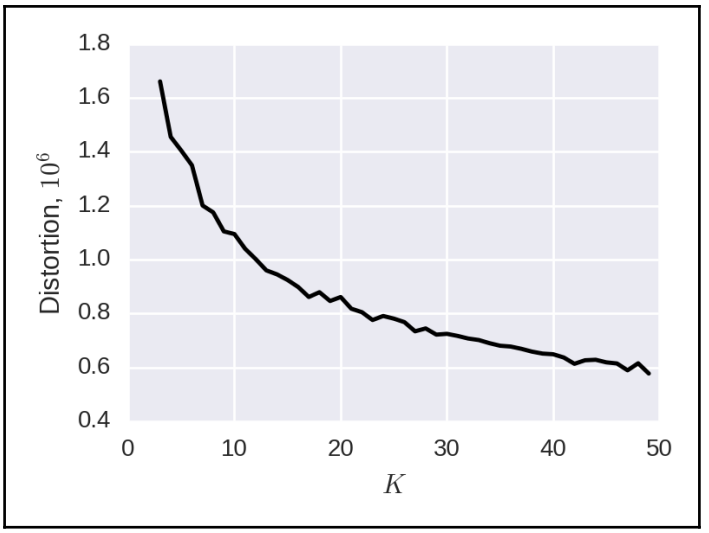

All Data
Hold-Out
Train
Test



Chapter 5: Unsupervised Learning - Clustering and Dimensionality Reduction




44
40
35
30
25
20
15
10
5
0



Chapter 6: Working with Text - Natural Language Processing and Information Retrieval

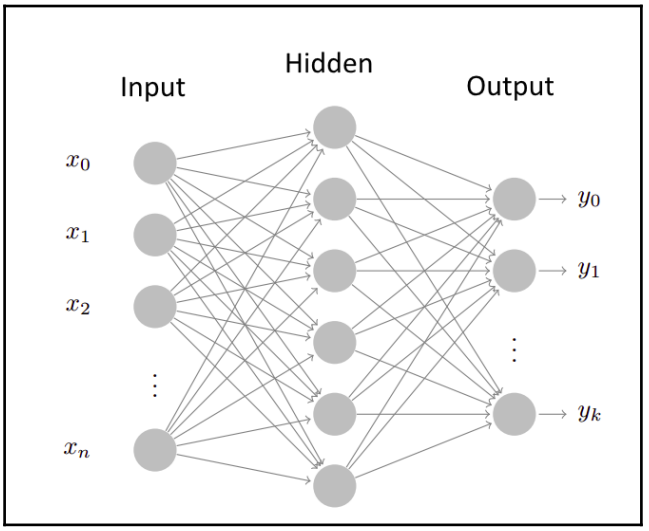
	because	data	development	enterprise	for	good	is	java	science	use	we
doc1	1	1			1			1	1	1	2
doc2			1	1	1	1	1	1			

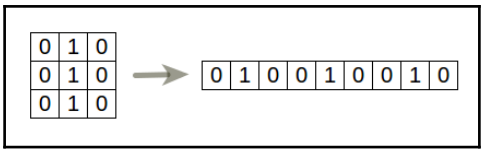
Chapter 7: Extreme Gradient Boosting

qid	1	1	1	1	2	2	2	3	3	3	3	3
size	4				3			5				

qid	query	url
1	adidas basketball shoes	http://www.adidas.com/us/men-basketball-shoes
1	adidas basketball shoes	http://www.adidas.com/us/basketball
1	adidas basketball shoes	http://www.eastbay.com/adidas/Basketball/Shoes/_-_/N-x8Z1e1Zne
2	angry birds	https://www.angrybirds.com/
2	angry birds	http://chrome.angrybirds.com/
2	angry birds	http://www.youtube.com/channel/UCYC2wjLop-S6Ld4raeoUVNA
2	angry birds	http://www.rovio.com/index.php?page=angry-birds
3	animal shelter	http://www.animalshelter.org/
3	animal shelter	https://www.petfinder.com/animal-shelters-and-rescues
3	animal shelter	https://www.animalhouseshelter.com/pets/berlin/
3	animal shelter	https://www.petfinder.com/
3	animal shelter	http://www.adoptapet.com/adoption_rescue/80291.html

Chapter 8: Deep Learning with DeepLearning4J




0 1 0
0 1 0
0 1 0
0 1 0 0 1 0 0 1 0

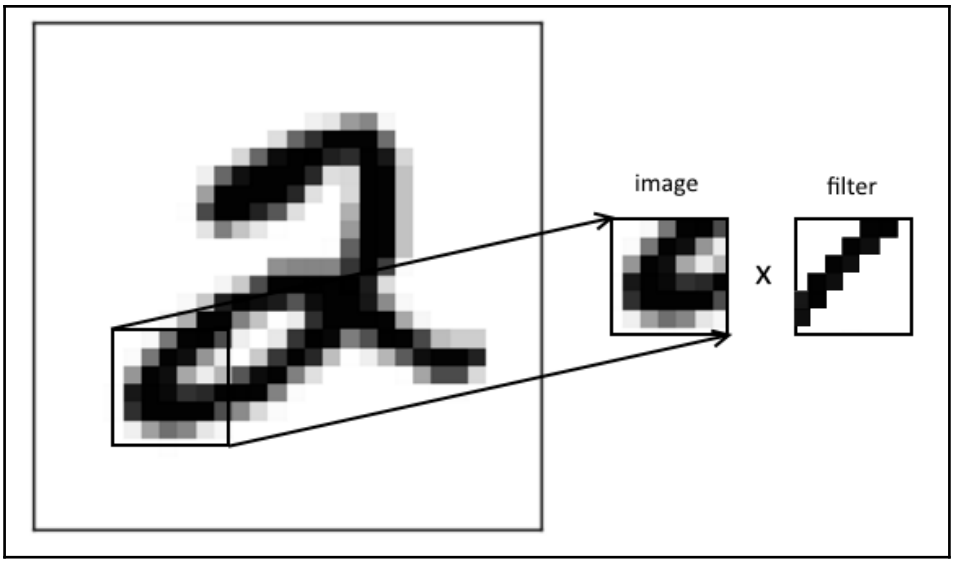
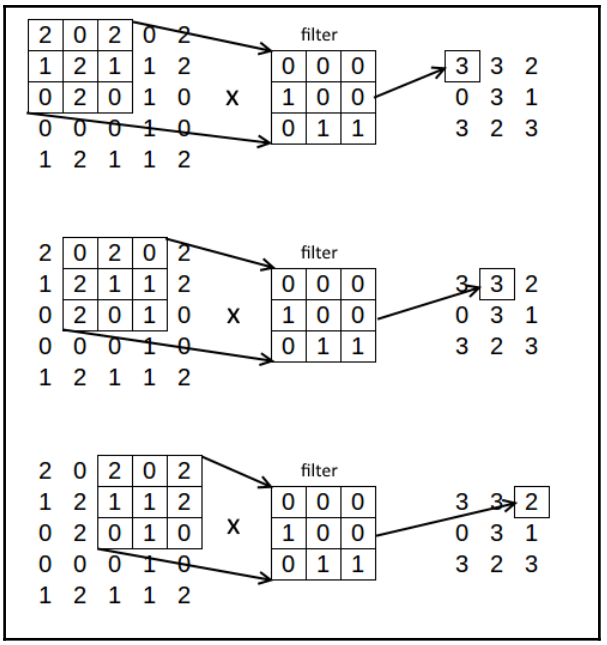
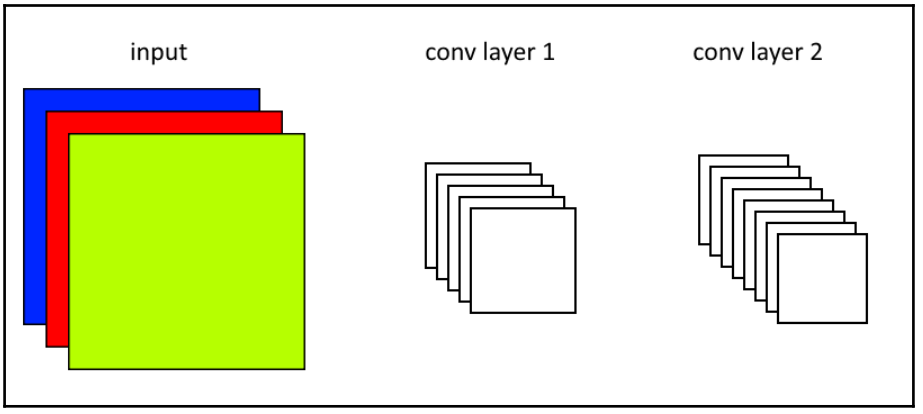
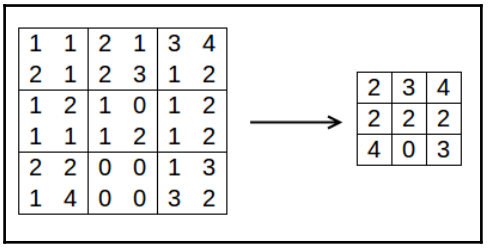
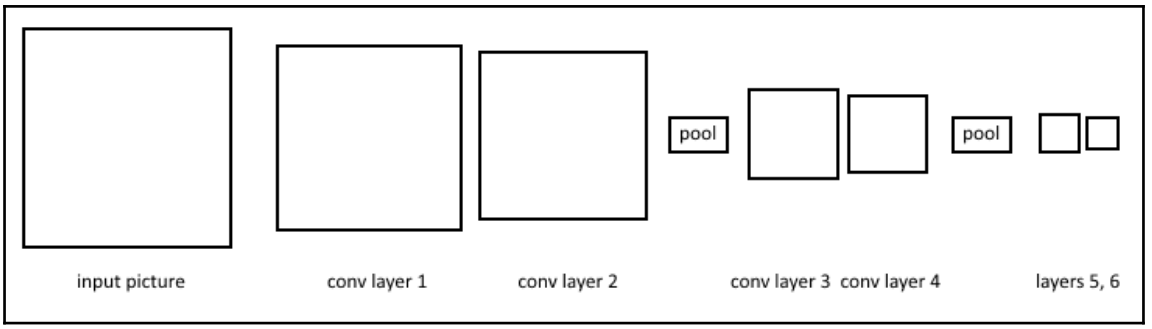


image
filter
X

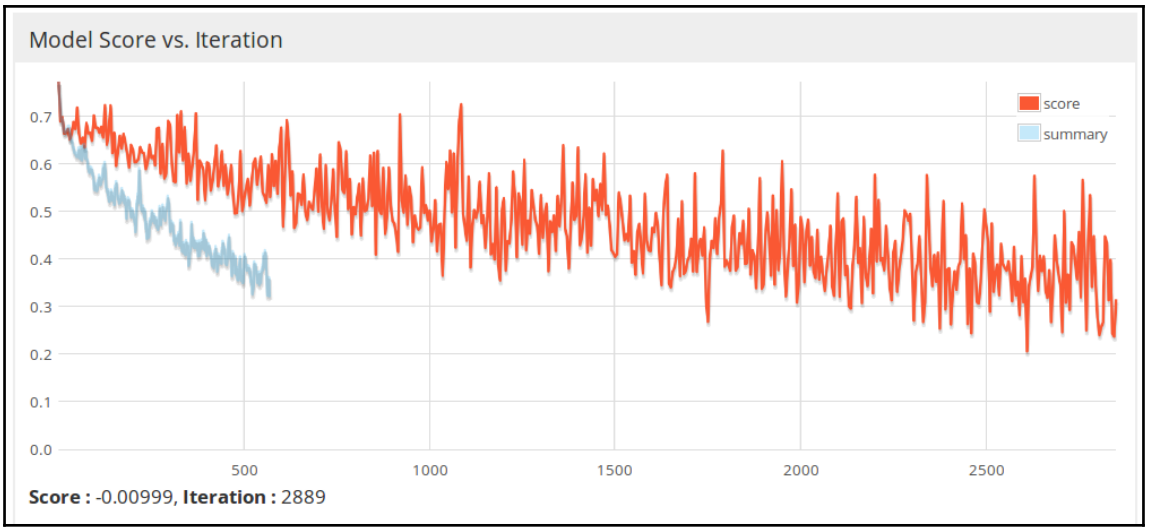

filter
filter
filter

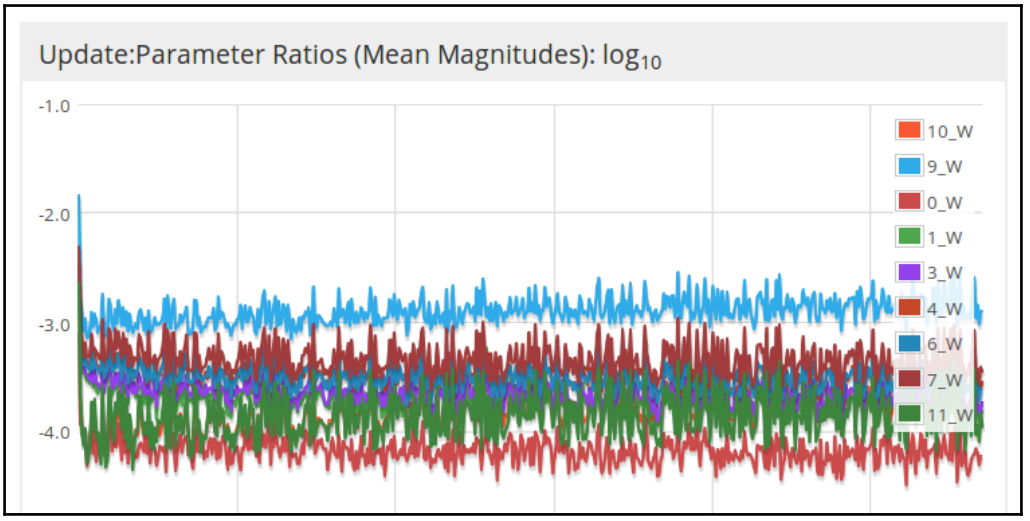

input
conv layer 1
conv layer 2

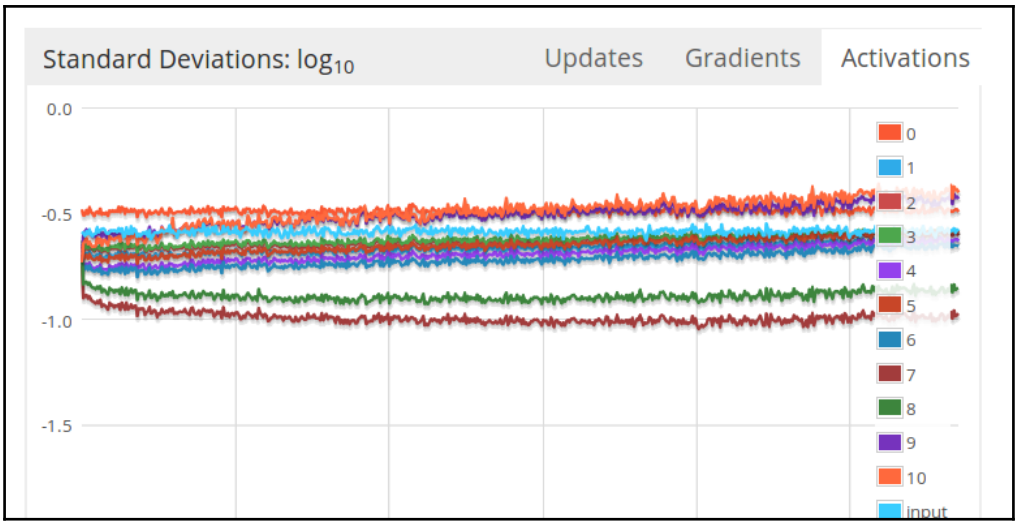


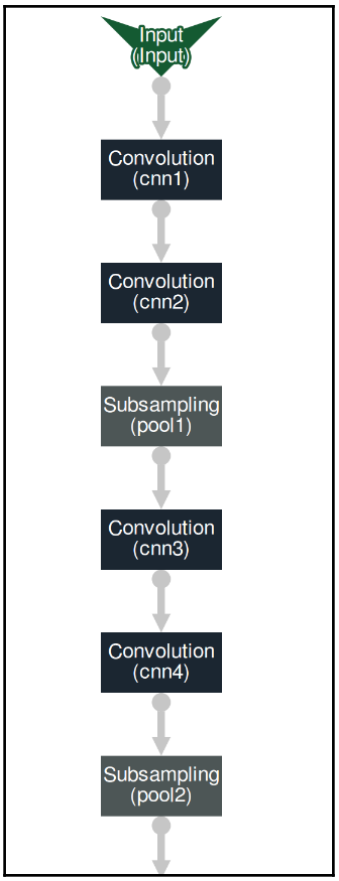

pool
pool
input picture
conv layer 1
conv layer 2
conv layer 3
conv layer 4
layers 5, 6

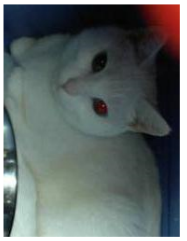
DL4J Training UI
Overview
Model
System
Language
Model Score vs. Iteration
score
summary
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
500
1000
1500
2000
2500
Score : -0.00999, Iteration : 2889
Model and Training Information
Model Type MultiLayerNetwork
Layers 12
Total Parameters 9856546
Start Time
Total Runtime
Last Update 2016-12-23 08:56:34
Total Parameter Updates 2851
Updates/sec .06
Examples/sec 2.86
Update:Parameter Ratios (Mean Magnitudes): log10
-1.0
-2.0
-3.0
-4.0
10_W
9_W
0_W
1_W
3_W
4_W
6_W
7_W
11_W
Standard Deviations: log10
Updates
Gradients
Activations
0.0
-0.5
-1.0
-1.5
0
1
2
3
4
5
6
7
8
9
10
input

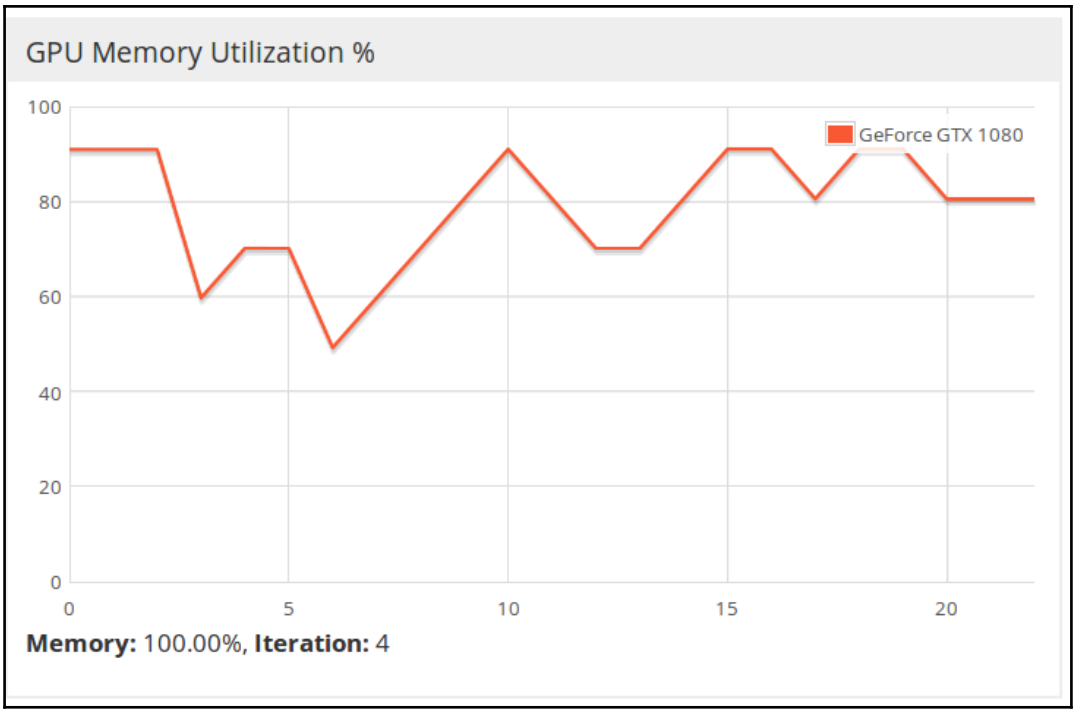

Model Score vs. Iteration
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
500
1000
1500
2000
2500
score
summary
Score : -0.00999, Iteration : 2889


Update:Parameter Ratios (Mean Magnitudes): log10
-1.0
-2.0
-3.0
-4.0
10_W
9_W
0_W
1_W
3_W
4_W
6_W
7_W
11_W

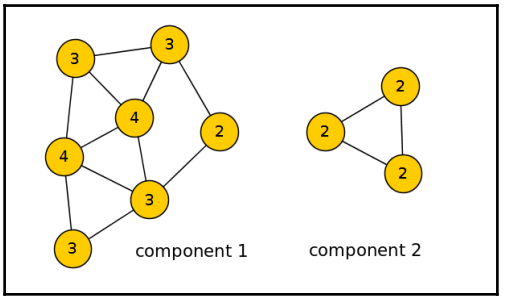

Standard Deviations: log10
Updates
Gradients
Activations
0.0
-0.5
-1.0
-1.5
0
1
2
3
4
5
6
7
8
9
10
input

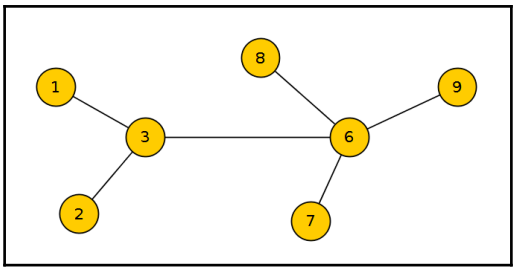

Input
(Input)
Convolution
(cnn1)
Convolution
(cnn2)
Subsampling
(pool1)
Convolution
(cnn3)
Convolution
(cnn4)
Subsampling
(pool2)

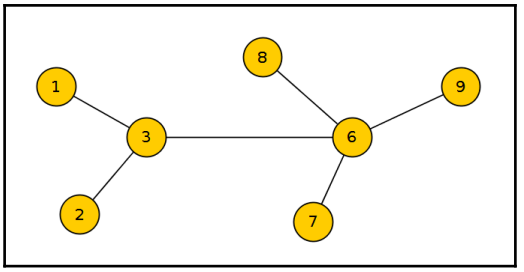




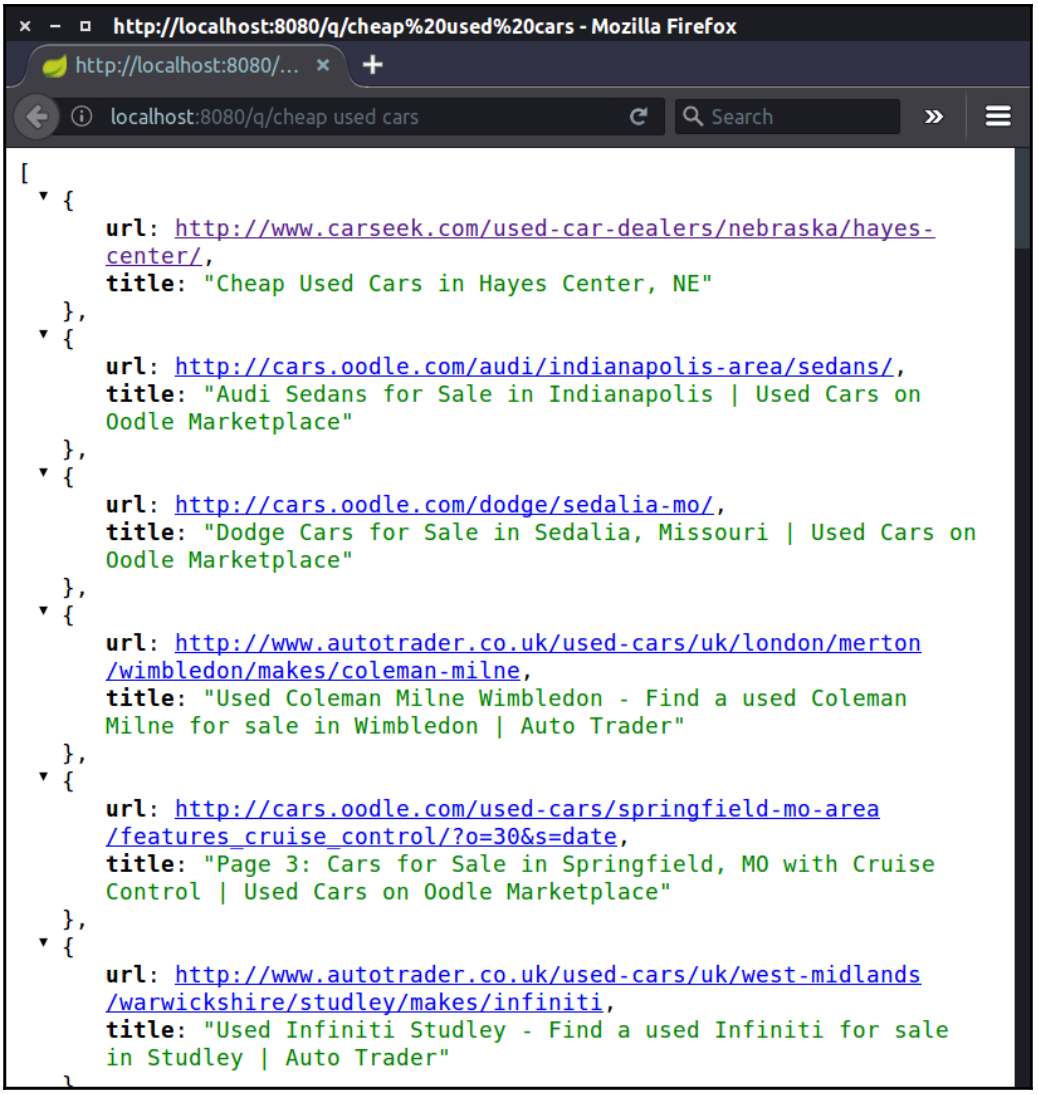
Chapter 9: Scaling Data Science







Chapter 10: Deploying Data Science Models


http://localhost:8080/q/cheap%20used%20cars - Mozilla Firefox
http://localhost:8080/...
localhost:8080/q/cheap used cars
Search
[
{
url: http://www.carseek.com/used-car-dealers/nebraska/hayes-center/,
title: "Cheap Used Cars in Hayes Center, NE"
},
{
url: http://cars.oodle.com/audi/indianapolis-area/sedans/,
title: "Audi Sedans for Sale in Indianapolis | Used Cars on Oodle Marketplace"
},
{
url: http://cars.oodle.com/dodge/sedalia-mo/,
title: "Dodge Cars for Sale in Sedalia, Missouri | Used Cars on Oodle Marketplace"
},
{
url: http://www.autotrader.co.uk/used-cars/uk/london/merton/wimbledon/makes/coleman-milne,
title: "Used Coleman Milne Wimbledon - Find a used Coleman Milne for sale in Wimbledon | Auto Trader"
},
{
url: http://cars.oodle.com/used-cars/springfield-mo-area/features_cruise_control/?o=30&s=date,
title: "Page 3: Cars for Sale in Springfield, MO with Cruise Control | Used Cars on Oodle Marketplace"
},
{
url: http://www.autotrader.co.uk/used-cars/uk/west-midlands/warwickshire/studley/makes/infiniti,
title: "Used Infiniti Studley - Find a used Infiniti for sale in Studley | Auto Trader"