

Connor Johnson

CS 1675

Assignment 5

1. (b) Predictions are the columns and actual are the rows

Test Data

	Positive	Negative
Negative	121	40
Positive	22	46

The model error on the test data was 27.07%

Sensitivity = 0.6765, Specificity = 0.7516

Training Data

	Positive	Negative
Negative	263	76
Positive	89	151

The model error on the training data was 30.61%

Sensitivity = 0.555, Specificity = 0.7758

(e)

Test Error

	0.05	$0.2/\sqrt{k}$	$0.05/k$	$2/\sqrt{k}$
4000 epochs	.2969	0.2140	0.2140	0.2664
500 epochs	.2969	0.2140	.2140	0.2751
2000 epochs	.2969	0.2140	0.2140	0.2707

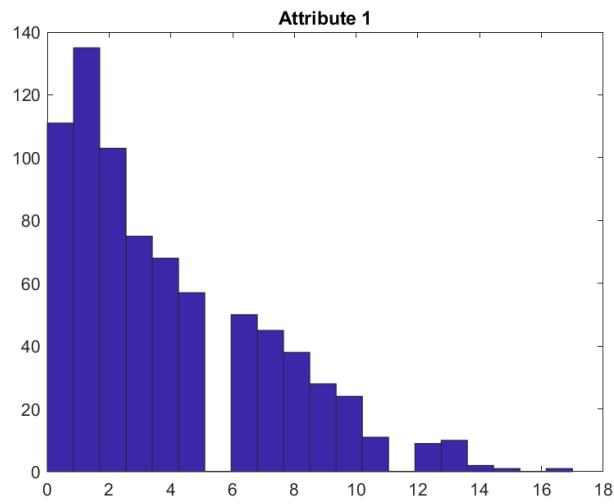
Training error

	0.05	$0.2/\sqrt{k}$	$0.05/k$	$2/\sqrt{k}$
4000 epochs	.2913	0.2301	.2301	0.3024
500 epochs	.2913	0.2301	.2301	0.3340
2000 epochs	.2913	0.2301	.2301	0.3061

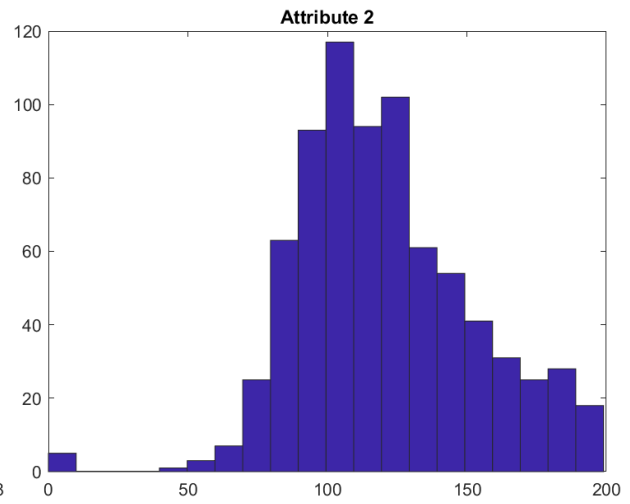
The best performing models are when the number of epochs is higher and when alpha is equal to $0.2/\sqrt{k}$ or $0.05/k$. This is true for both the test and training data.

2.1

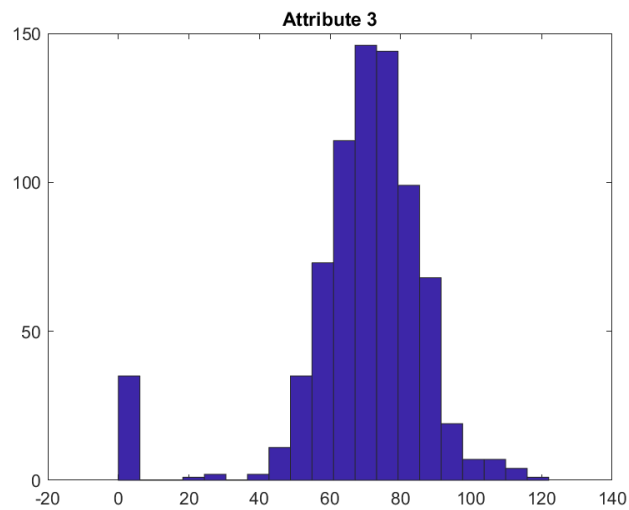
(b)



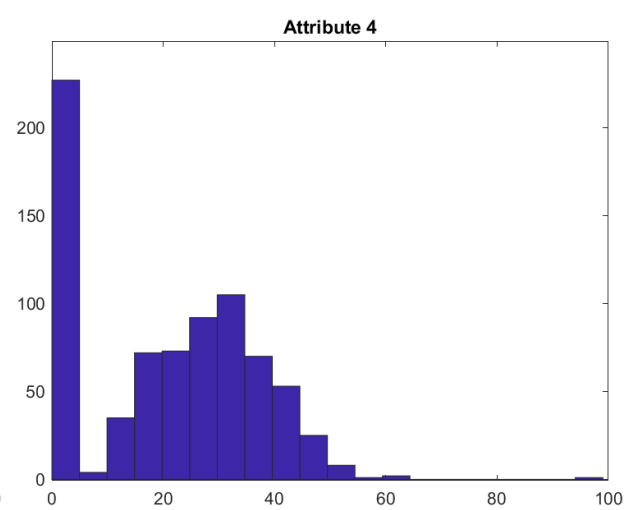
Gamma Distribution



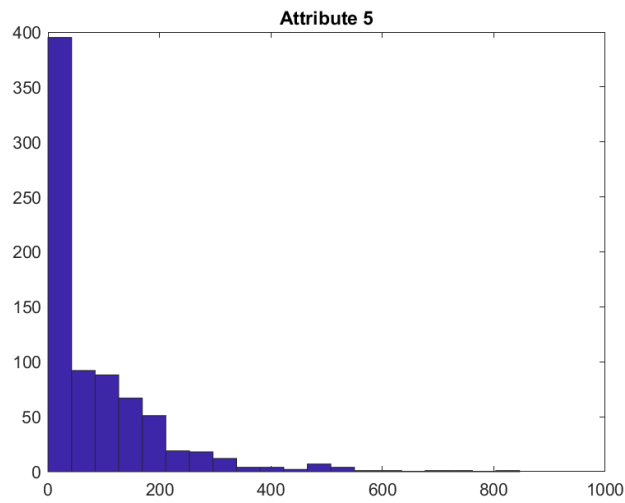
Gamma Distribution



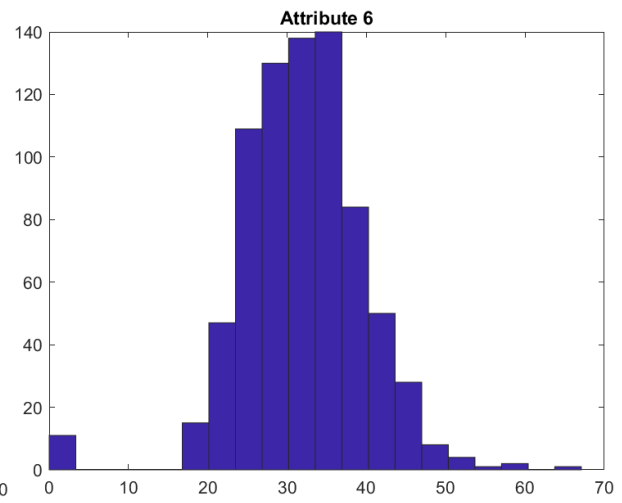
Binomial Distribution



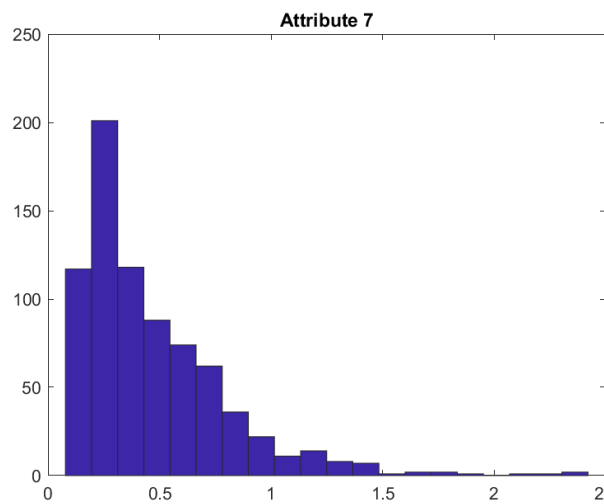
Multinomial distribution



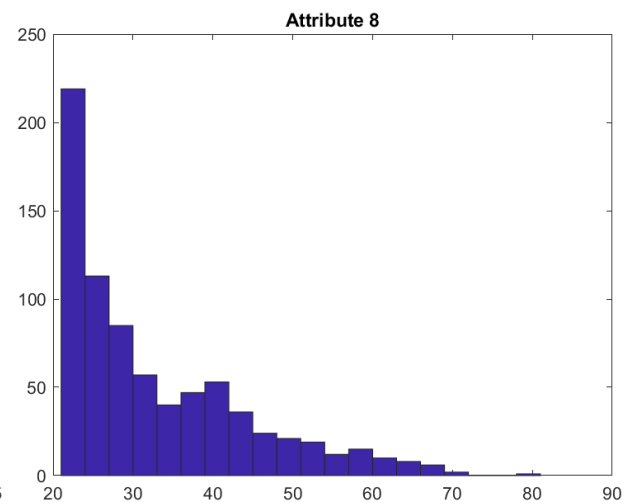
Exponential Dist.



Normal Distribution



Gamma Distribution



Exponential Distribution

2.2b Blank sigma means it's an exponential distribution

	Class 0 mu	Class 0 sigma	Class 1 mu	Class 1 sigma
Attribute 1	3.2419		4.71	
Attribute 2	109.6254	26.2304	141.395	33.6655
Attribute 3	67.5339	18.6683	70.19	21.6213
Attribute 4	19.7316	14.5828	22.935	17.8275
Attribute 5	67.7168		103.72	
Attribute 6	30.3059	7.7258	35.258	7.3286
Attribute 7	0.4164		0.5491	
Attribute 8	31.1032		37.12	

2.3b

Predictions are the columns and actual are the rows

Test Data

	Positive	Negative
Negative	119	42
Positive	18	50

The model error on the test data was 26.2%

Sensitivity = 0.7353, Specificity = 0.7391

Training Data

	Positive	Negative
Negative	235	104
Positive	48	152

The model error on the training data was 28.2%

Sensitivity = 0.76, Specificity = 0.6932

(c)

Overall, the naïve bayes model performs slightly better than the logistic regression. The regression only works when alpha is set to certain values which could just be because of the data we have. Both models have very similar specificities on the test set. The bayes model performs better because of its higher sensitivity. This shows that the bayes model is better at predicting positives when compared to the logistic model.