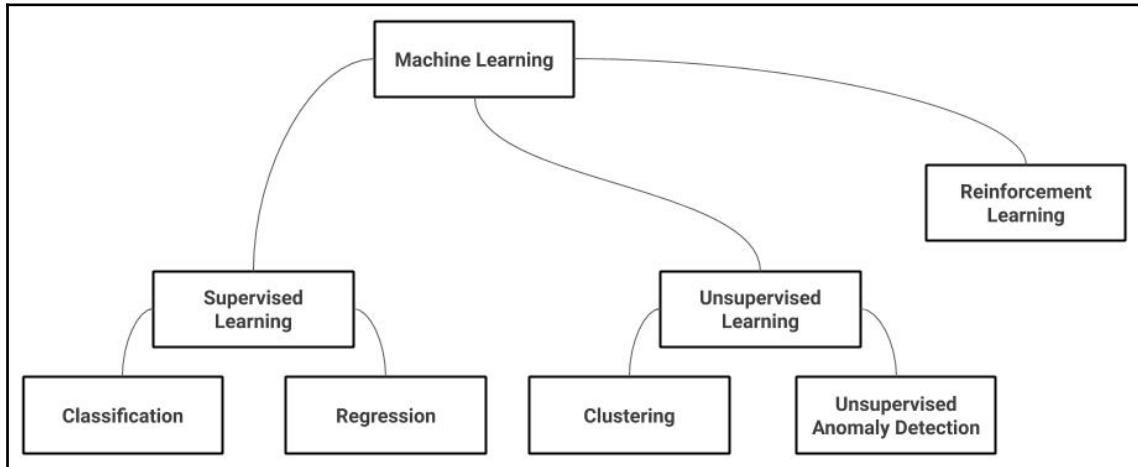


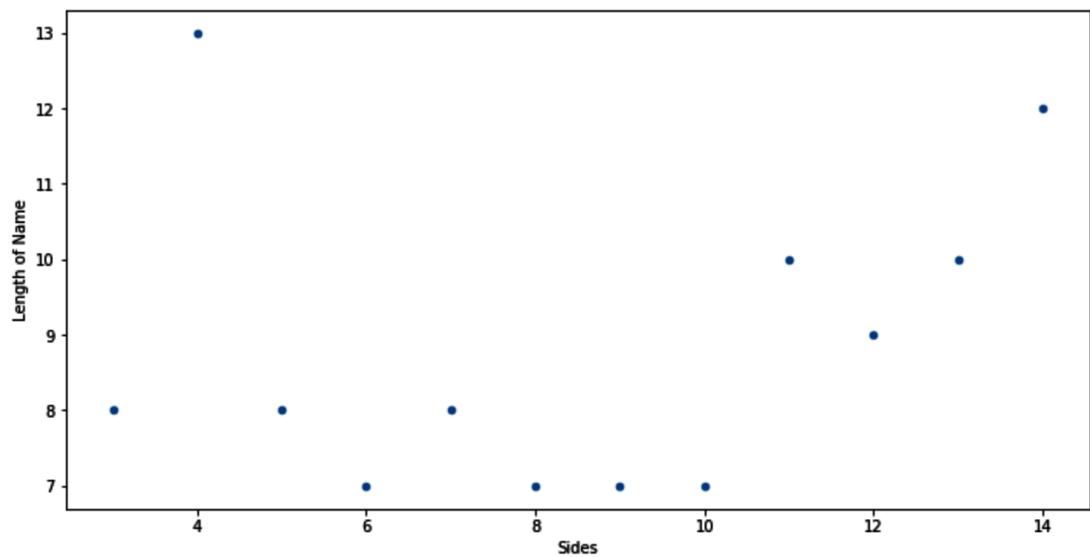
Chapter 1: Introduction to Machine Learning



	Name	Sides
0	Triangle	3
1	Quadrilateral	4
2	Pentagon	5

	Name	Sides
7	Decagon	10
9	Dodecagon	12
8	Hendecagon	11
4	Heptagon	7
3	Hexagon	6

Sides versus Length of the Name



Chapter 2: Making Decisions with Trees

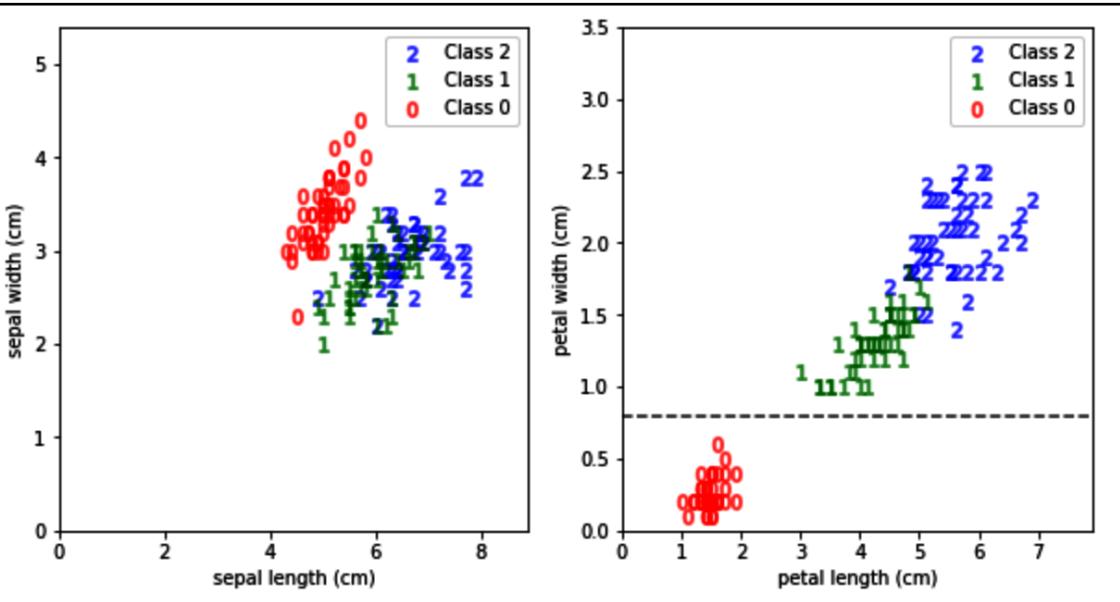
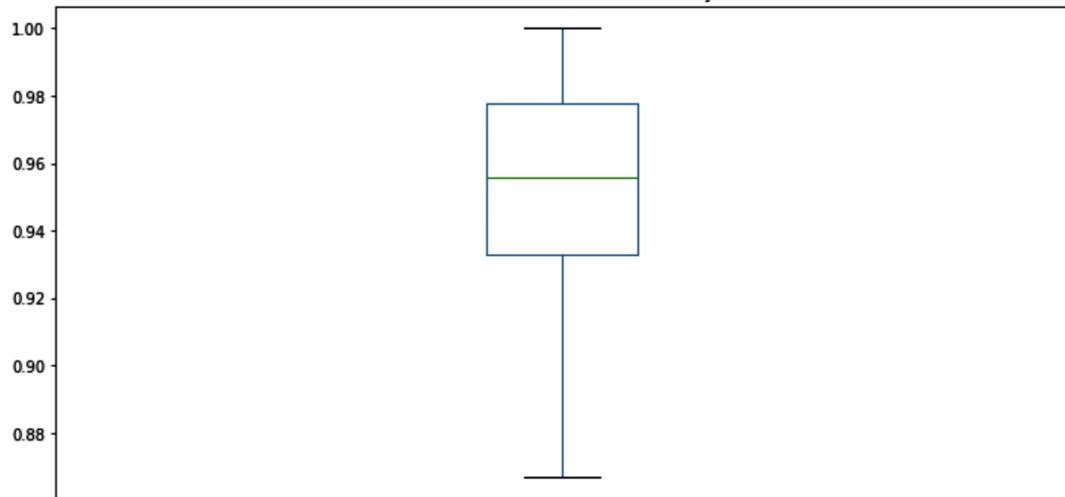
Chance of Rain	UV Index	Umbrella
0.1	11	True
0.9	1	True
0.3	3	False
0.1	2	False

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target	target_names
26	5.0	3.4	1.6	0.4	0	setosa
101	5.8	2.7	5.1	1.9	2	virginica
113	5.7	2.5	5.0	2.0	2	virginica
52	6.9	3.1	4.9	1.5	1	versicolor
134	6.1	2.6	5.6	1.4	2	virginica
11	4.8	3.4	1.6	0.2	0	setosa

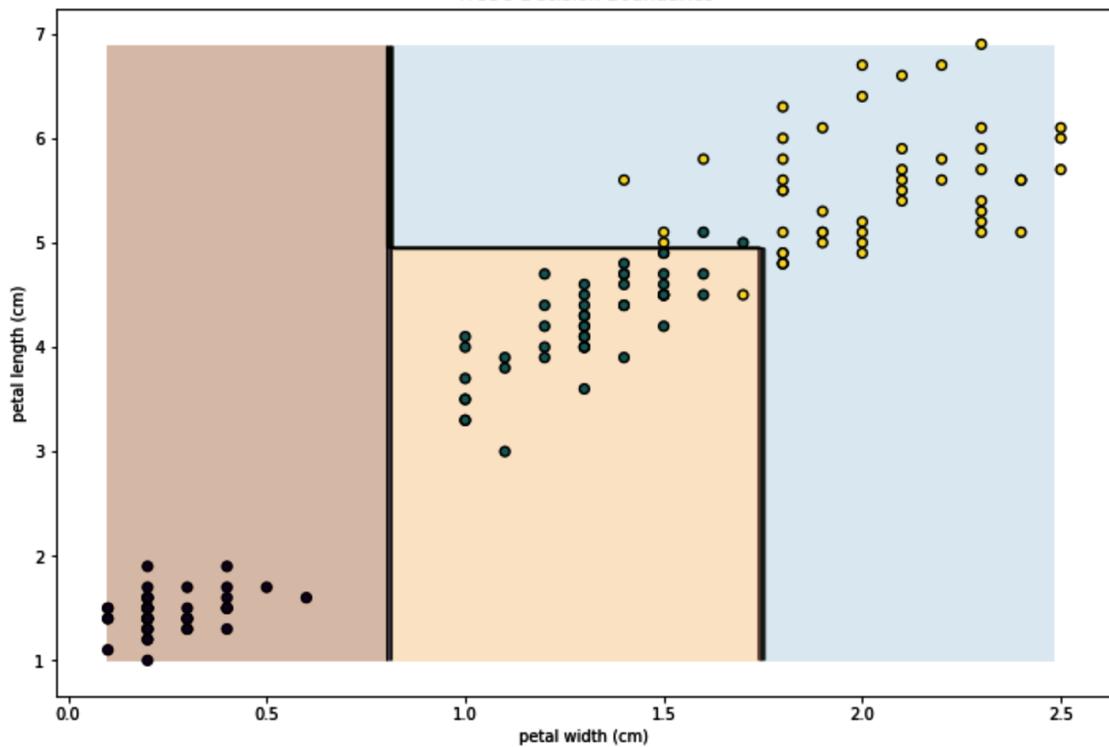
feature_importances	
feature_names	
petal width (cm)	0.594961
petal length (cm)	0.405039
sepal length (cm)	0.000000
sepal width (cm)	0.000000

Average Score: 0.945 [5th percentile: 0.889 & 95th percentile: 0.979]

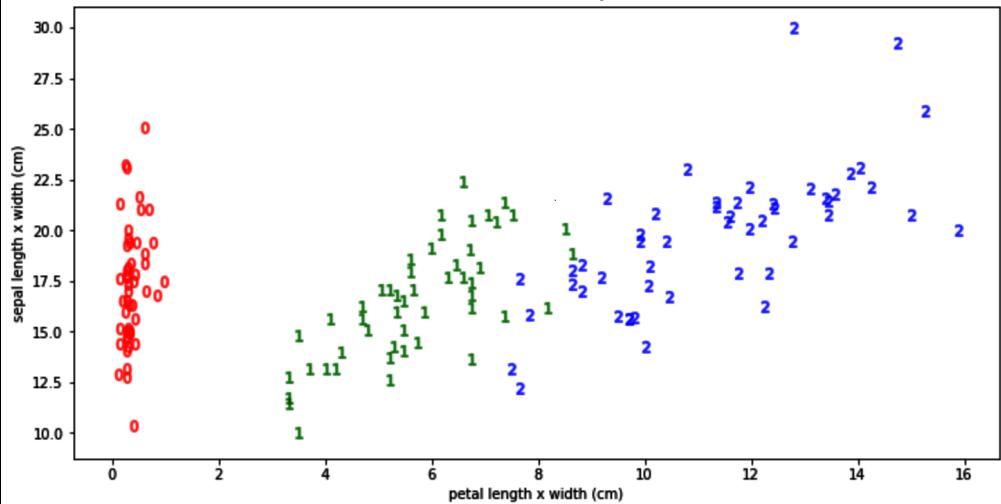
Distribution of classifier accuracy

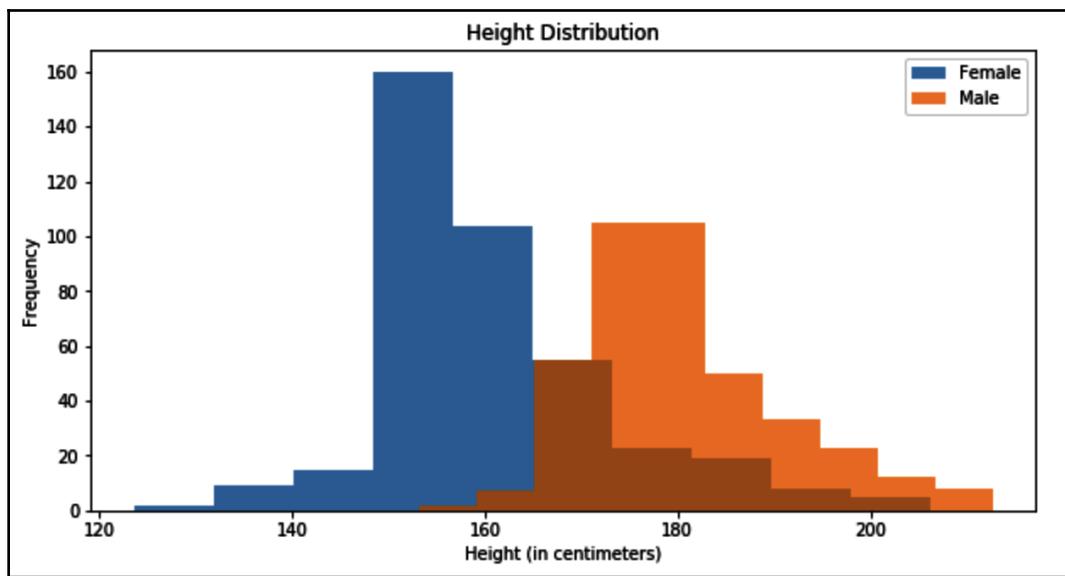
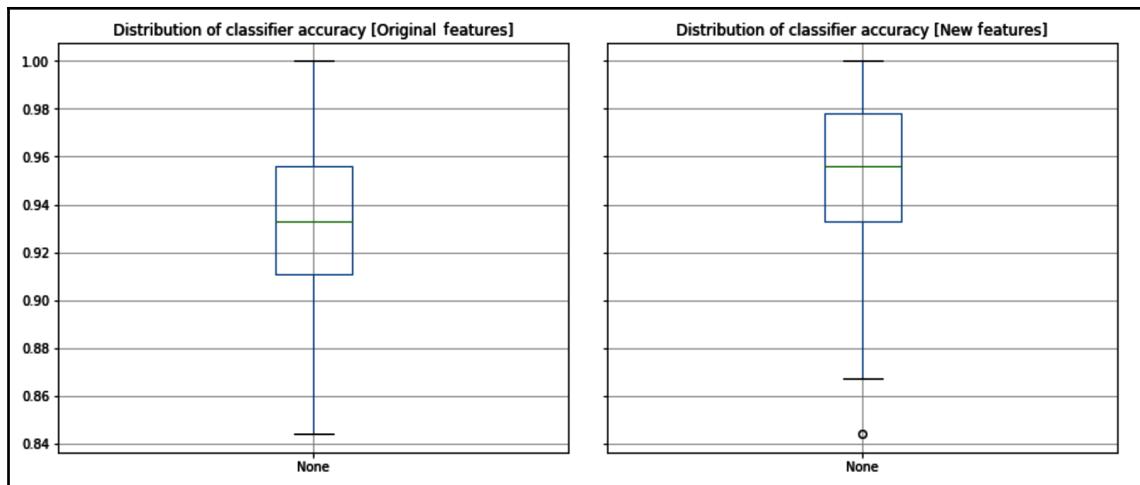


Tree's Decision Boundaries



Class distribution versus the newly derived features

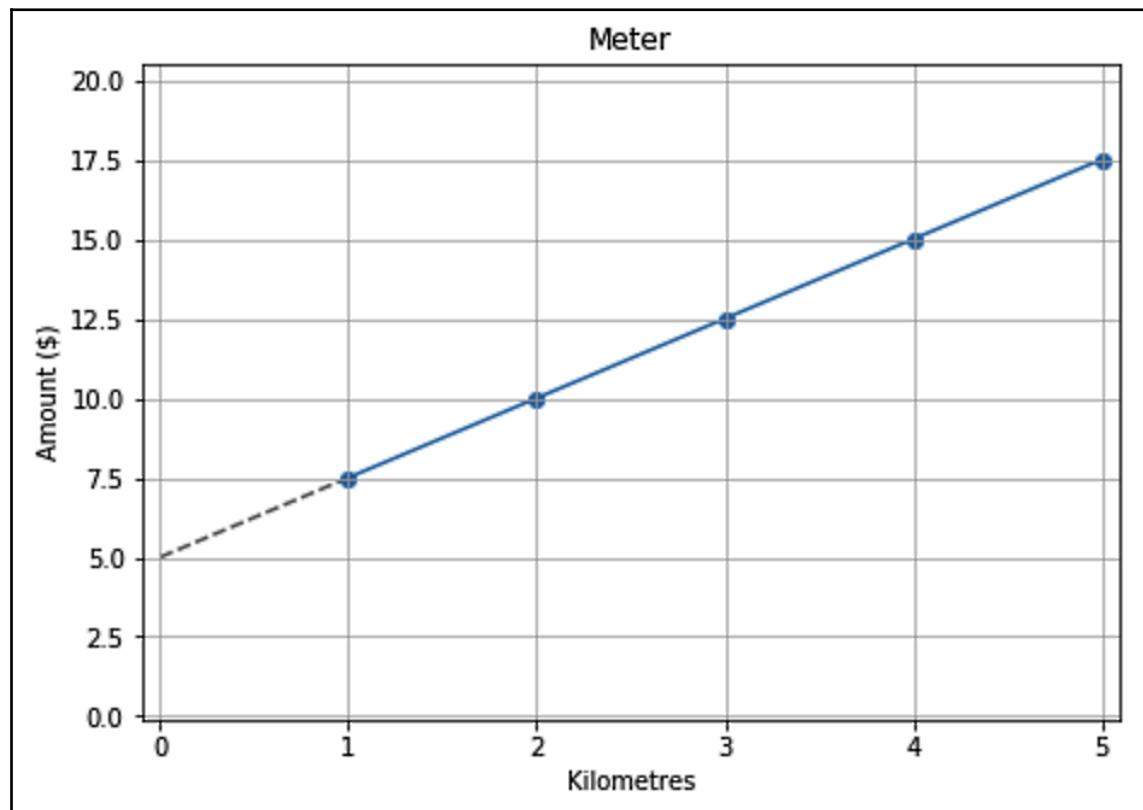




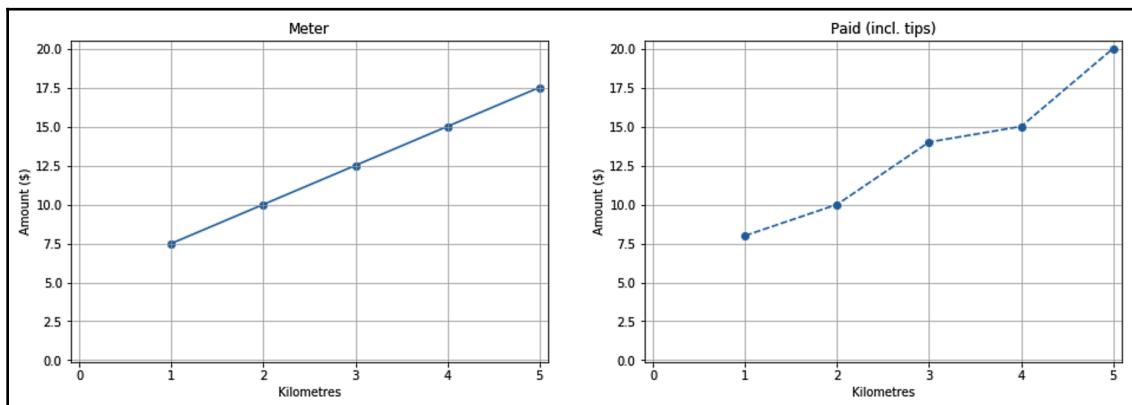
Gender	Height	
	mean	median
1	160.5	157.3
2	180.5	178.4

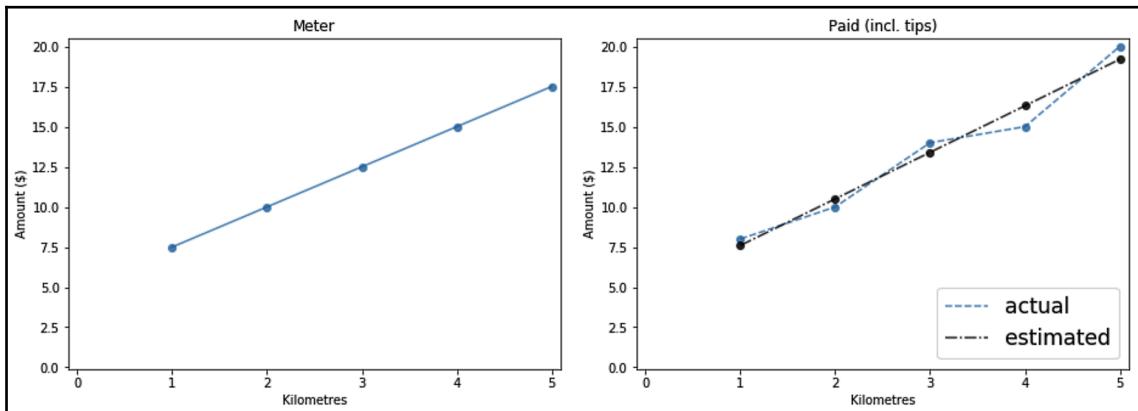
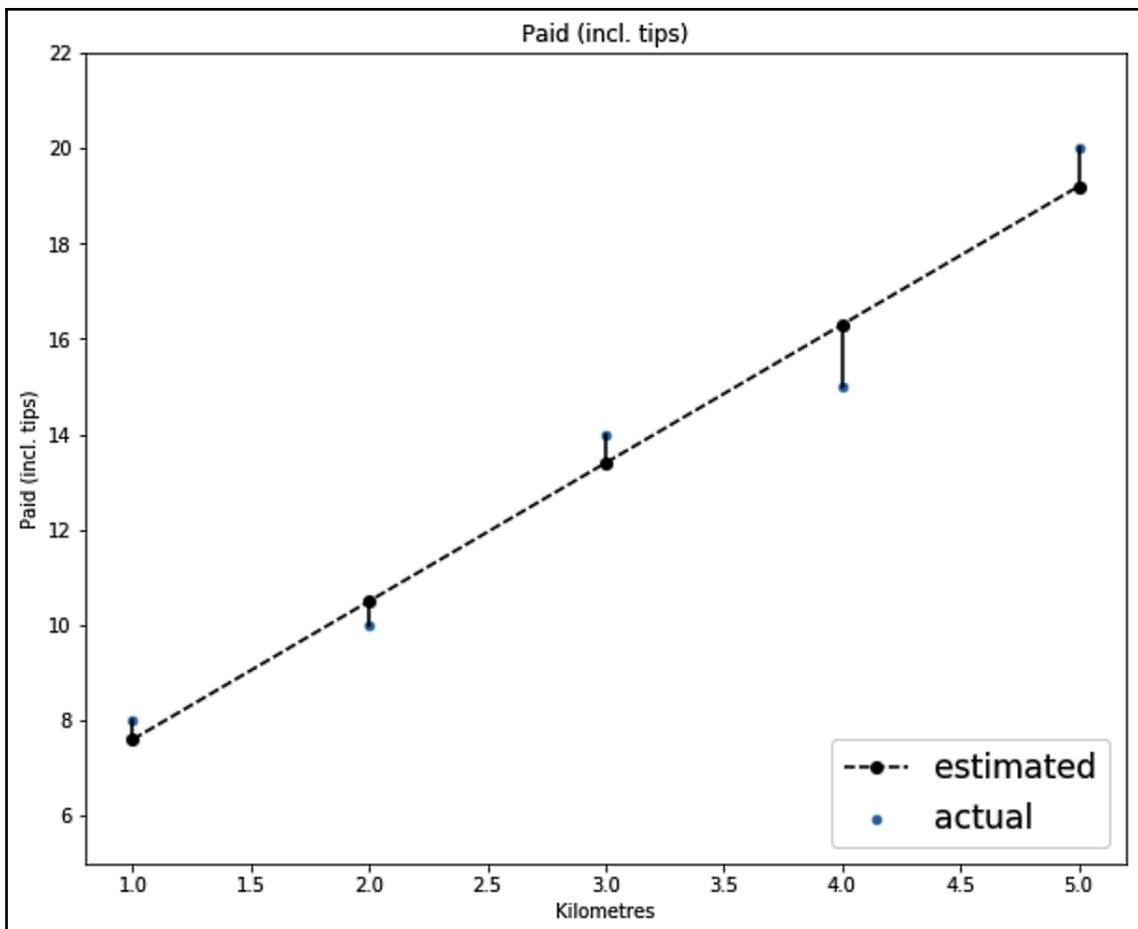
Chapter 3: Making Decisions with Linear Equations

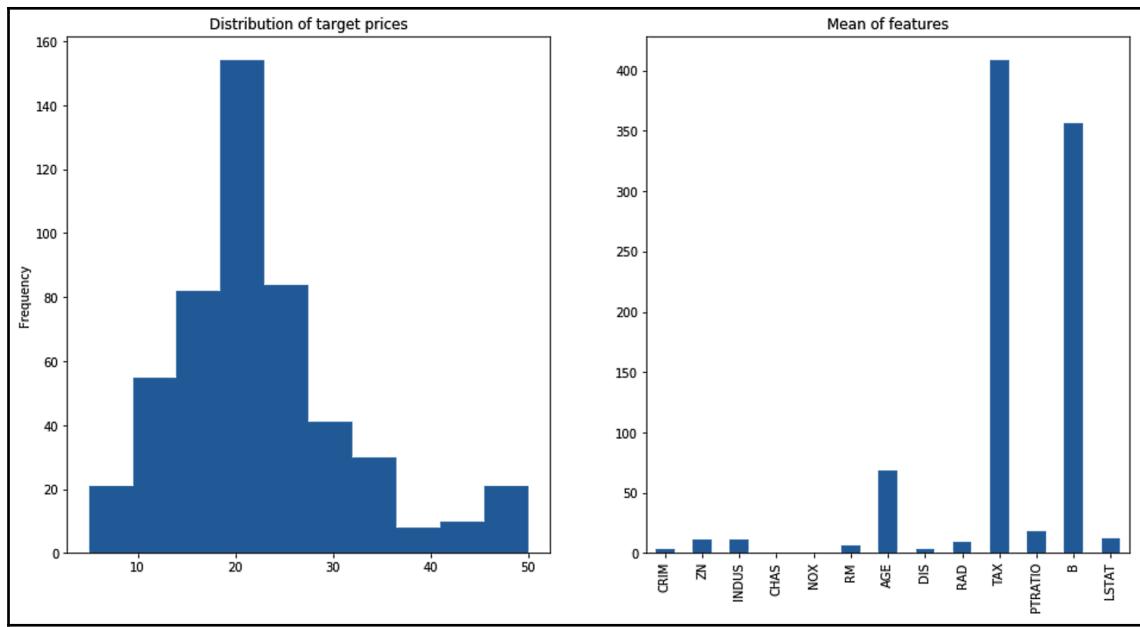
Meter	
Kilometres	
1	7.5
2	10.0
3	12.5
4	15.0
5	17.5



	Meter	Paid (incl. tips)
Kilometres		
1	7.5	8
2	10.0	10
3	12.5	14
4	15.0	15
5	17.5	20

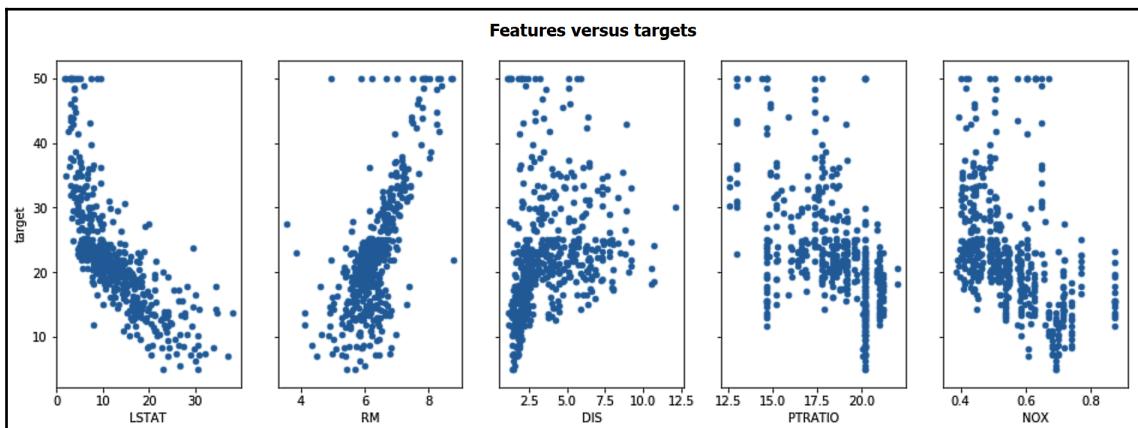


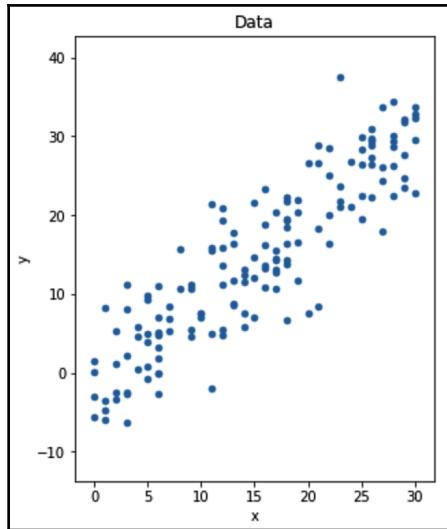
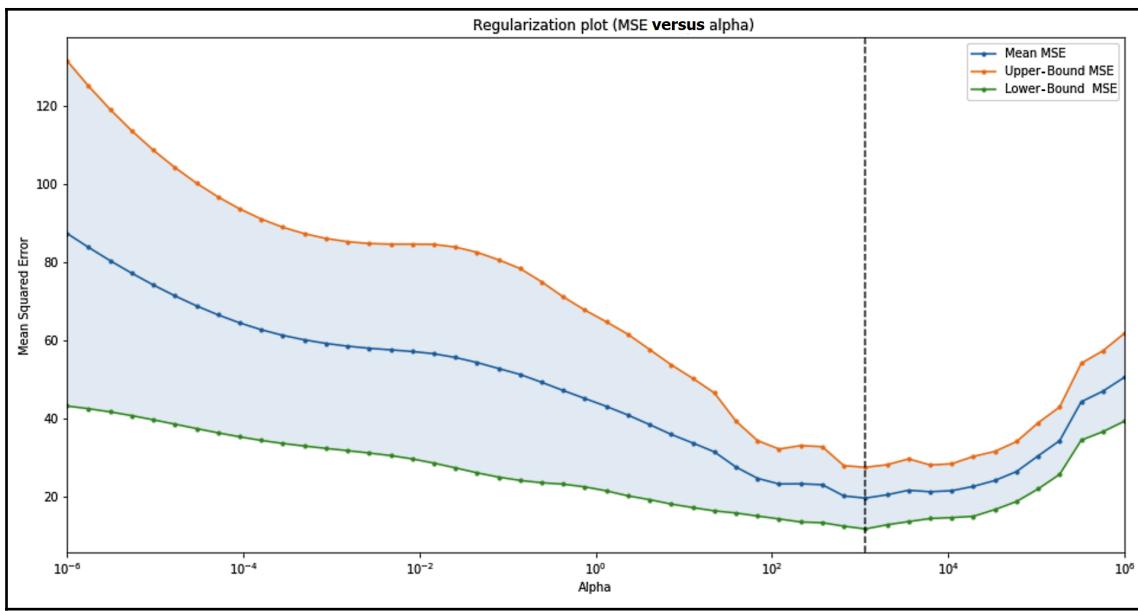


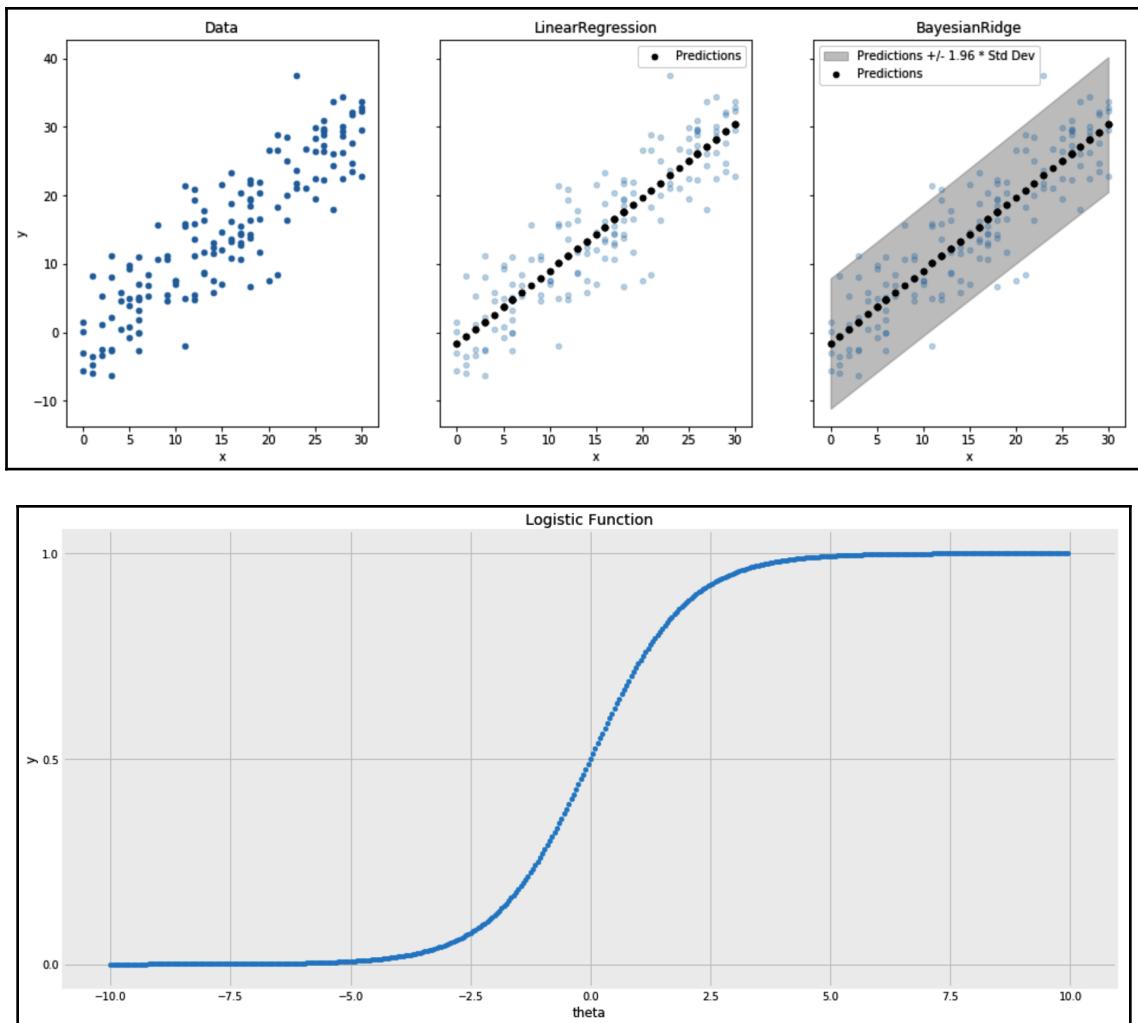


Features	Coeff	ABS(Coeff)
RM	3.771080	3.771080
CHAS	2.300030	2.300030
RAD	0.306593	0.306593
INDUS	0.048167	0.048167
ZN	0.041997	0.041997
B	0.010654	0.010654
AGE	0.008924	0.008924
TAX	-0.010705	0.010705
CRIM	-0.080001	0.080001
LSTAT	-0.646811	0.646811
PTRATIO	-1.036784	1.036784
DIS	-1.314393	1.314393
NOX	-15.325464	15.325464

	Coeff	ABS(Coeff)
Features		
RM	19.473855	19.473855
RAD	7.051632	7.051632
B	4.200840	4.200840
ZN	4.199695	4.199695
CHAS	2.300030	2.300030
INDUS	1.313999	1.313999
AGE	0.866509	0.866509
TAX	-5.598546	5.598546
CRIM	-5.882282	5.882282
NOX	-7.340897	7.340897
PTRATIO	-9.331056	9.331056
DIS	-12.592804	12.592804
LSTAT	-22.800083	22.800083

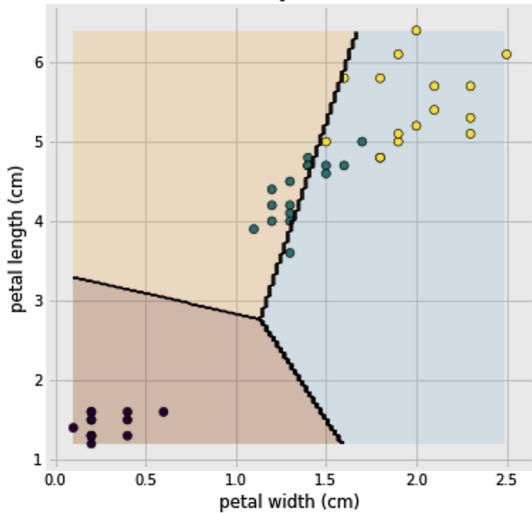




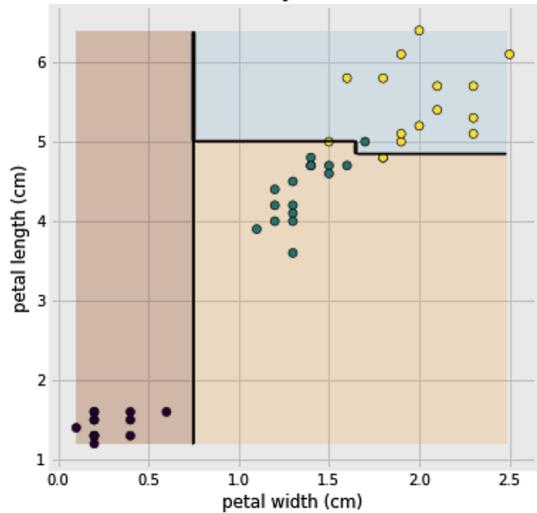


	Coef [Class 0]	Coef [Class 1]	Coef [Class 2]
sepal length (cm)	-1.404926	0.261358	1.143569
sepal width (cm)	1.560021	-1.437230	-0.122791
petal length (cm)	-2.860633	0.418396	2.442238
petal width (cm)	-2.787136	-0.478438	3.265574
intercept	2.203698	1.229151	-3.432848

Logistic Regression Classifier
Accuracy: 88.89%

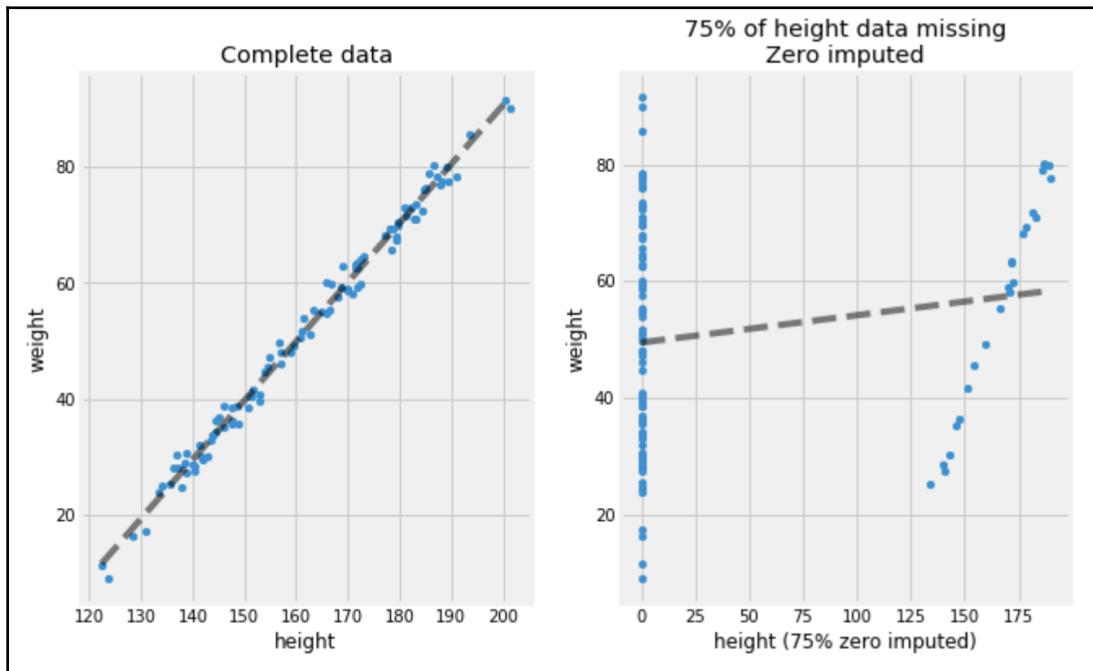


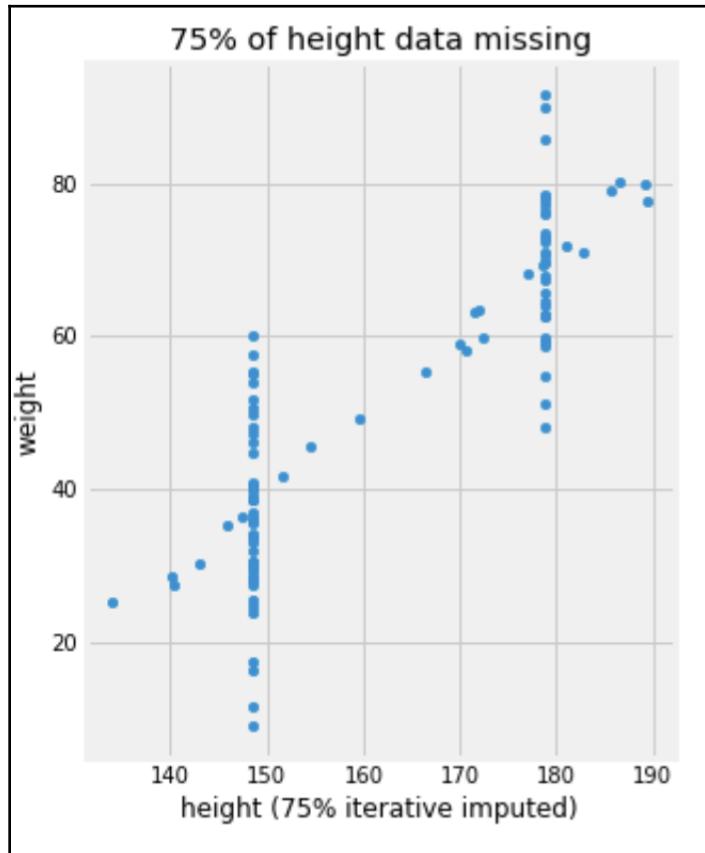
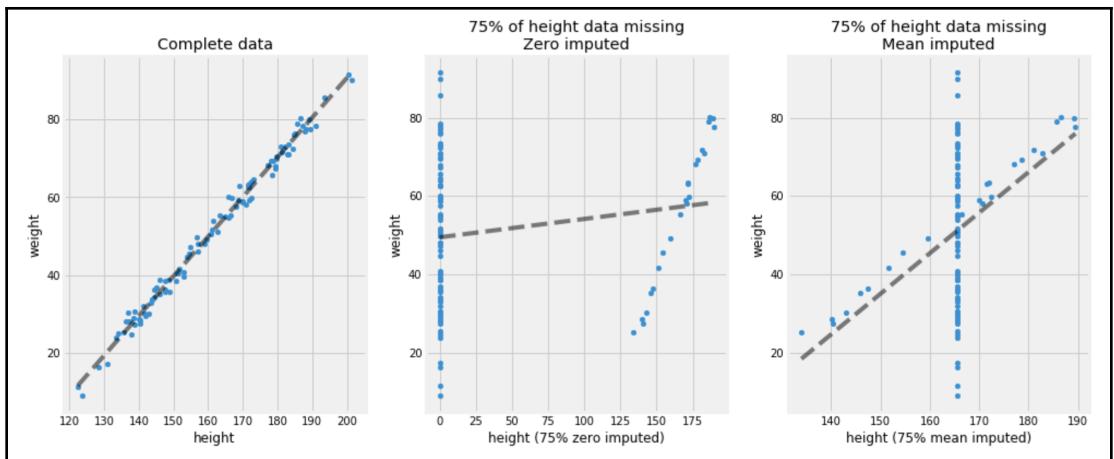
Decision Tree Classifier
Accuracy: 91.11%



Chapter 4: Preparing Your Data

	height (with 75% NaN)	weight
0	NaN	30.417204
1	172.458316	59.818478
2	NaN	47.255345
3	NaN	50.682057





	Size_XL	Size_L	Size_XXL	Size_M	Brand_Reebok	Brand_Le Coq	Brand_Puma	Brand_Nike
4	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
5	NaN	NaN	NaN	NaN	1.0	0.0	0.0	0.0
1	NaN	NaN	NaN	NaN	1.0	0.0	0.0	0.0
2	NaN	NaN	NaN	NaN	0.0	1.0	0.0	0.0
8	0.0	0.0	1.0	0.0	NaN	NaN	NaN	NaN

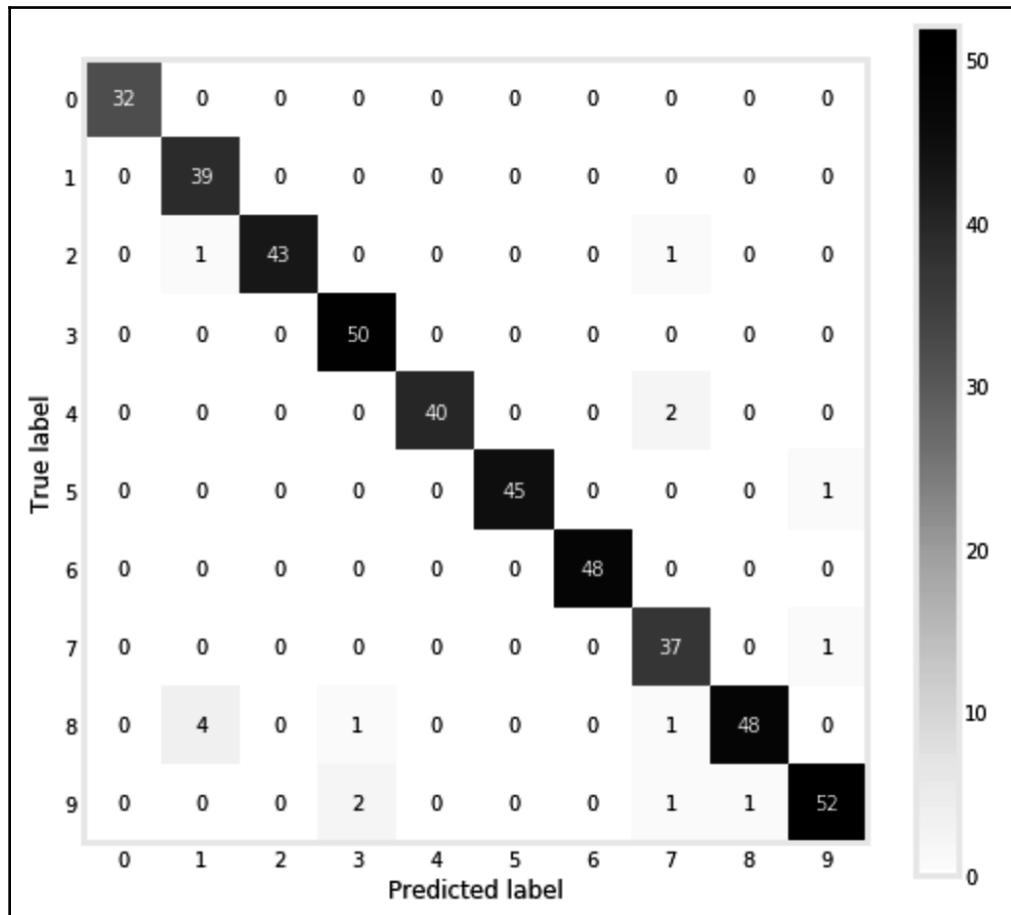
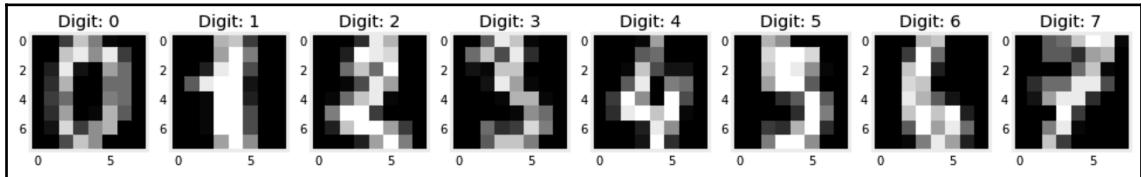
	Size	Brand	Size [Ordinal Encoded]
4	XL	Nike	5.0
5	S	Reebok	2.0
1	S	Reebok	2.0
2	XS	Le Coq	1.0
8	XXL	Adidas	-1.0

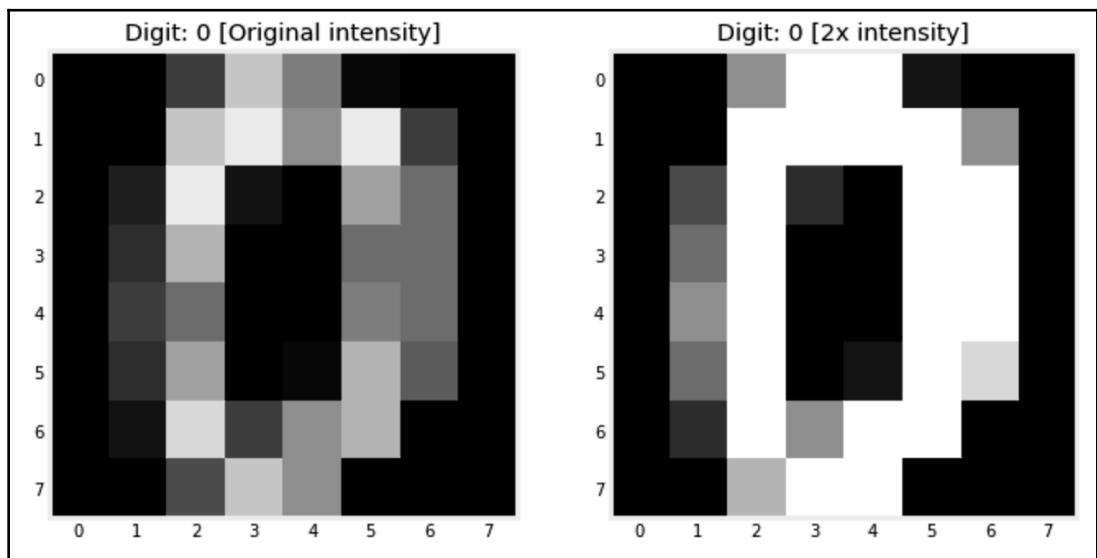
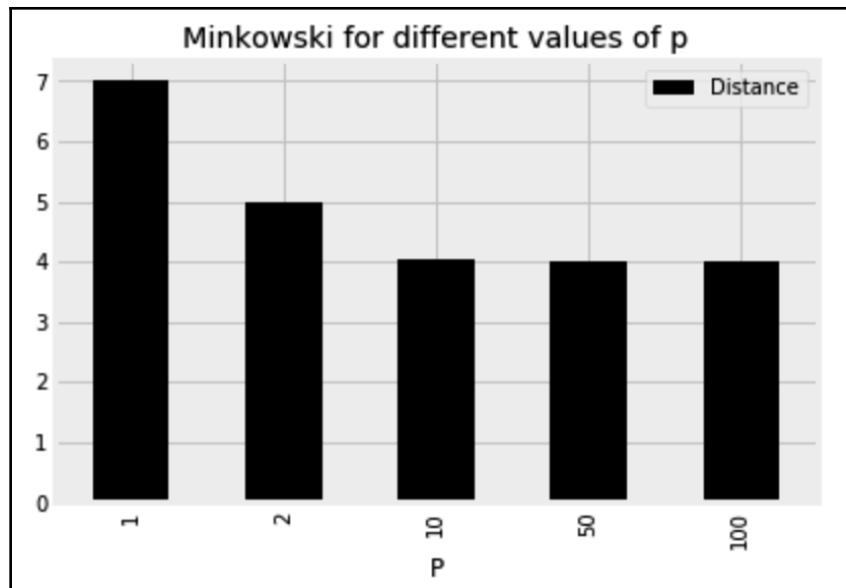
	Feature	Variance
7	nonflavanoid_phenols	0.013845
10	hue	0.055709
2	ash	0.068327
8	proanthocyanins	0.306081
5	total_phenols	0.369353
11	od280/od315_of_diluted_wines	0.478116
0	alcohol	0.679757
6	flavanoids	0.925865
1	malic_acid	0.987016
9	color_intensity	5.612742
3	alcalinity_of_ash	8.914152
4	magnesium	92.000000
12	proline	1402.000000

	Feature	F	MI
6	flavanoids	119.566	0.685479
9	color_intensity	74.0814	0.622413
12	proline	133.289	0.53515
11	od280/od315_of_diluted_wines	90.6593	0.508954
10	hue	66.7377	0.461271
0	alcohol	82.6288	0.453612
5	total_phenols	54.6519	0.392575
1	malic_acid	23.0863	0.333107
4	magnesium	7.65196	0.27917
3	alcalinity_of_ash	25.8039	0.248394
8	proanthocyanins	15.1593	0.22716
7	nonflavanoid_phenols	9.79048	0.127077
2	ash	8.69691	0.126323

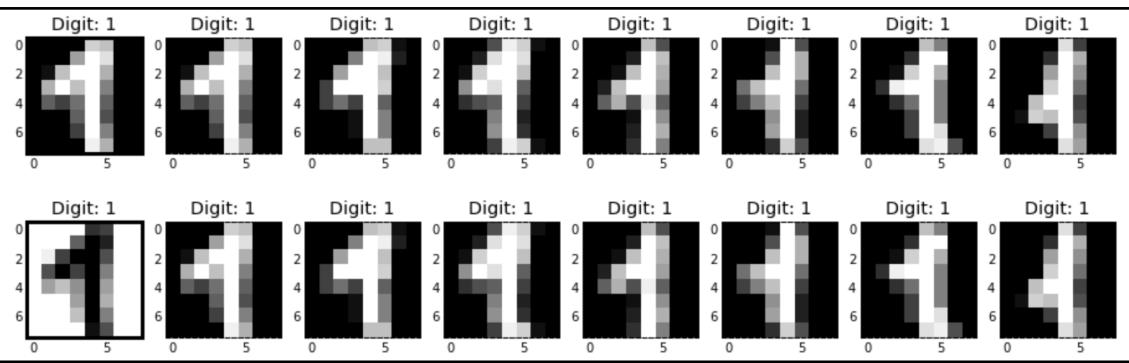
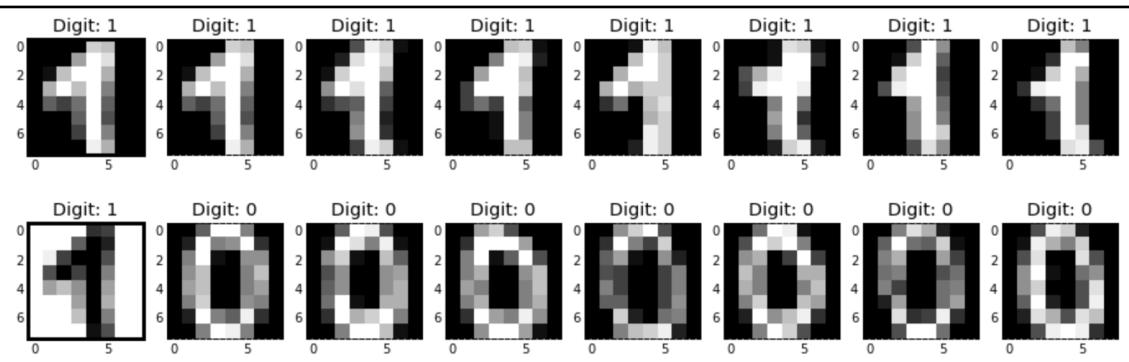
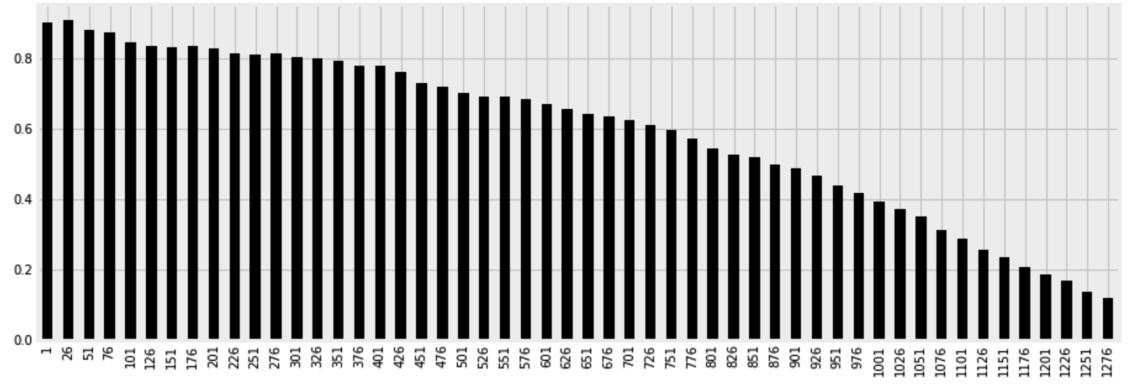
	Logistic Regression	Decision Tree Classifier
SelectKBest(f_classif)	0.97	0.89
SelectKBest(mutual_info_classif)	0.9	0.93

Chapter 5: Image Processing with Nearest Neighbors

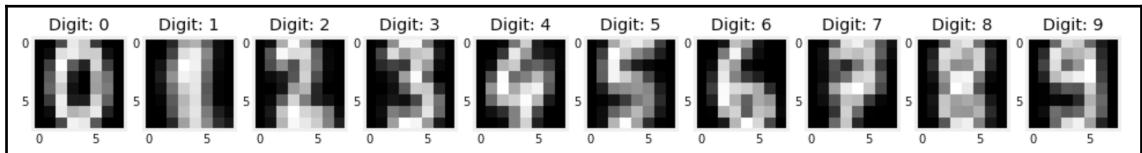




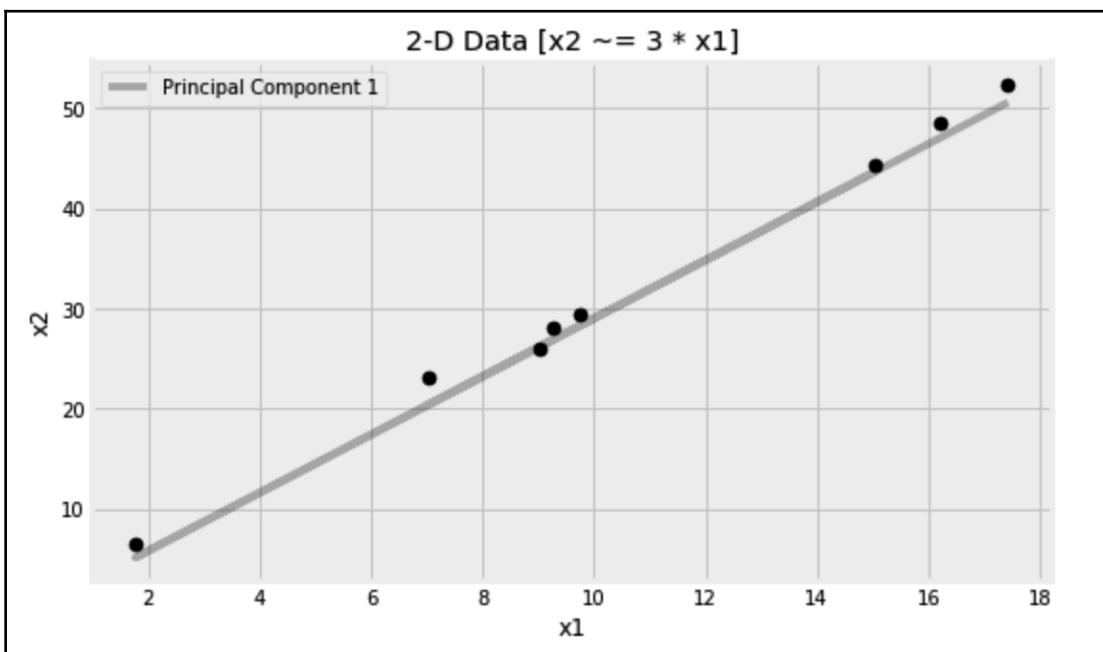
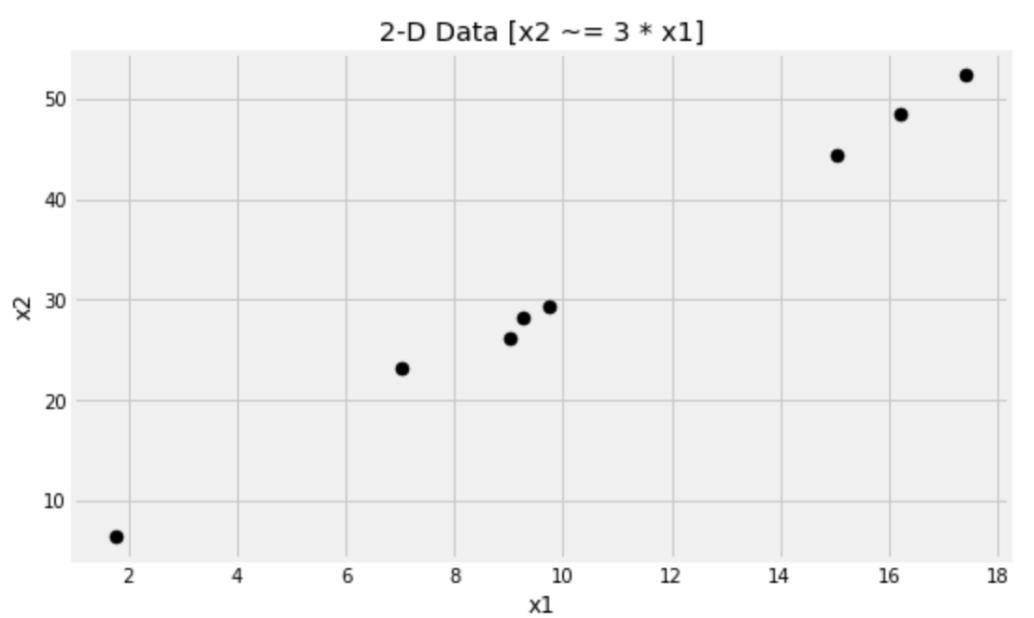
Accuracy versus K

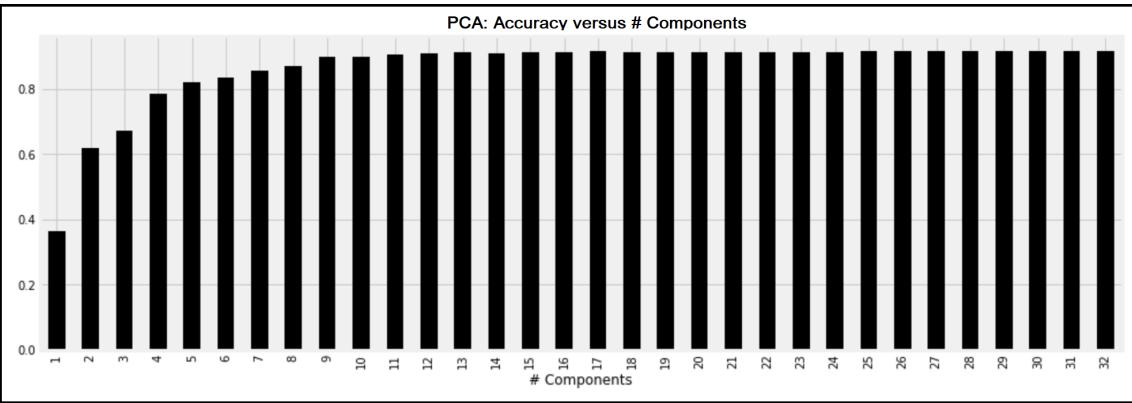
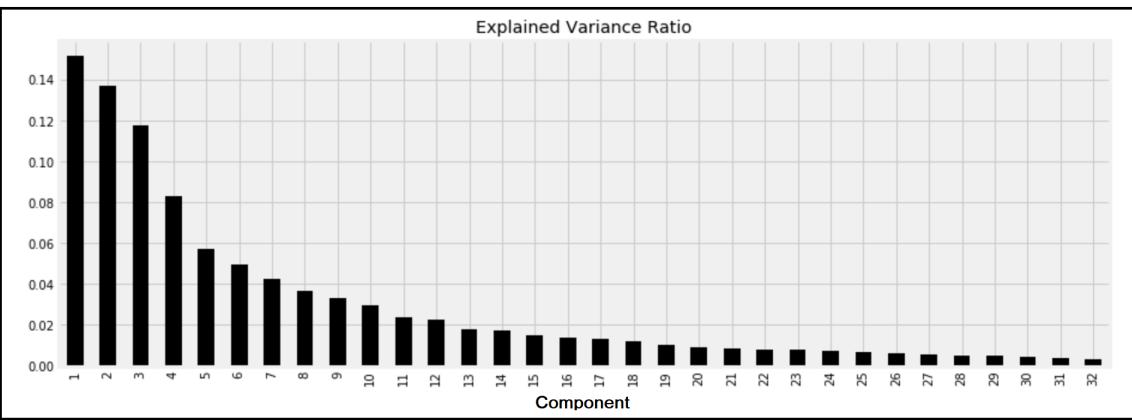
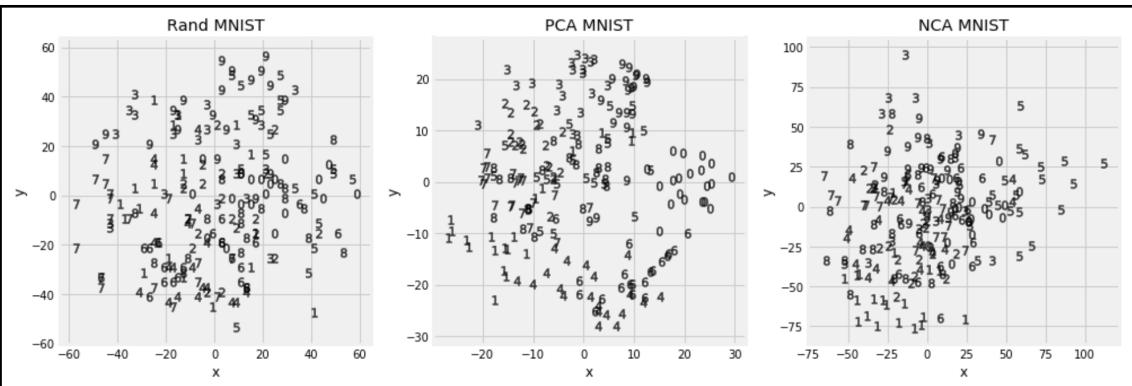


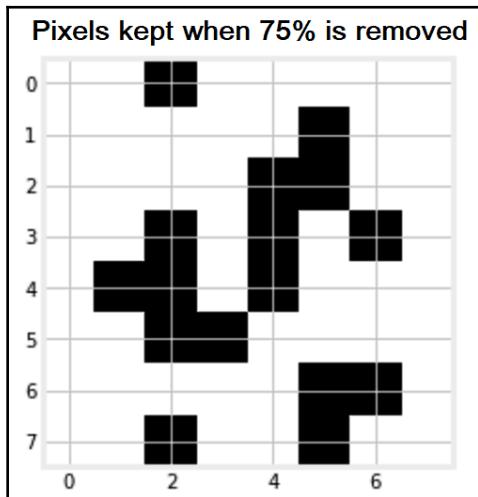
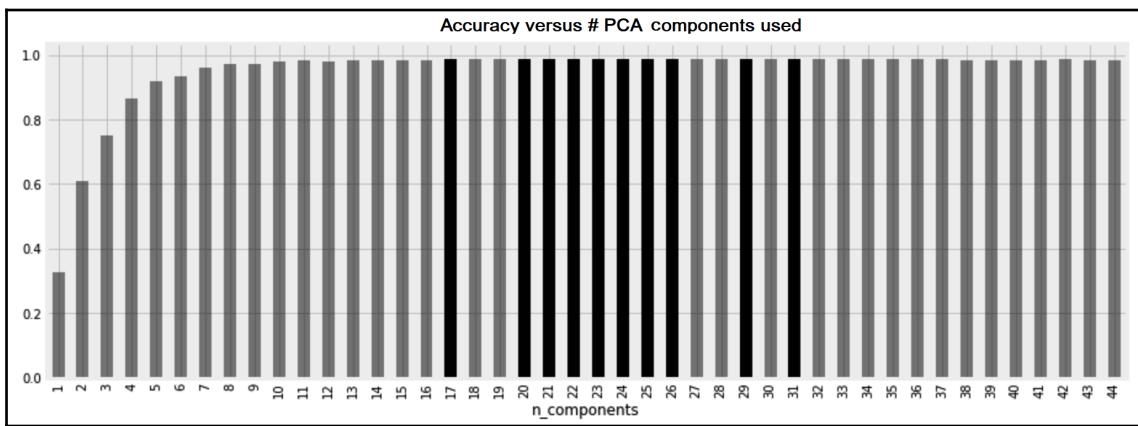
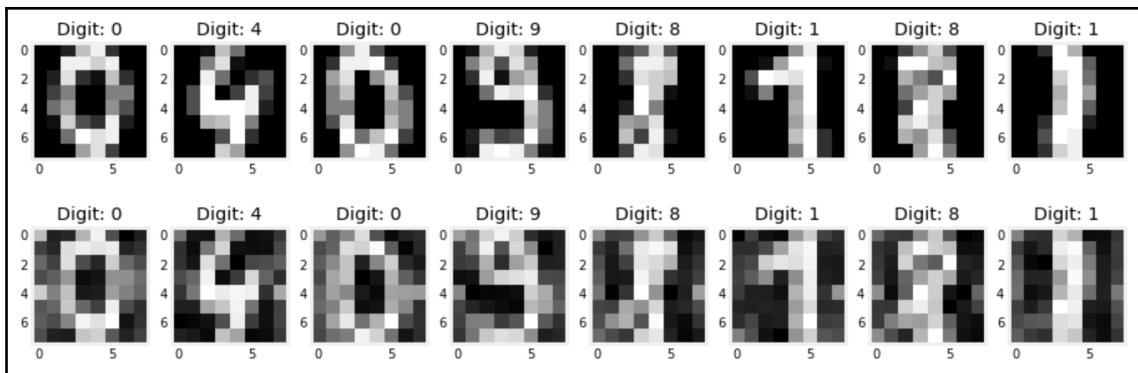
y_test	y_test_pred
117	3 3.666667
215	8 8.333333
304	7 7.666667



	x1	x2
0	9.259916	28.129327
1	7.021191	23.122587
2	9.744407	29.393867
3	15.047500	44.300523
4	1.782589	6.445454
5	9.014174	26.074613
6	16.200623	48.464456
7	17.430507	52.386115







Chapter 6: Classifying Text Using Naive Bayes

doc-id	_hashtag_	_num_	_url_	a	accessible	boss	followers	google	how	is	like	new	to	tokenize	via
0	0	0	0	1	0	1	0	0	1	0	1	0	1	0	1
1	0	0	1	0	1	0	0	1	0	1	0	1	0	0	0
2	1	1	0	0	0	0	1	0	0	0	0	0	1	0	0

doc-id	am	delayed	flight	happy	not	was
0	1	1	1	1	1	1
1	1	1	1	1	1	1

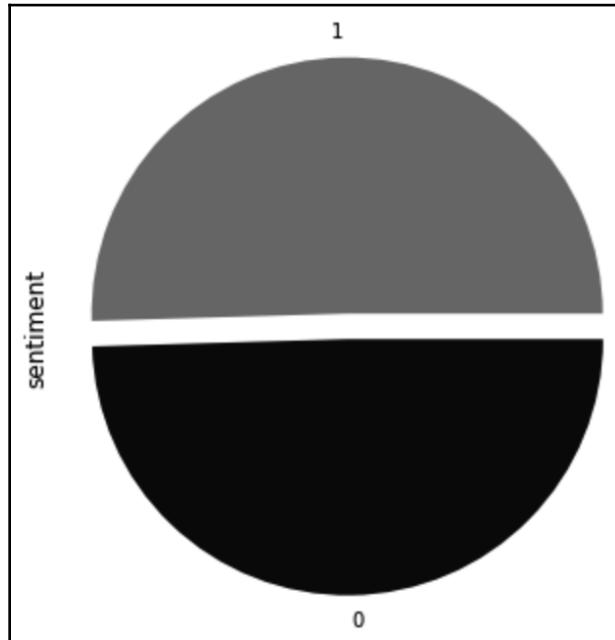
doc-id	am happy	am not	delayed am	flight was	not delayed	not happy	was delayed	was not
0	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1

doc-id	am	del	hap	ia	not	was	,i	am h	am n	appy	...	not	ot d	ot h	s de	s no	t de	tha	twa	was	yed,
0	1	1	1	1	1	1	1	1	0	1	1	...	1	0	1	1	0	0	1	1	1
1	1	1	1	1	1	1	1	1	1	0	1	...	1	1	0	0	1	1	0	1	1

CountVectorizer	apples i like oranges pears						apples i like oranges pears					
	0	1.0	1.0	1.0	0.0	0.0	0	0.77	0.45	0.45	0.00	0.00
1	0.0	1.0	1.0	1.0	1.0	0.0	1	0.00	0.45	0.45	0.77	0.00
2	0.0	1.0	1.0	1.0	0.0	1.0	2	0.00	0.45	0.45	0.00	0.77

	apples	i	like	oranges	pears
CountVectorizer					
0	1.0	1.0	1.0	0.0	0.0
1	0.0	1.0	1.0	1.0	0.0

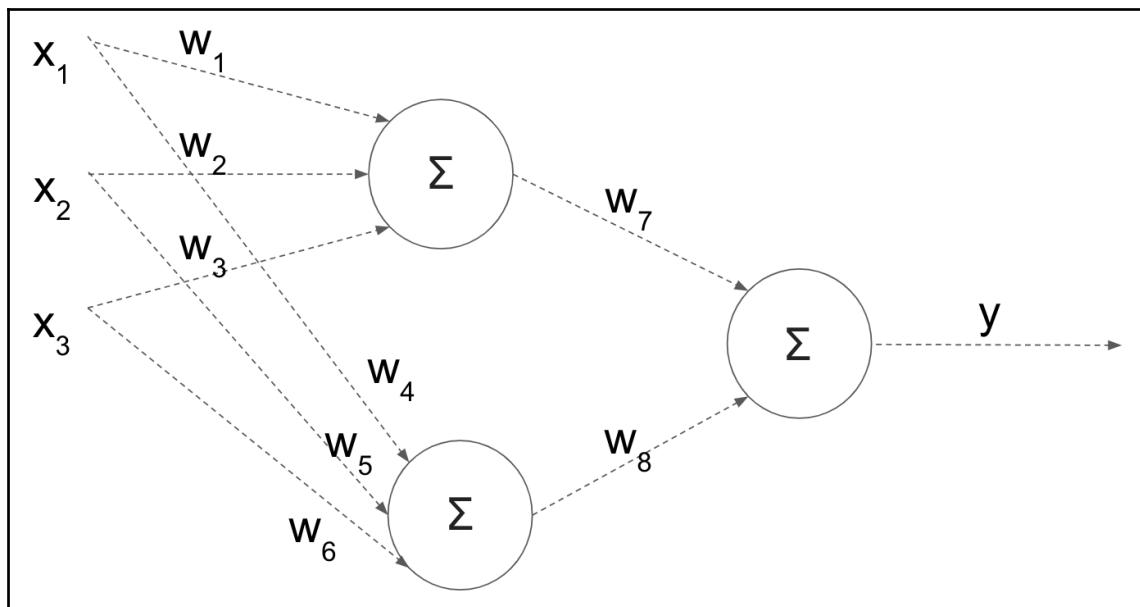
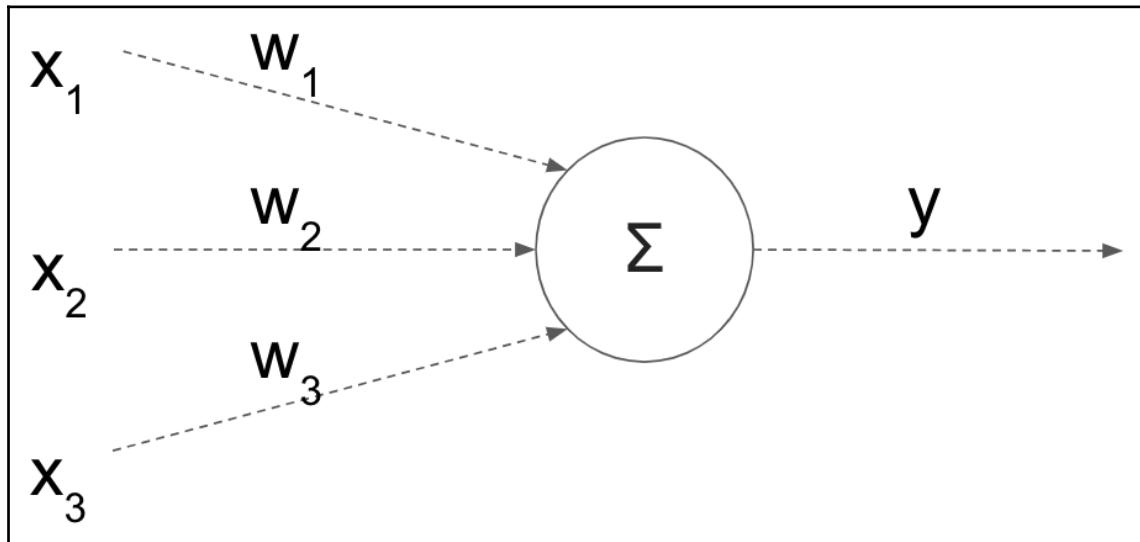
	I	like	apples	oranges	pears
I	1	0.555	0.204	0.188	0.204
like	0.555	1	0.33	0.277	0.33
apples	0.204	0.33	1	0.778	1
oranges	0.188	0.277	0.778	1	0.778
pears	0.204	0.33	1	0.778	1

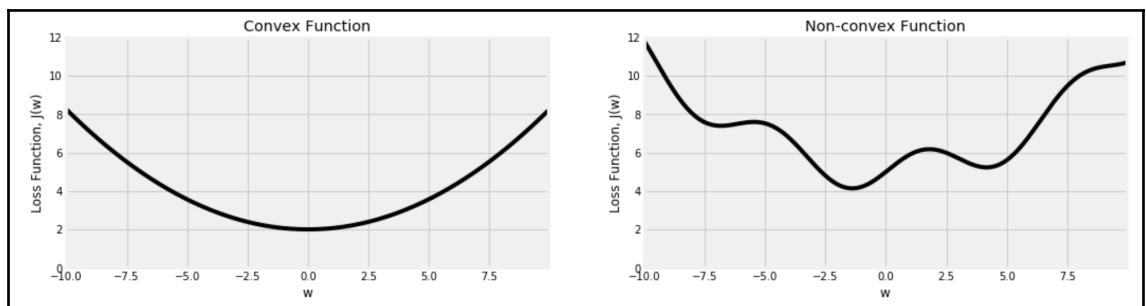
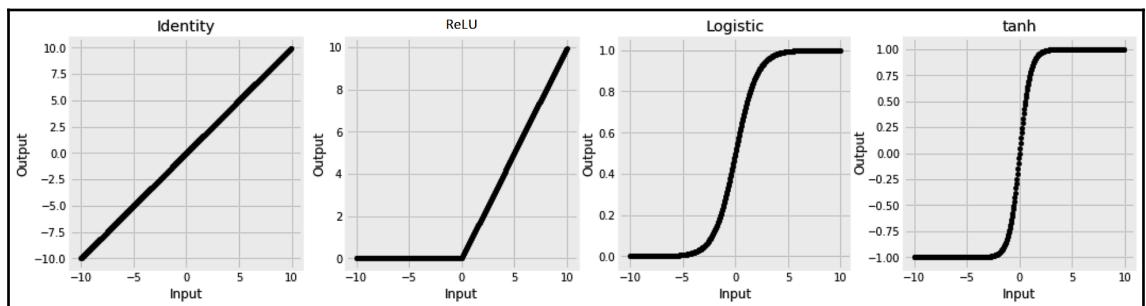
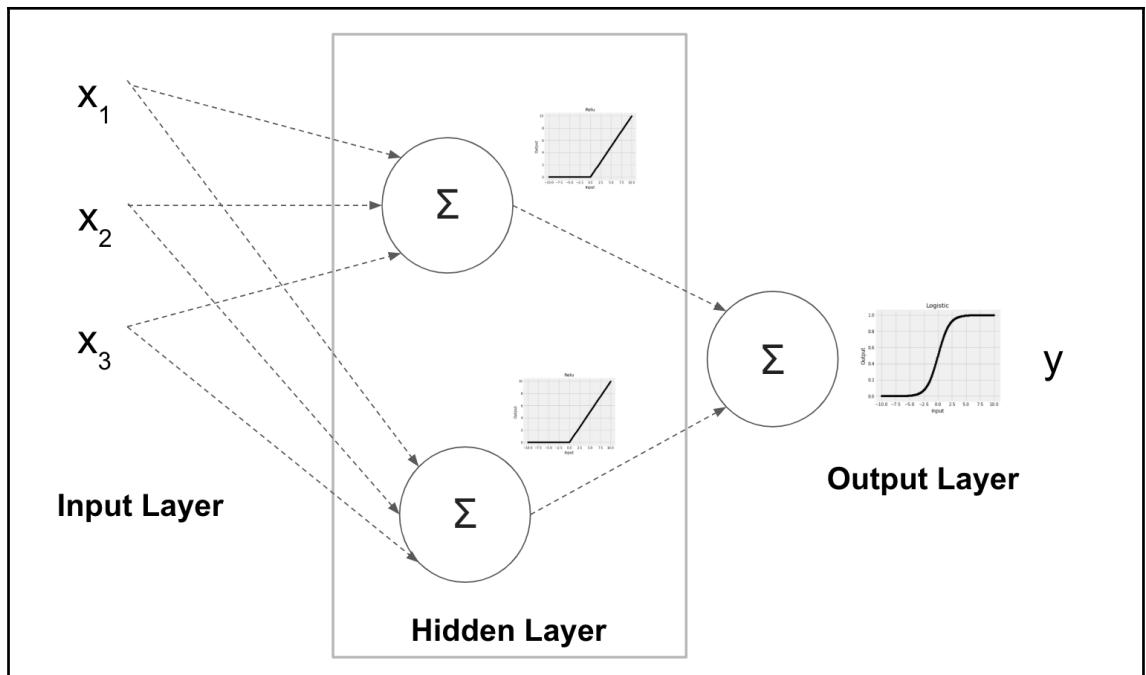


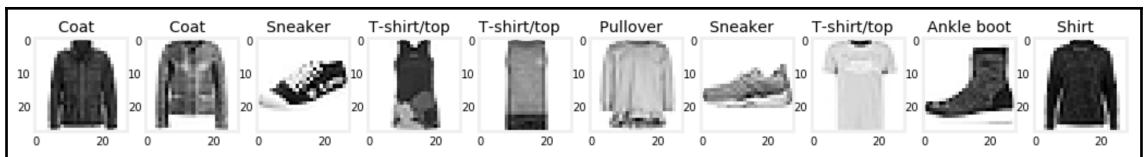
		text	sentiment
471		This is a stunning movie.	1
278	I had the mac salad and it was pretty bland so I will not be getting that again.		0
20		The food, amazing.	1
150		Audio Quality is poor, very poor.	0
430	His acting alongside Olivia De Havilland was brilliant and the ending was fantastic!		1

	Precision	Recall	F	Support
0	0.81	0.78	0.79	565
1	0.77	0.80	0.79	533

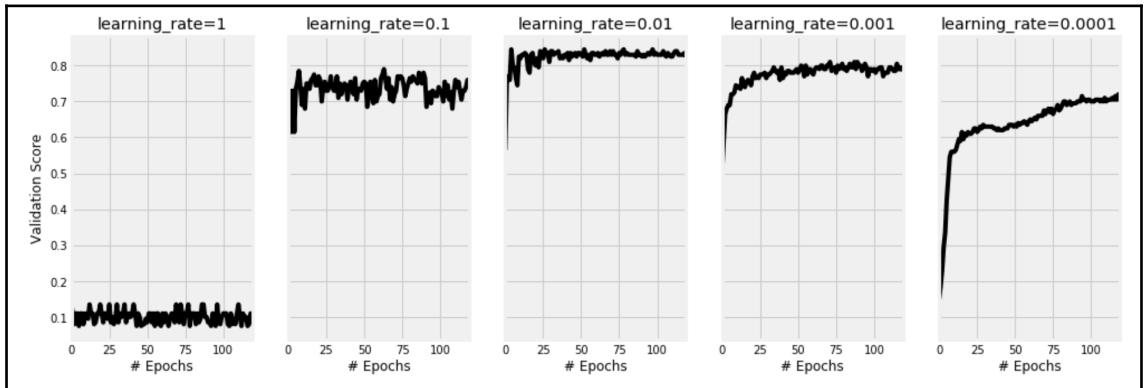
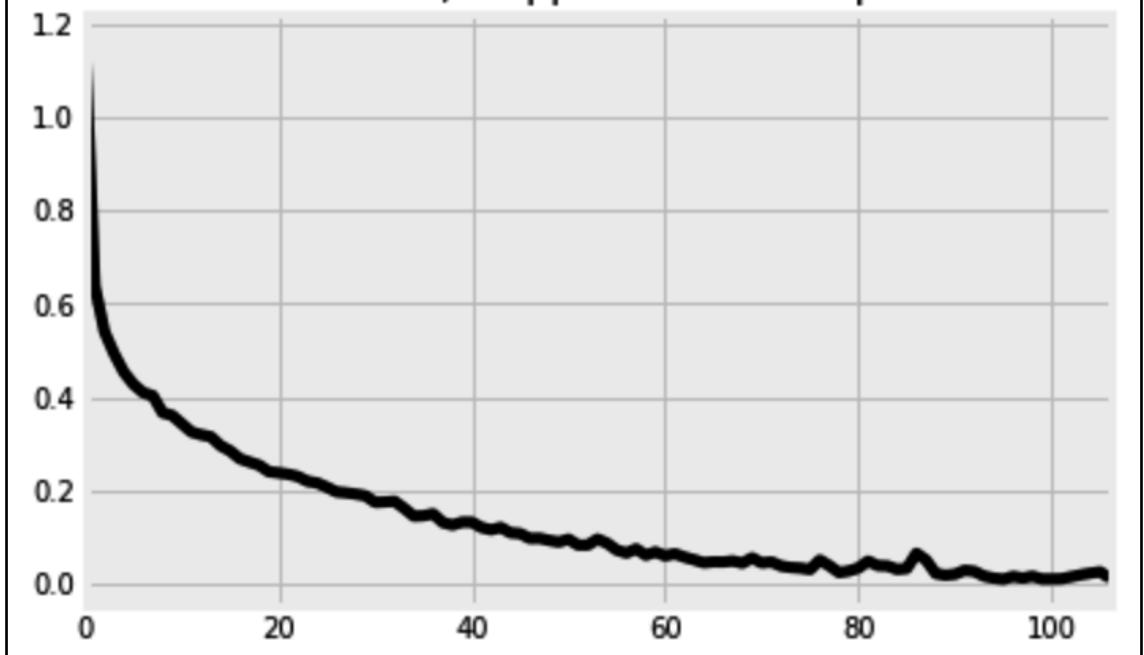
Chapter 7: Neural Networks - Here Comes Deep Learning

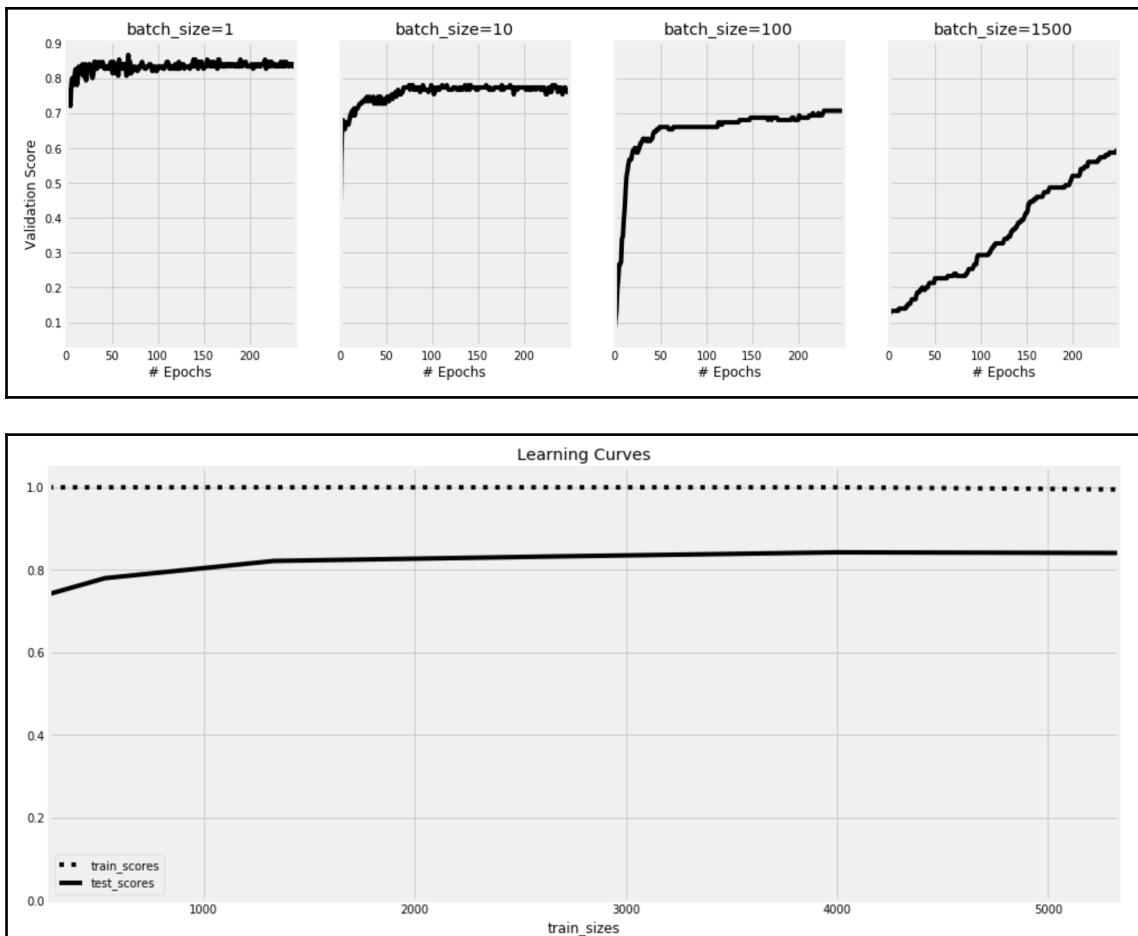


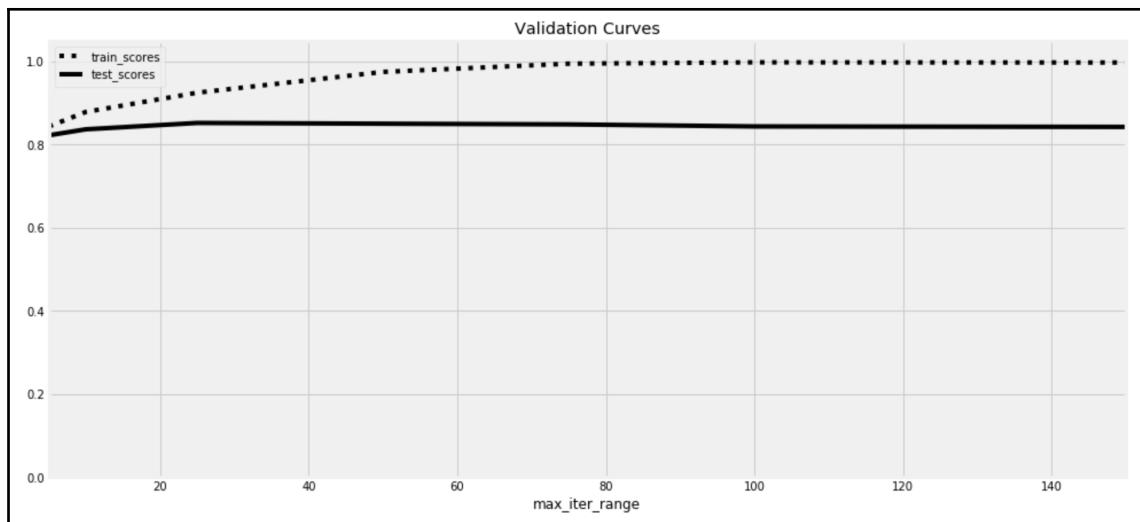




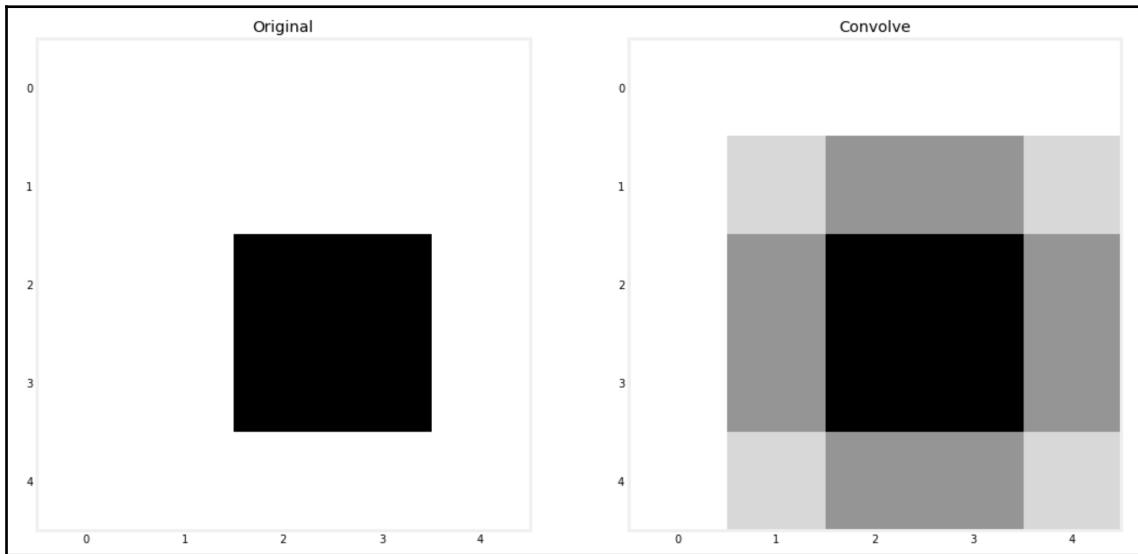
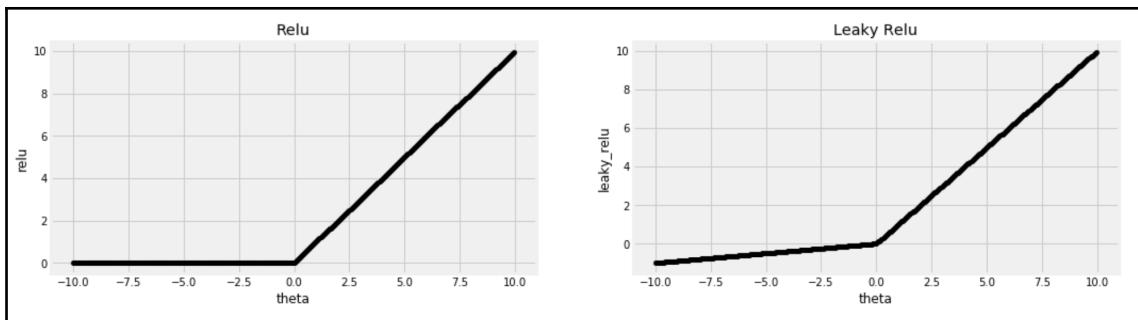
Loss curve; stopped after 107 epochs

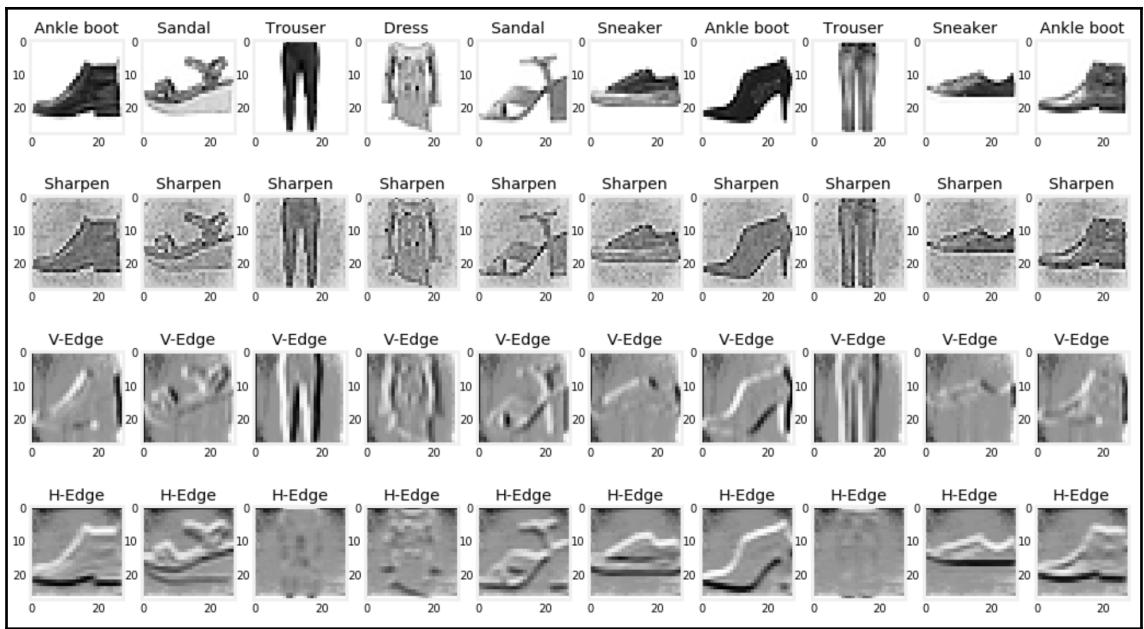






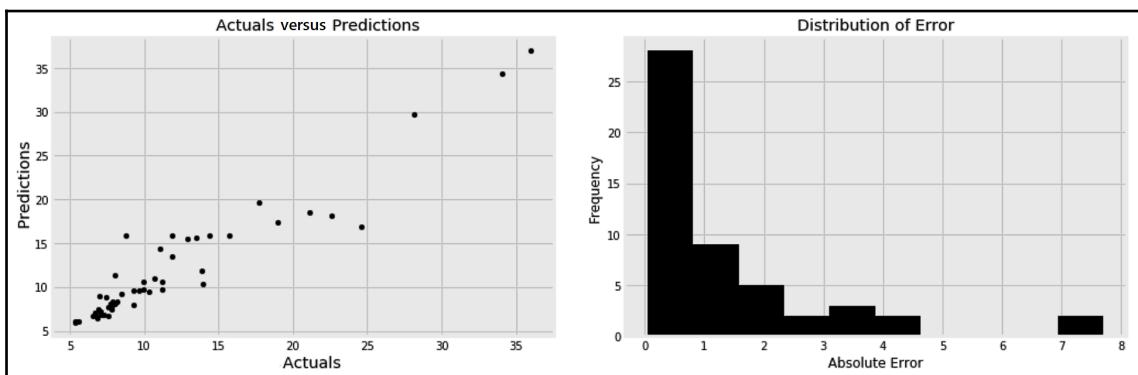
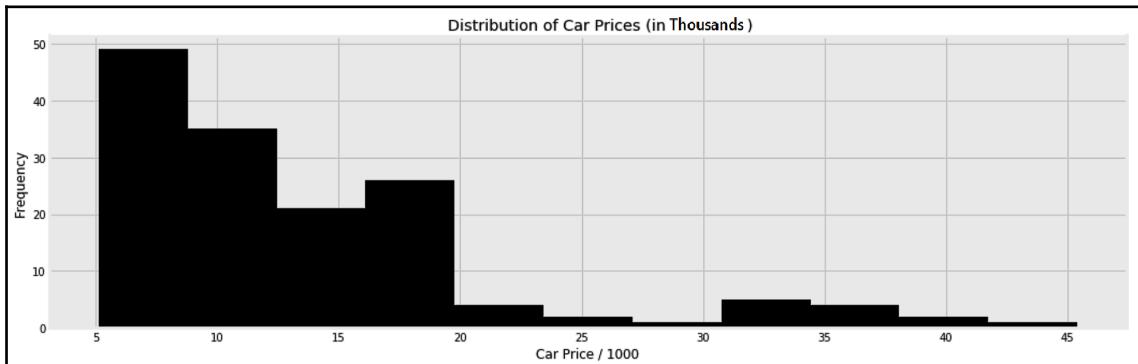
	Precision	Recall	F	Support
Ankle boot	0.93	0.93	0.93	199
Bag	0.98	0.93	0.95	200
Coat	0.84	0.8	0.82	214
Dress	0.88	0.87	0.87	209
Pullover	0.75	0.75	0.75	185
Sandal	0.89	0.93	0.91	202
Shirt	0.64	0.71	0.67	201
Sneaker	0.91	0.91	0.91	210
T-shirt/top	0.79	0.78	0.79	187
Trouser	0.97	0.94	0.96	193

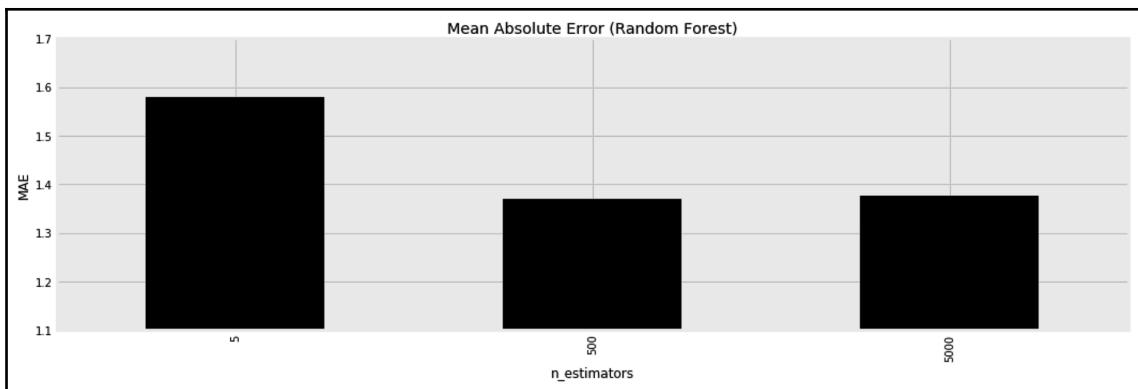




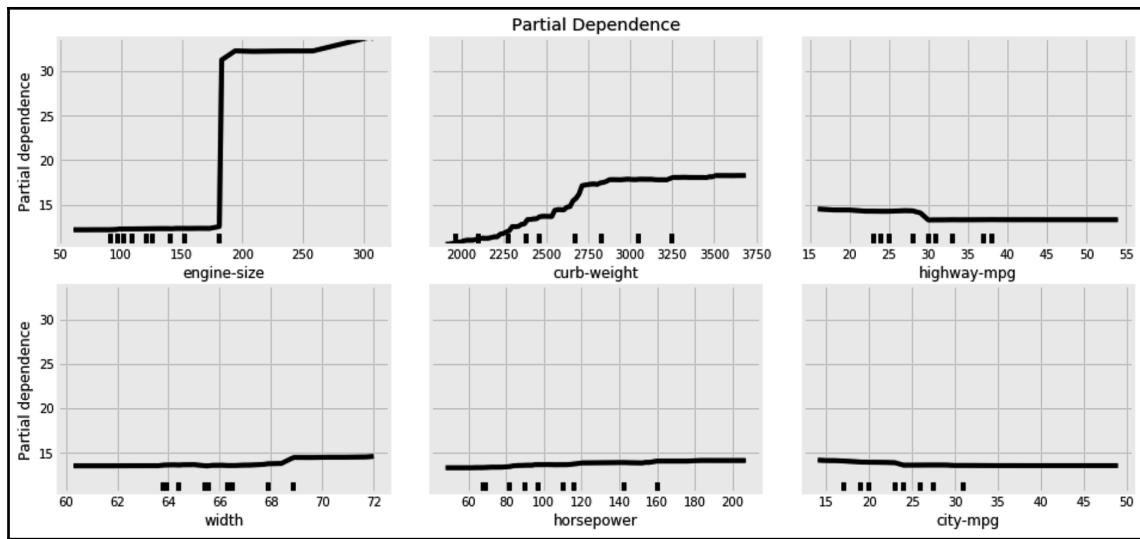
Chapter 8: Ensembles - When One Model Is Not Enough

	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location	engine-type	num-of-cylinders	fuel-system
169	toyota	gas	std	two	hatchback	rwd	front	ohc	four	mpfi
175	toyota	gas	std	four	hatchback	fwd	front	ohc	four	mpfi
46	isuzu	gas	std	two	hatchback	rwd	front	ohc	four	spfi

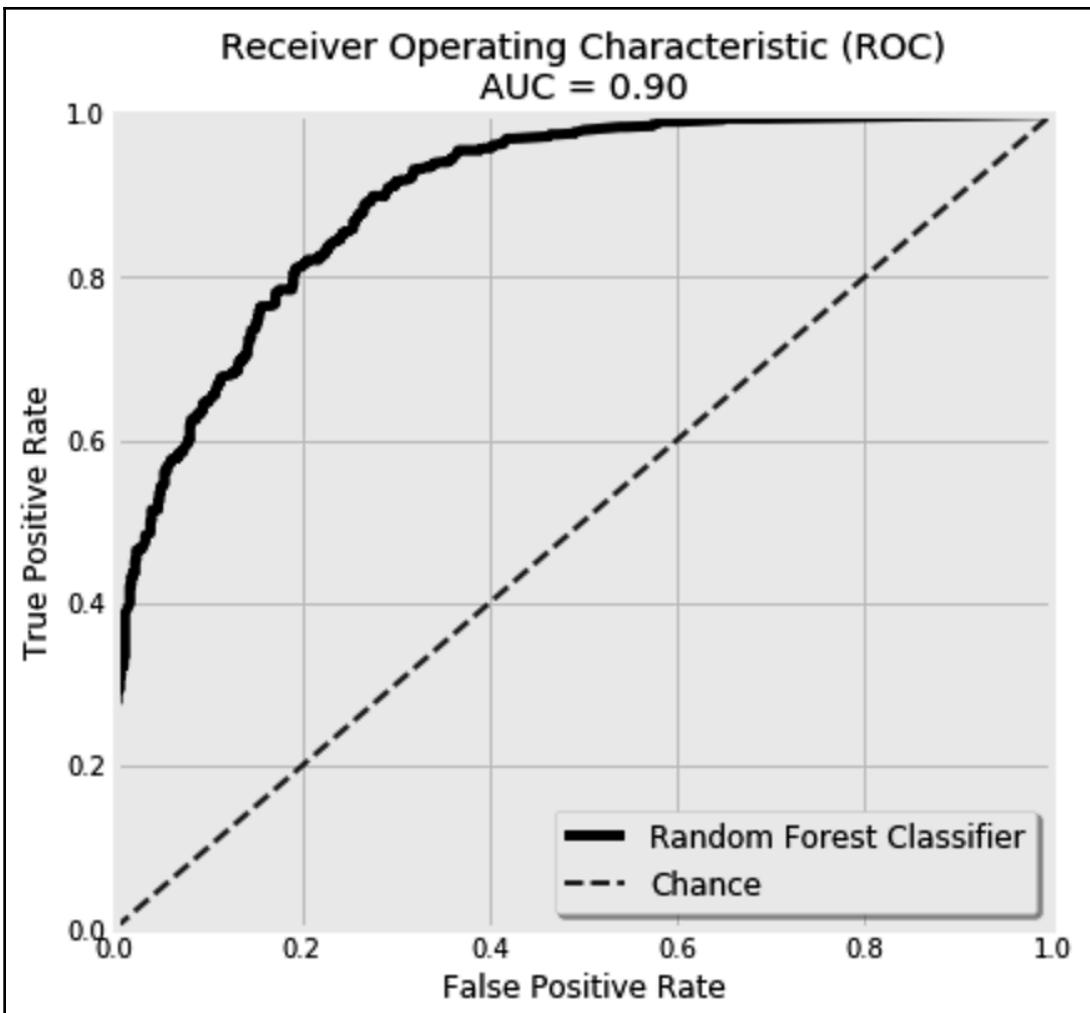


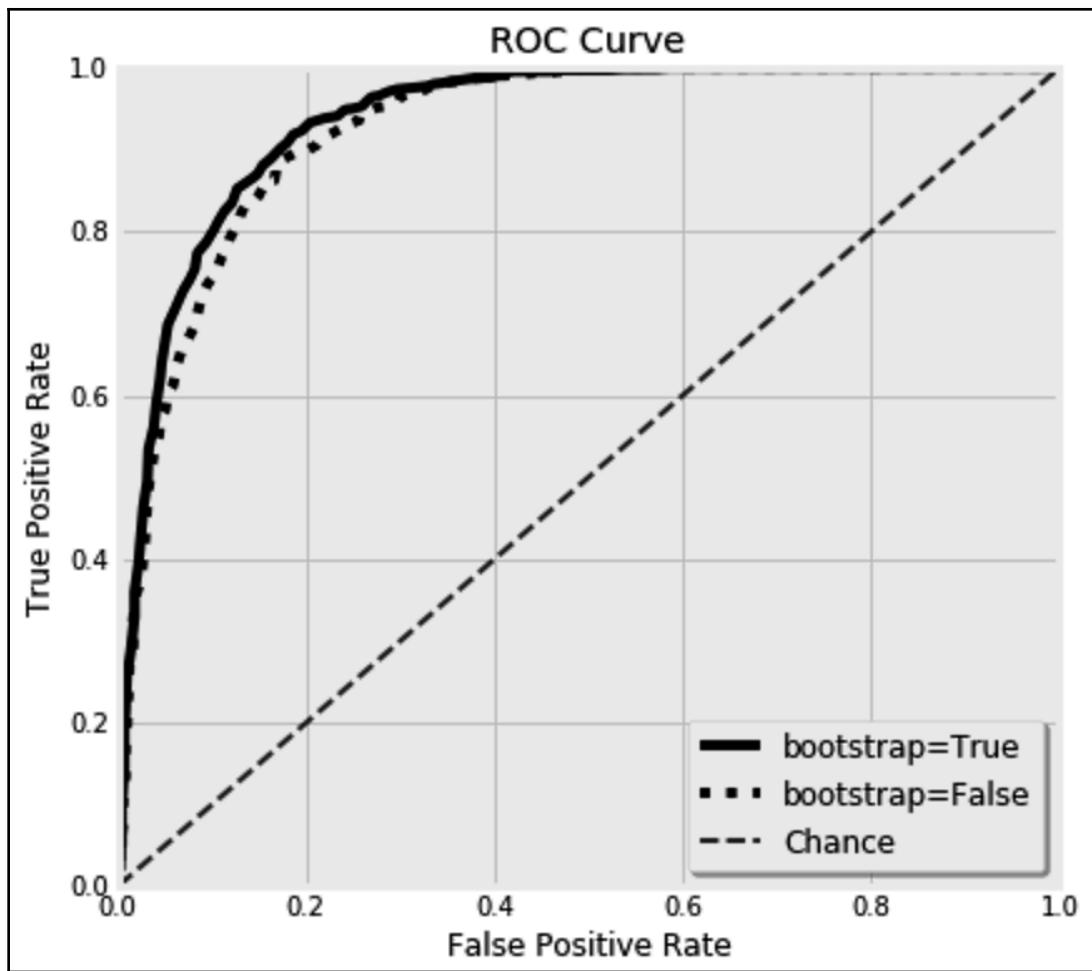


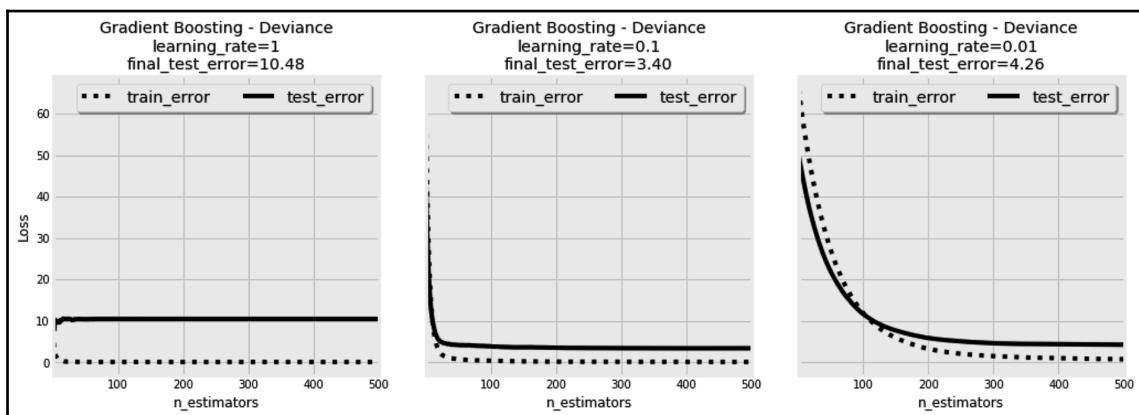
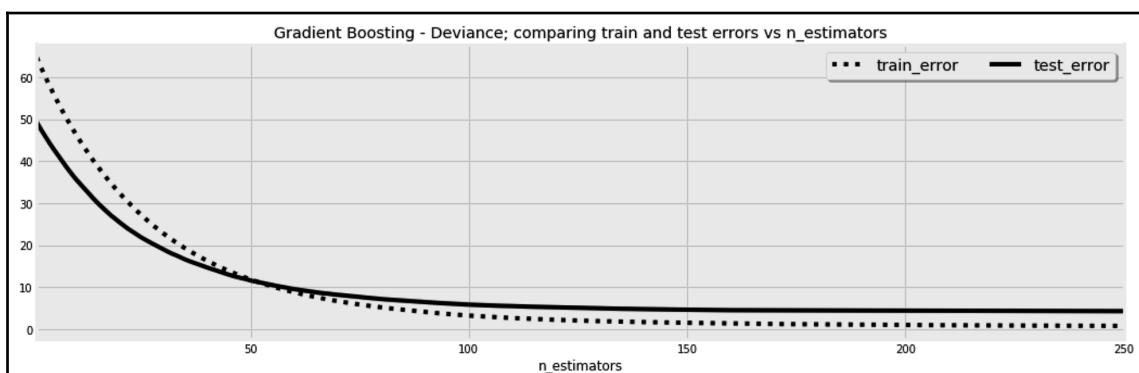
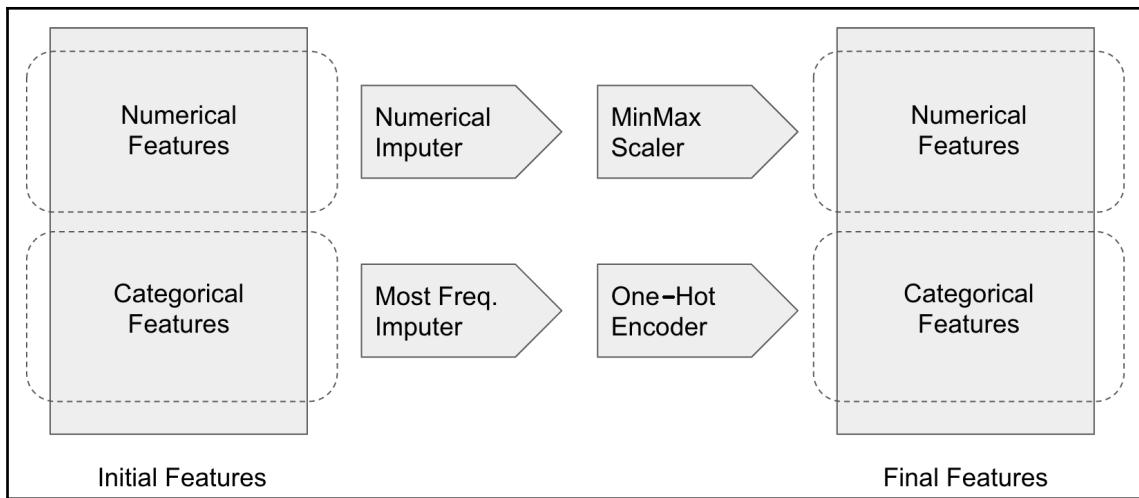
	Feature	Importance
16	engine-size	0.508621
15	curb-weight	0.223342
20	horsepower	0.064801
23	highway-mpg	0.059187
...
10	symboling	0.001612
7	engine-type	0.001577
1	fuel-type	0.000707
6	engine-location	0.000643

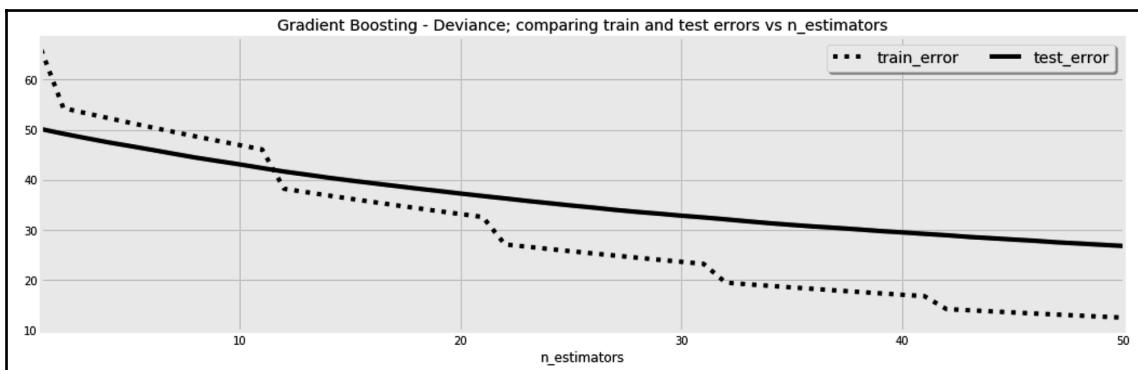
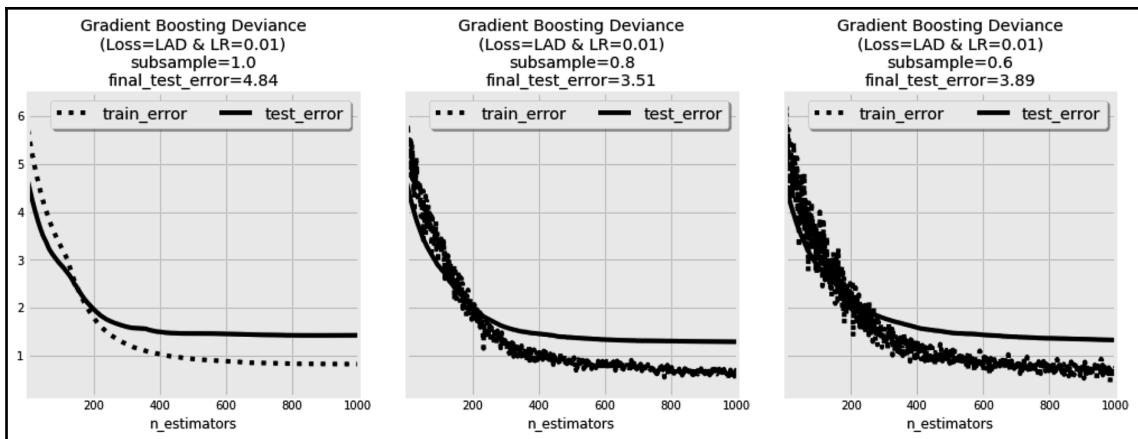


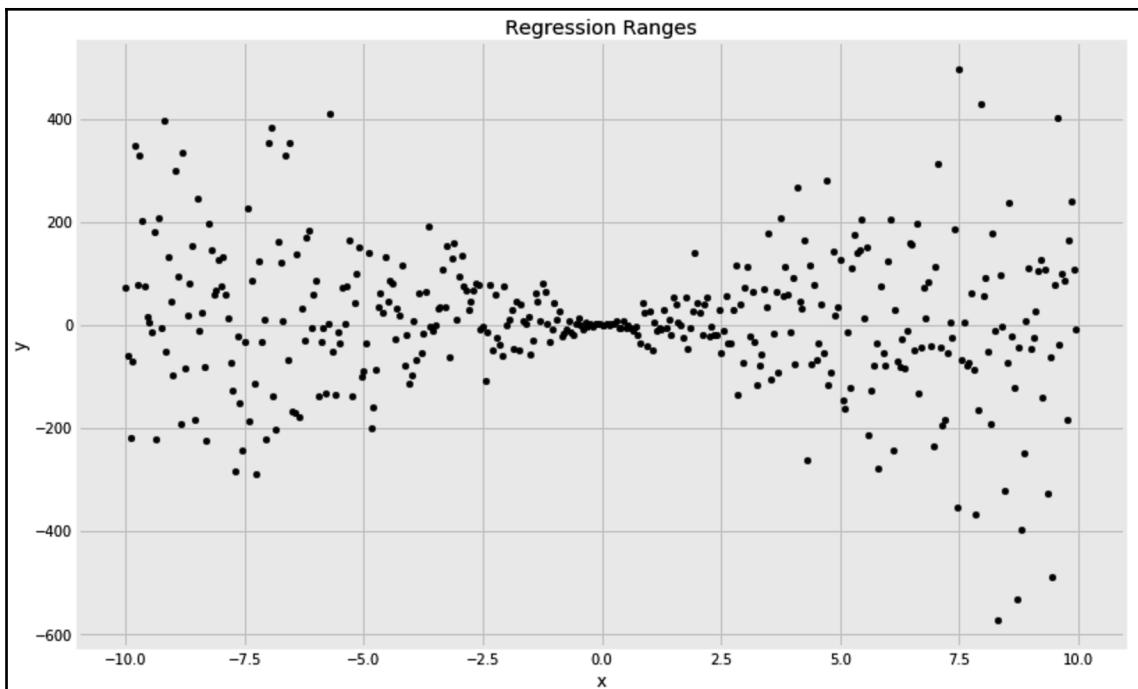
Threshold	True Positive Rate	False Positive Rate
0.710364	0.001305	0.000000
0.549833	0.259791	0.000000
0.549220	0.262402	0.000000
0.547614	0.274151	0.000000
0.547542	0.274151	0.001362
...
0.428332	0.994778	0.671662
0.426702	0.994778	0.685286
0.425484	0.994778	0.688011
0.425381	0.994778	0.693460
0.422533	1.000000	1.000000

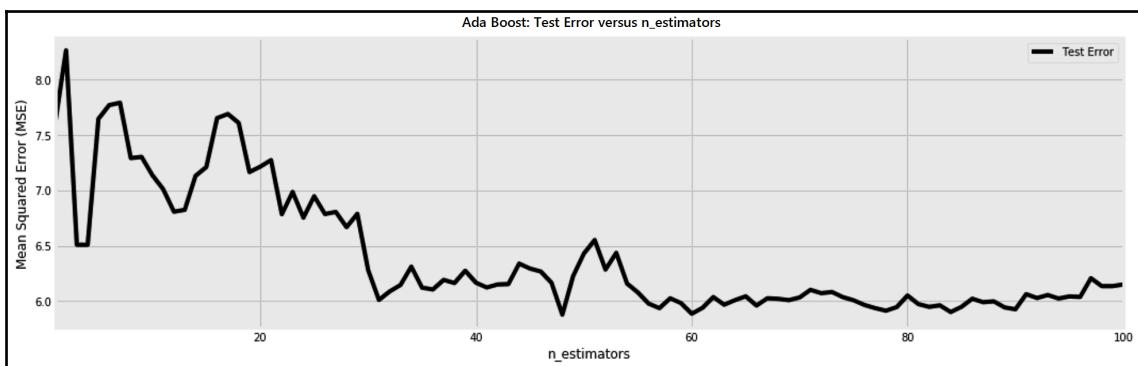
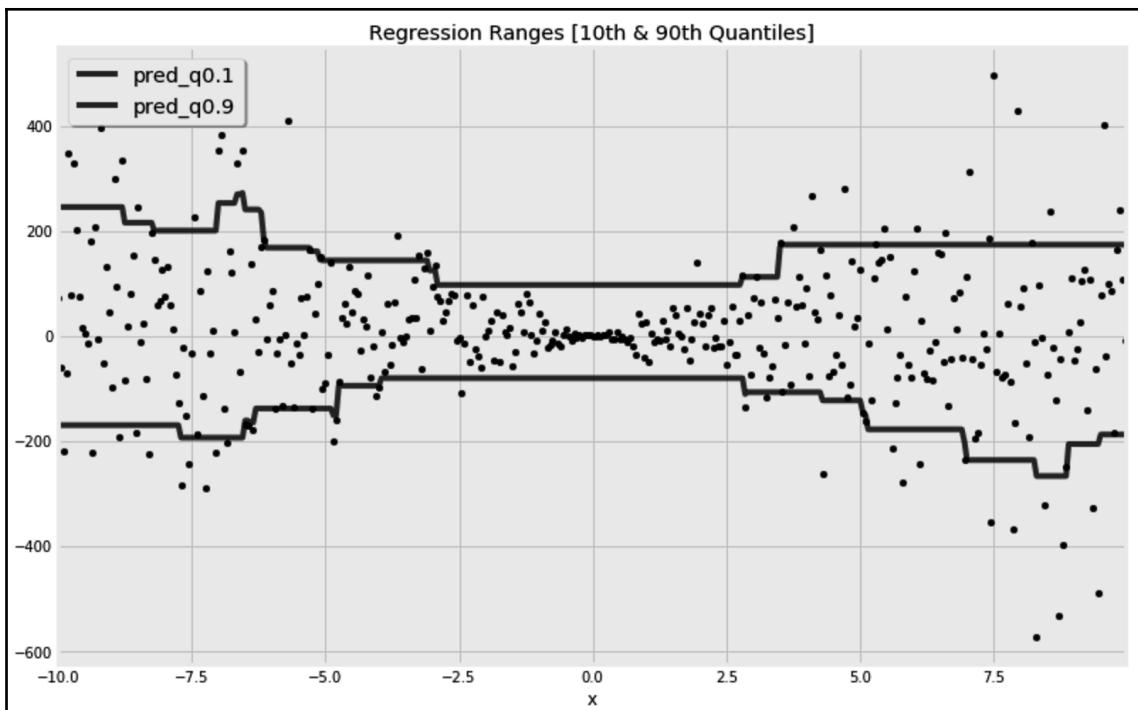




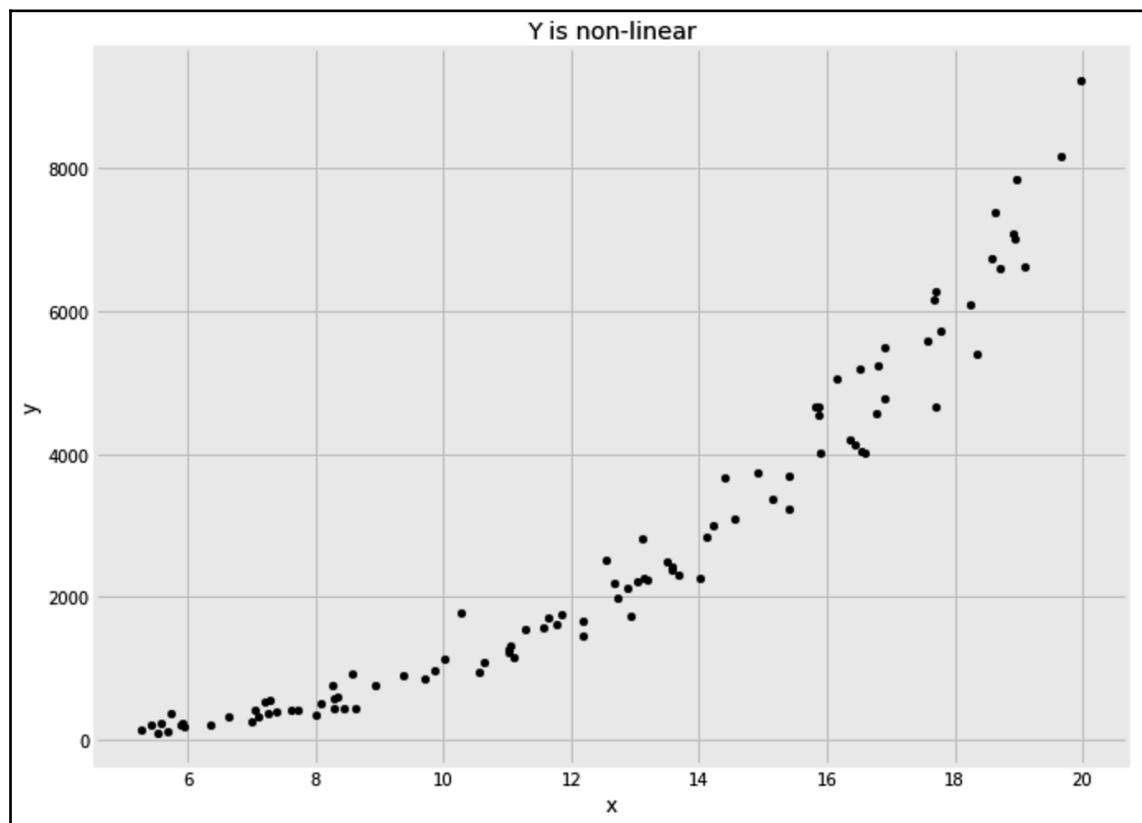






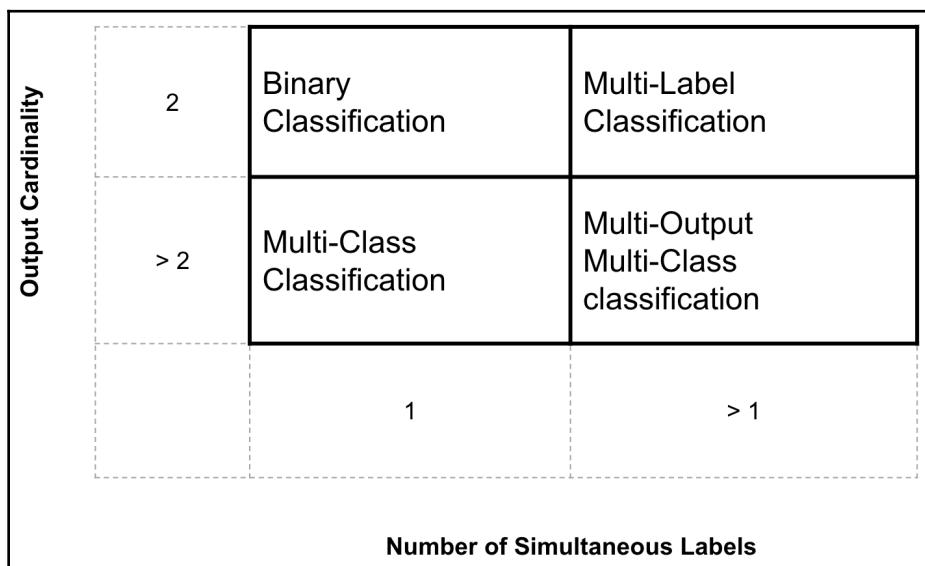


Chapter 9: The Y is as Important as the X

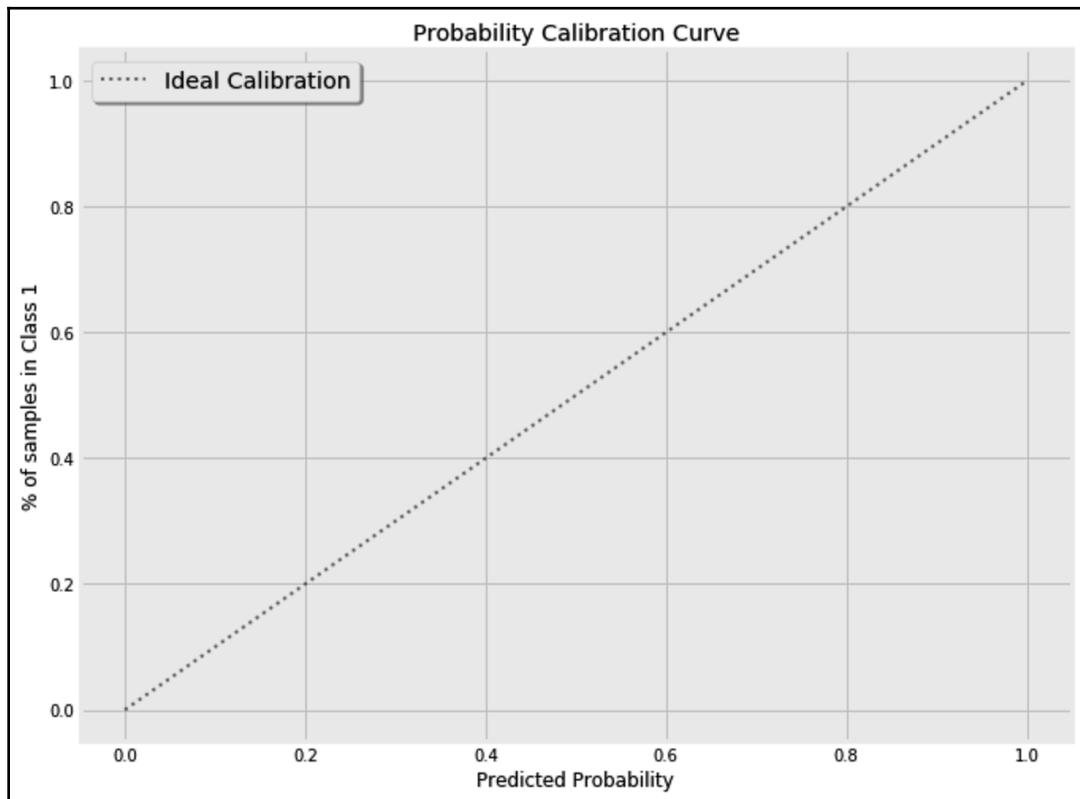


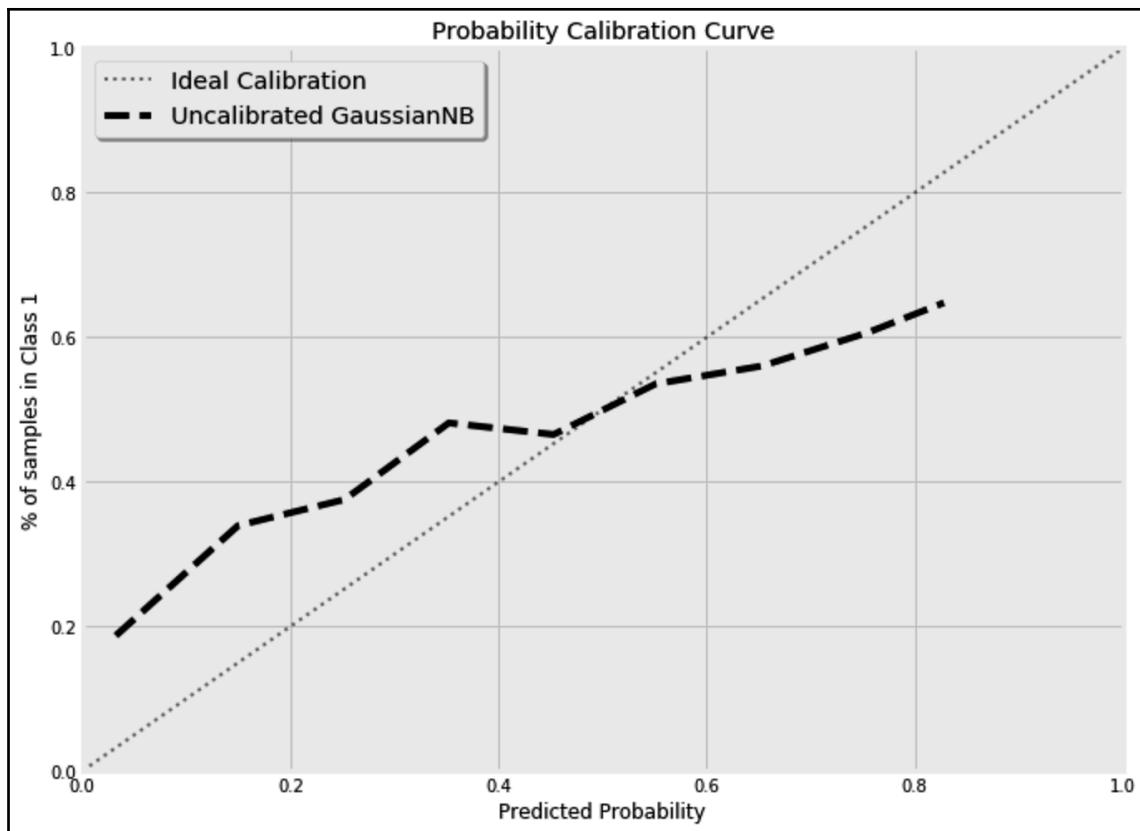
	Target # 0	Target # 1	Target # 2
0	-467.676944	-211.388640	-337.080270
1	-126.855710	-21.526441	-40.367112
2	-16.770193	49.015083	-169.969358
3	-240.225467	-158.054893	-242.705575

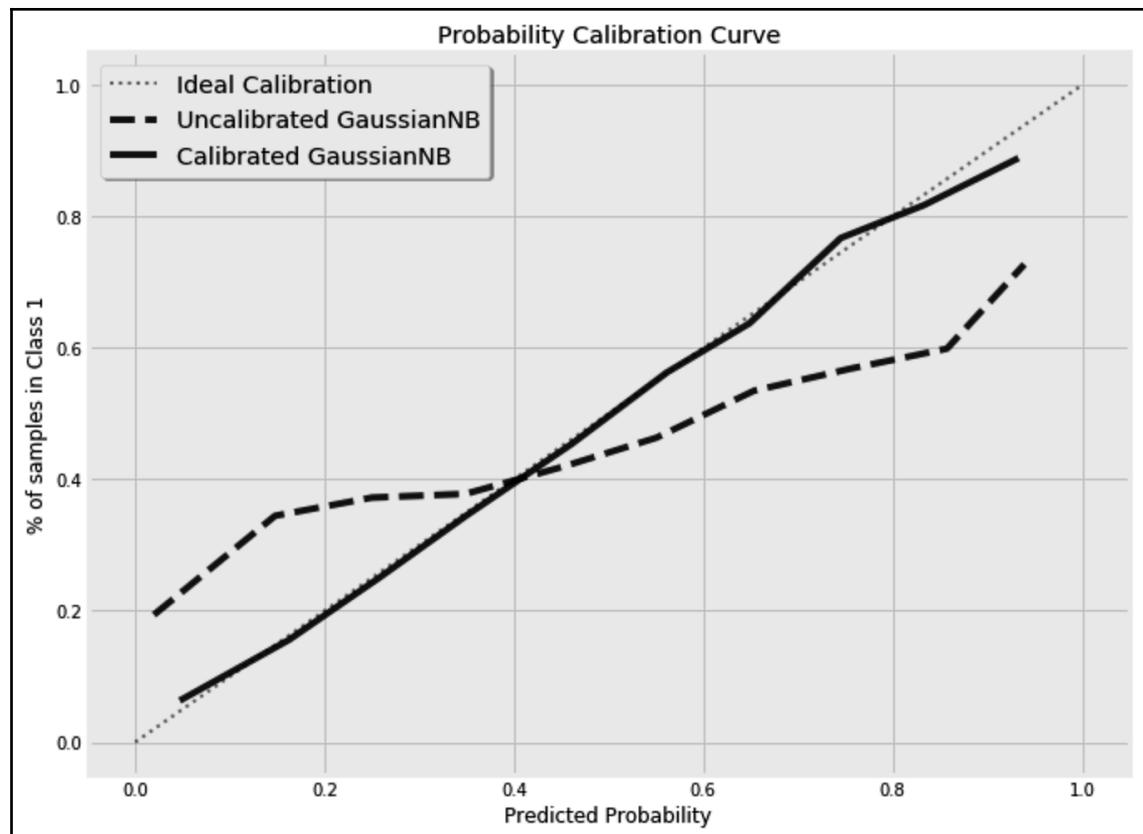
	Feature	Coeff
0	Feature # 0	70.3604
1	Feature # 1	1.83524
2	Feature # 2	92.8441
3	Feature # 3	89.2213
4	Feature # 4	41.7166
5	Feature # 5	70.8224
6	Feature # 6	27.6219
7	Feature # 7	14.9658
8	Target # 0	0.0901208
9	Target # 1	-0.0268364



Feature	Coeff
0 Feature # 0	0
1 Feature # 1	0
2 Feature # 2	0
3 Feature # 3	0
4 Feature # 4	0
5 Feature # 5	0
6 Feature # 6	0
7 Feature # 7	0
8 Target # 0	1
9 Target # 1	0

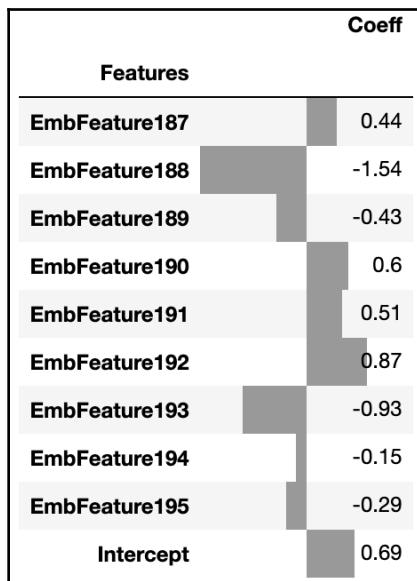




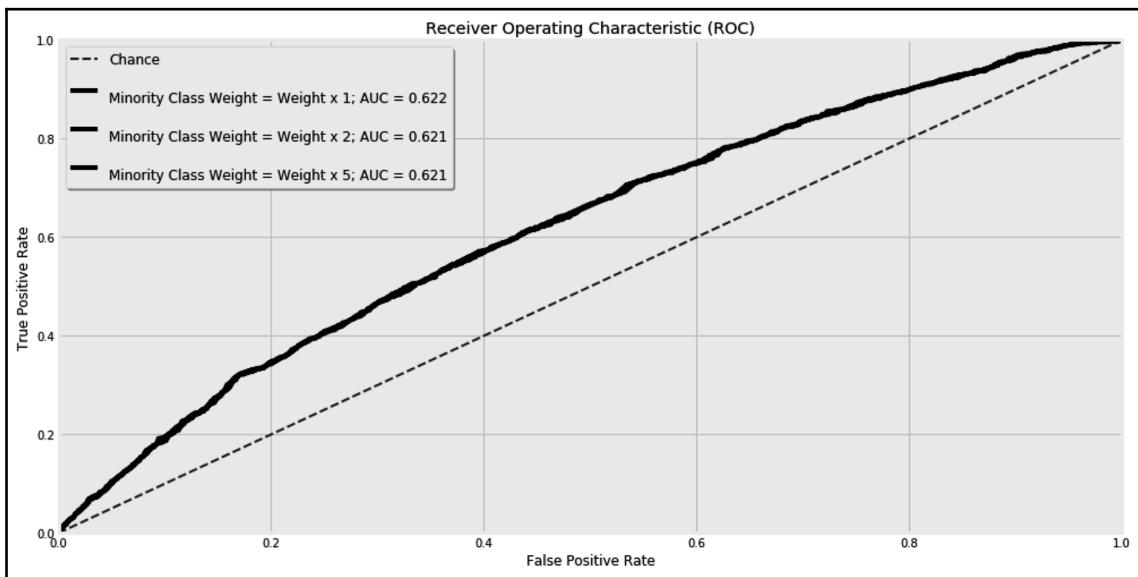


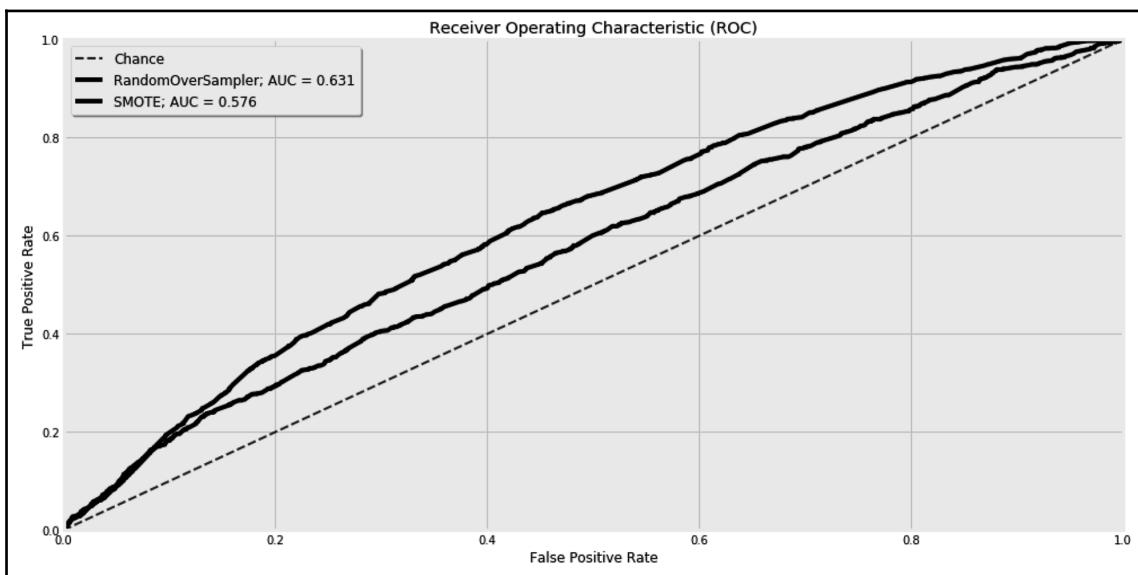
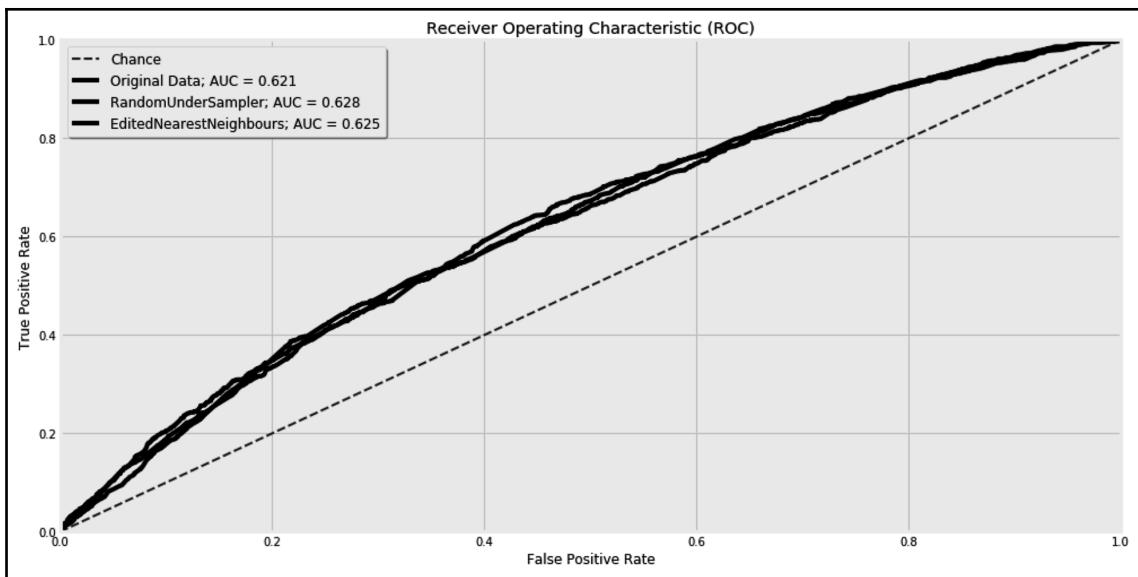
Chapter 10: Imbalanced Learning - Not Even 1% Win the Lottery

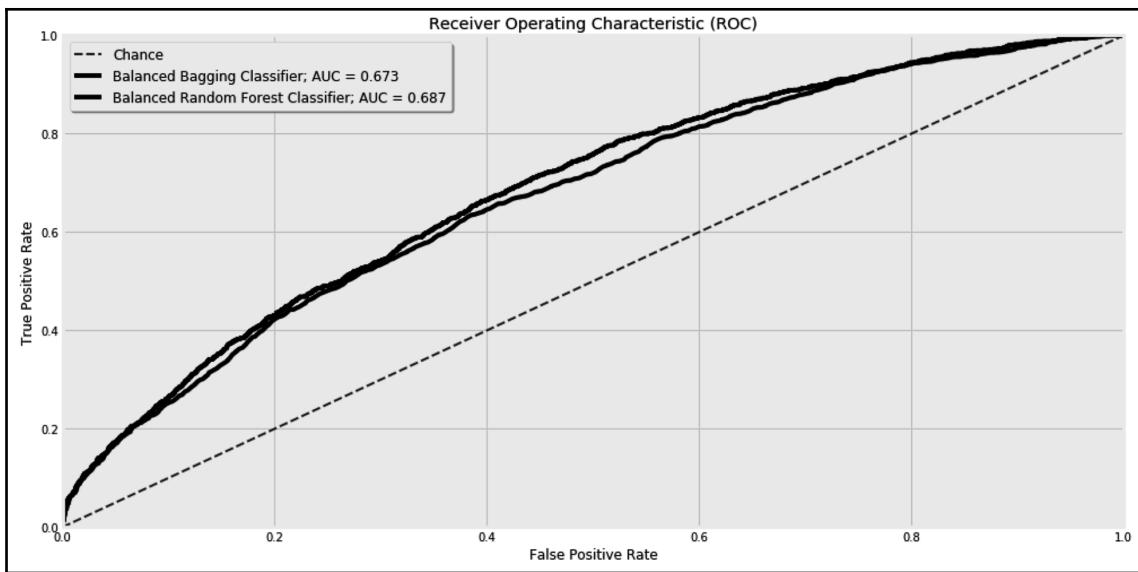
	impression	ad_id	advertiser_id	depth	position	keyword_id	title_id	description_id	user_id	target
12831	1.0	20884487.0	22234.0	1.0	1.0	19303.0	83270.0	74248.0	70.0	0
34147	29.0	10593104.0	1268.0	2.0	1.0	3275.0	2136.0	128.0	0.0	1
7656	1.0	21319021.0	20551.0	2.0	2.0	4766.0	15510.0	15125.0	11569401.0	0
8659	1.0	3831882.0	27486.0	2.0	1.0	1007.0	2869.0	387.0	12581245.0	0
17460	1.0	21442048.0	37039.0	3.0	1.0	35349.0	399482.0	328000.0	0.0	0



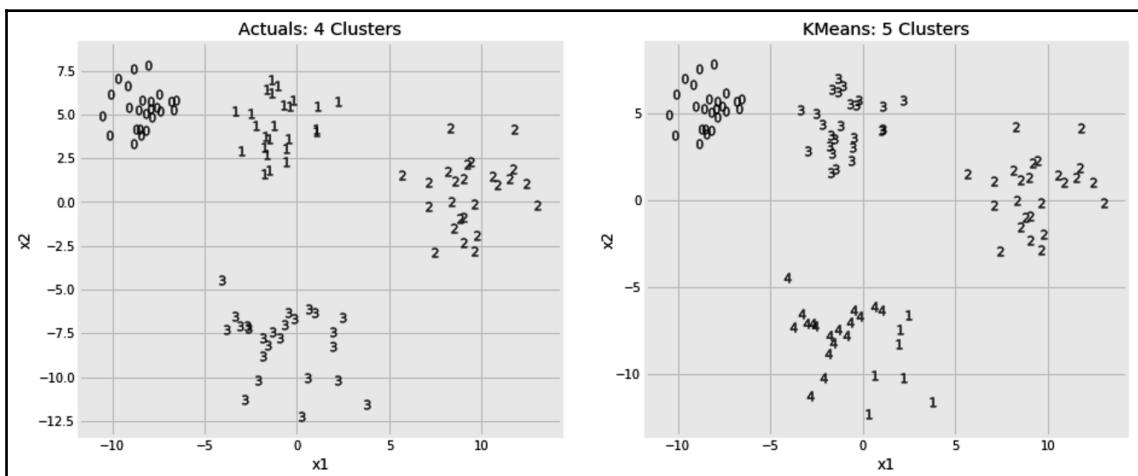
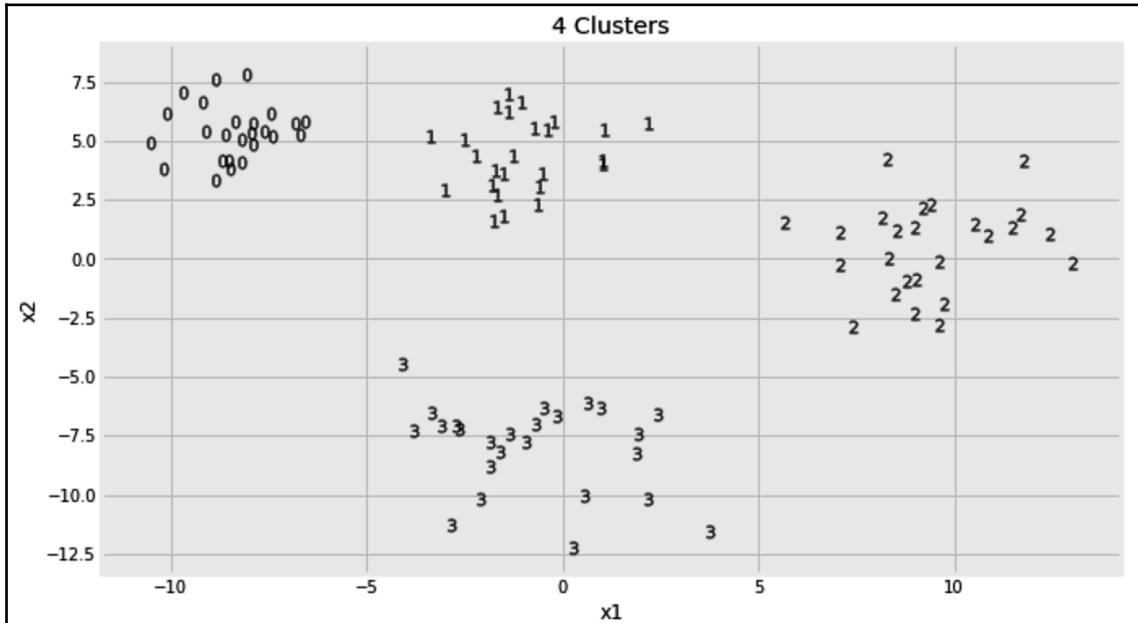
Features	Coeff [w=1]	Coeff [w=2]	Coeff [w=5]
EmbFeature187	0.09	0.17	0.44
EmbFeature188	-1.18	-1.37	-1.54
EmbFeature189	-0.17	-0.26	-0.43
EmbFeature190	0.37	0.45	0.6
EmbFeature191	0.37	0.44	0.51
EmbFeature192	0.6	0.76	0.87
EmbFeature193	-0.92	-0.88	-0.93
EmbFeature194	-0.24	-0.2	-0.15
EmbFeature195	-0.13	-0.18	-0.29
Intercept	-0.15	0.1	0.69

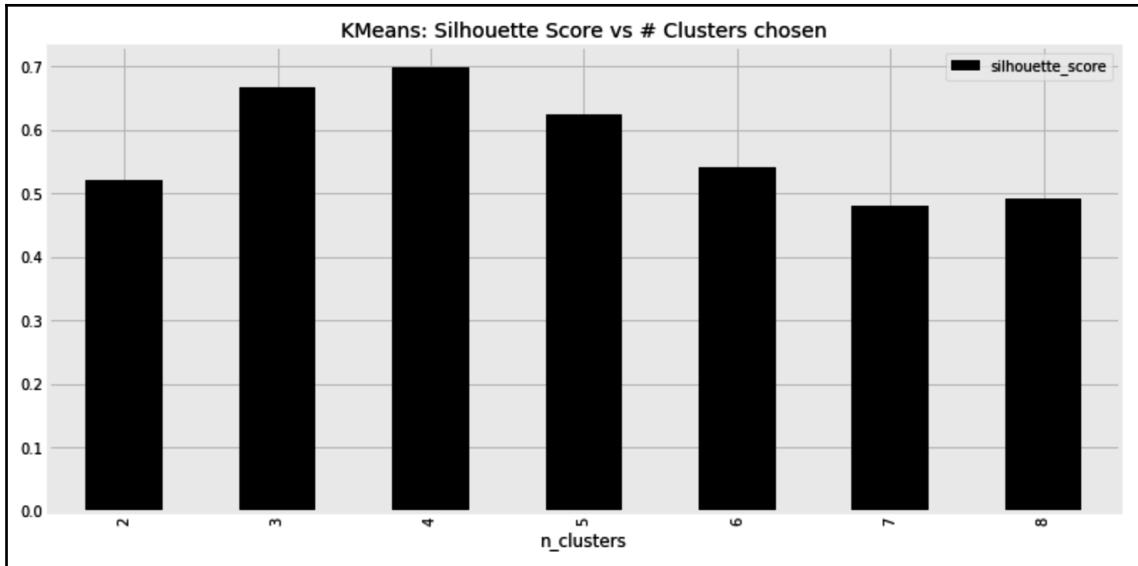
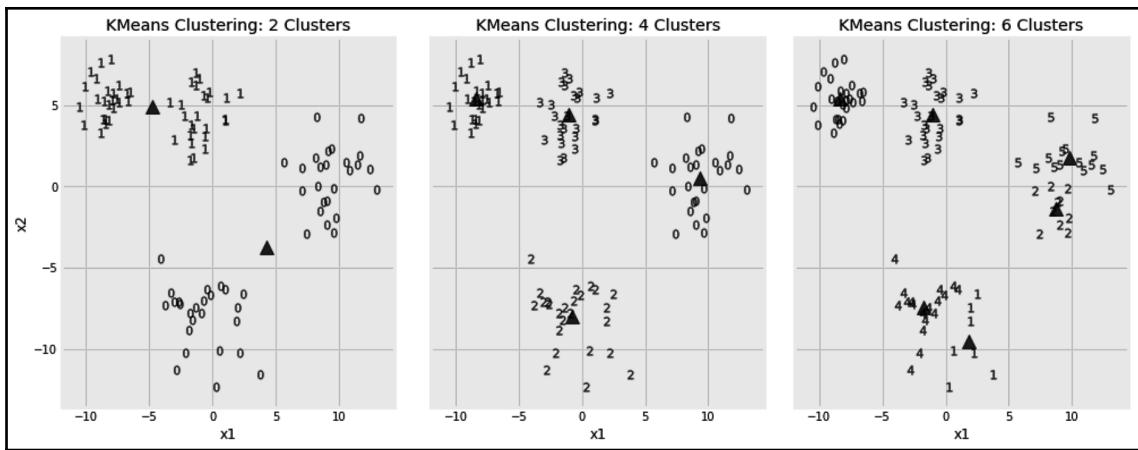


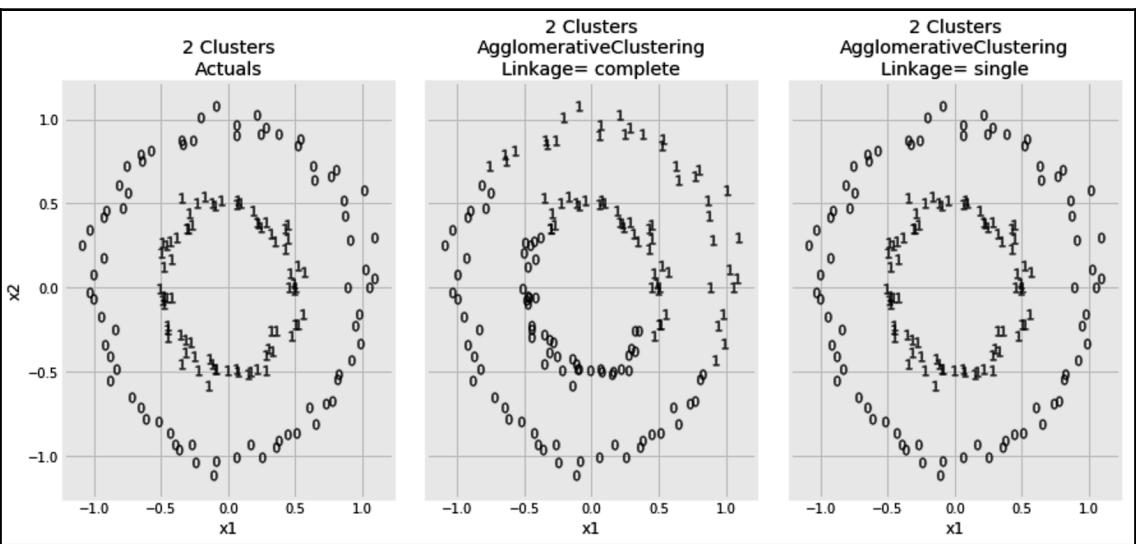
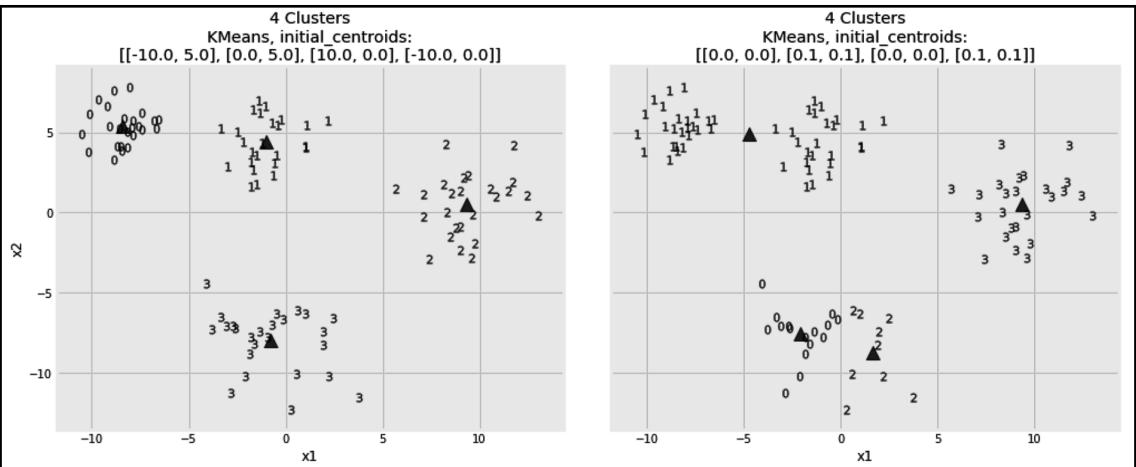


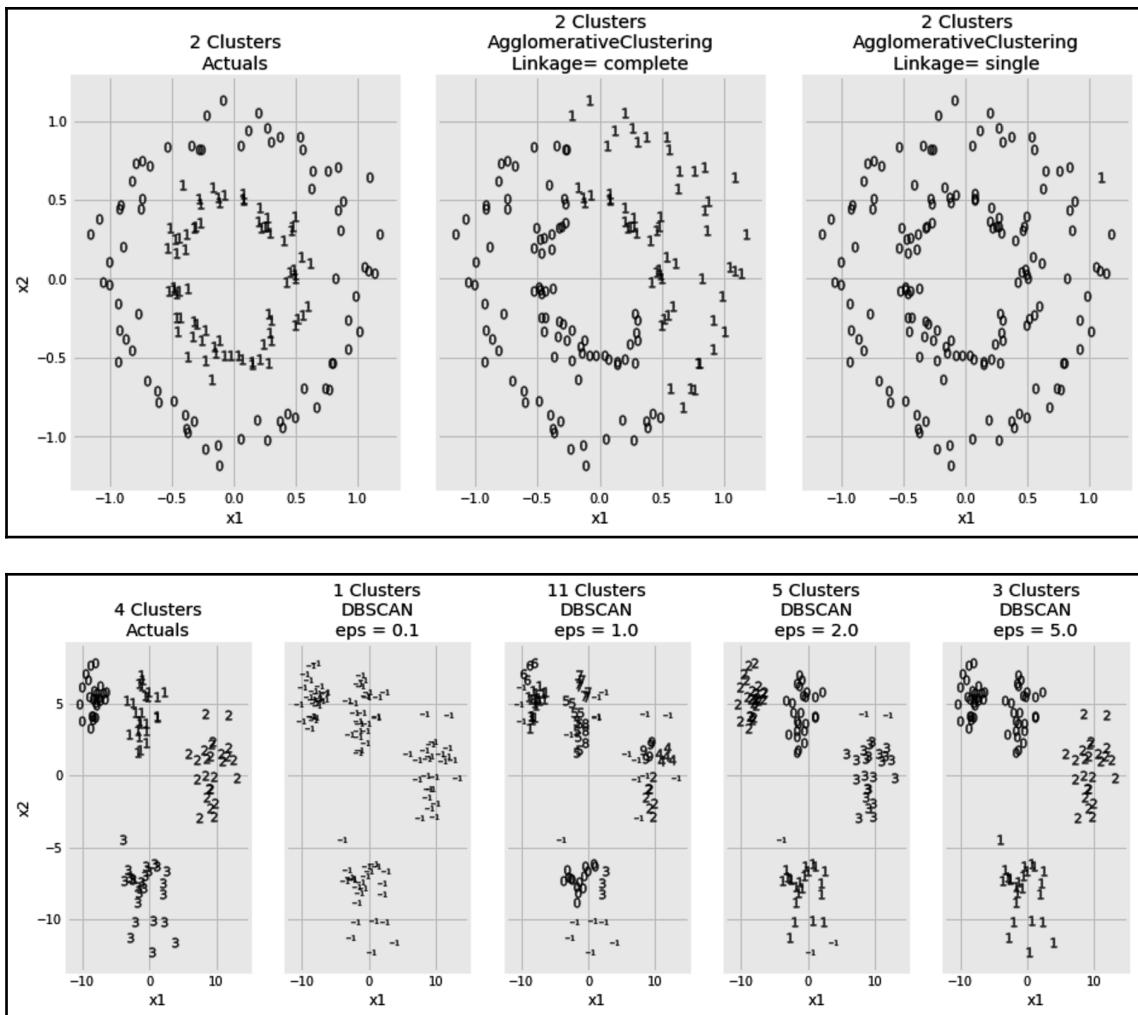


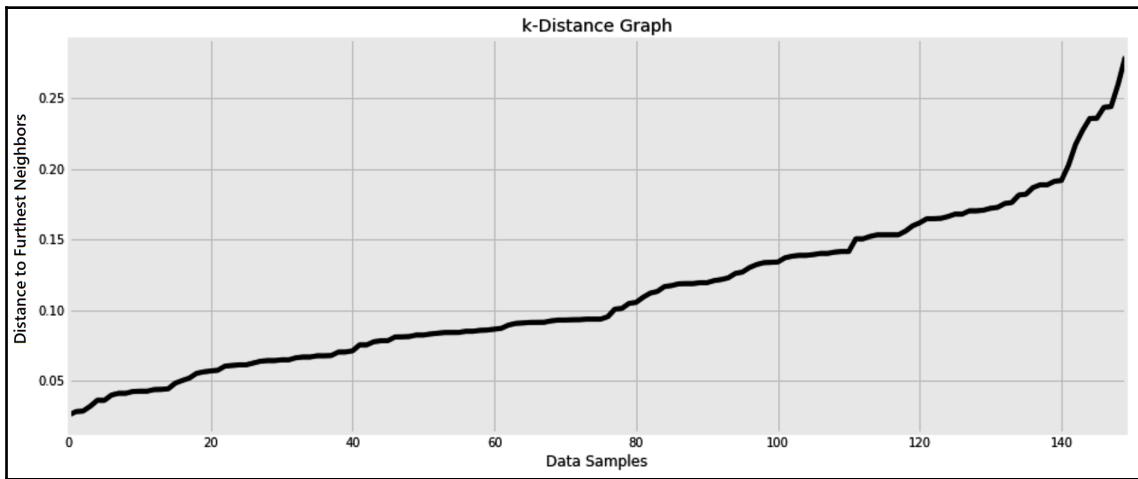
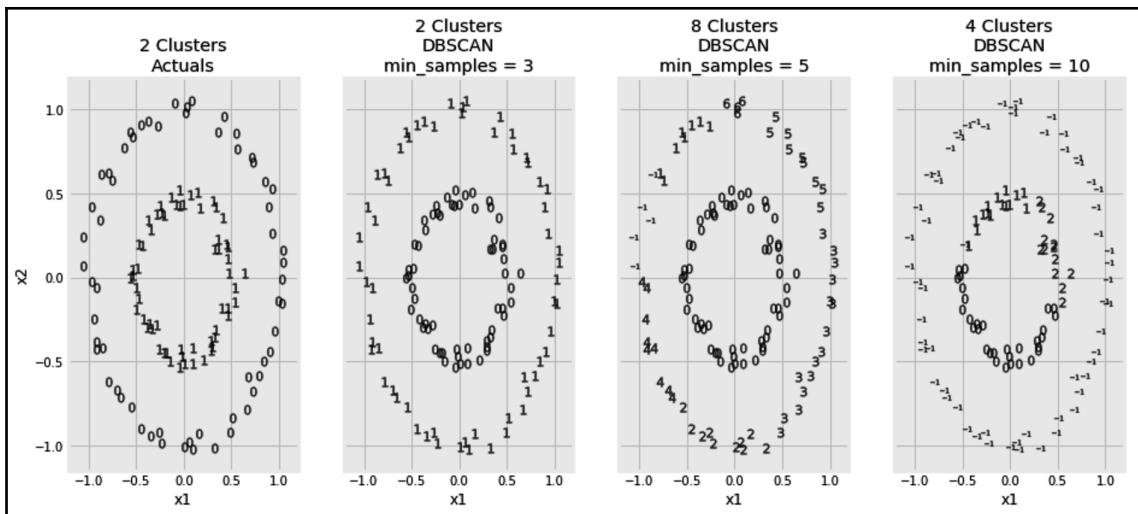
Chapter 11: Clustering - Making Sense of Unlabeled Data







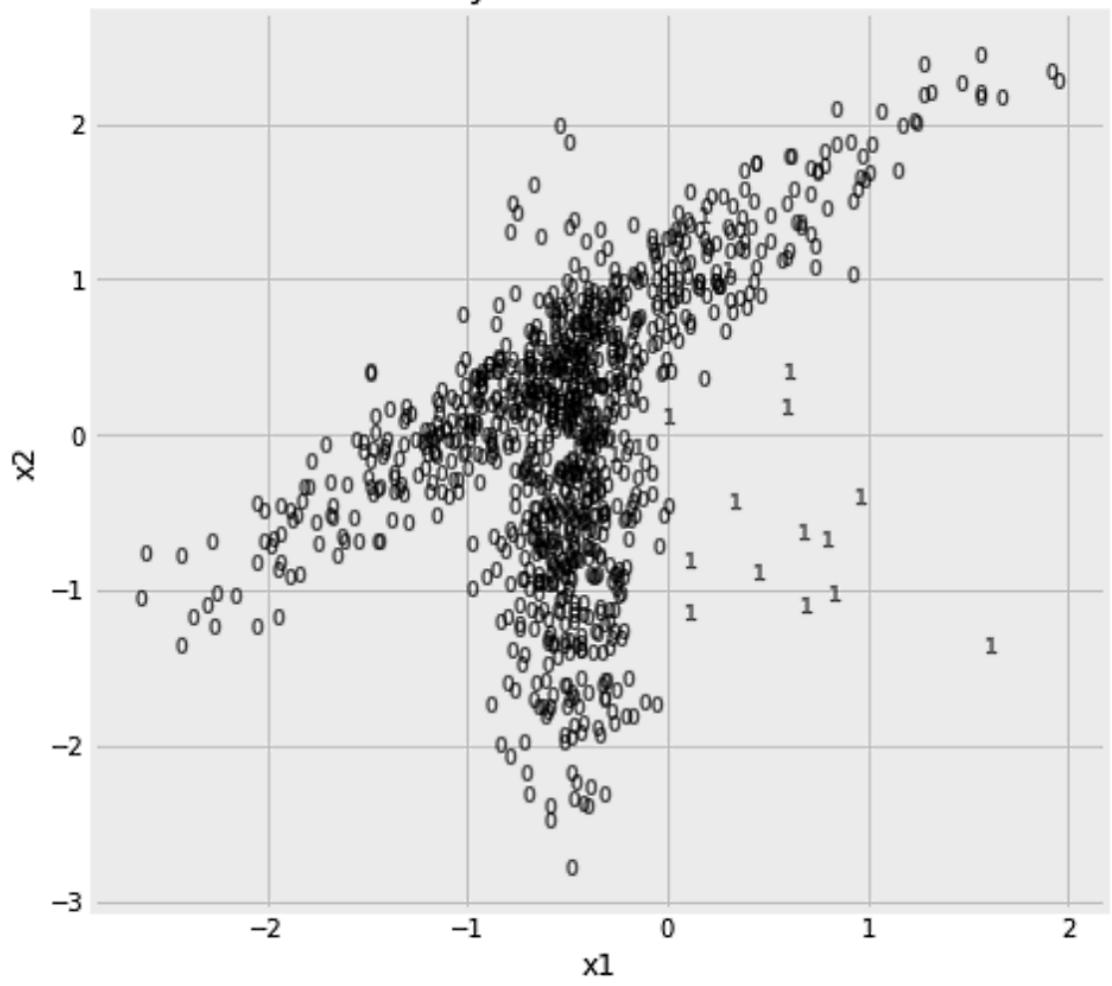




Chapter 12: Anomaly Detection - Finding Outliers in Data

	user_height	is_outlier
0	120	False
1	123	False
2	124	False
3	119	False
4	196	False
5	121	False
6	118	False
7	117	False
8	500	True
9	132	False

Synthetic Dataset



Chapter 13: Recommender System - Getting to Know Their Taste

	user_id	ad_id	user_rating
0	0	8343295	0
1	562934	20017077	1
2	11621116	21348354	0
3	8778348	20366086	0

ad_id	6803526	8343295	9027213	20017077	20366086	20886690	21186478	21348354	21367376	21811752
user_id										
0	0	0	0	0	0	0	0	0	0	0
562934	0	0	0	1	0	0	0	0	0	0
579253	0	0	0	0	0	0	0	0	0	0
2886008	0	0	0	0	0	0	1	0	0	0
5277279	0	0	0	0	0	0	0	0	0	0
7589739	0	0	0	0	0	0	0	0	0	0
8778348	0	0	0	0	0	0	0	0	0	0
11621116	0	0	0	0	0	0	0	0	0	0
11808635	0	0	1	0	0	0	0	0	0	0
12118311	0	0	0	0	0	0	0	0	0	0

Artist	Aly & Fila	Metallica	Paul van Dyk	Rammstein	Slipknot	Tiesto
User						
U1	0	1	0	1	0	0
U2	0	1	0	1	0	0
U3	1	0	1	0	0	1
U4	0	0	1	0	0	1
U5	0	1	0	0	1	0
U6	1	0	0	0	0	1

User	SV1	SV2
U1	0	1.37
U2	0	1.37
U3	1.71	-0
U4	1.21	-0
U5	0	1
U6	1.21	-0

	U1	U2	U3	U5
U1	1	1	0	1
U2	1	1	-0	1
U3	0	-0	1	-0
U5	1	1	-0	1

	U1	U2	U3	U5
U1	1	1	0	1
U2	1	1	-0	1
U3	0	-0	1	-0
U5	1	1	-0	1