

8.2 Binary Classifier

Brandon Sams

2/1/2020

Fit a logistic regression model to the binary-classifier-data.csv dataset from the previous assignment.

```
binary_data <- read.csv(url("http://content.bellevue.edu/cst/dsc/520/id/resources/binary-classifier-data.csv"))
log_mod_binary <- glm(label ~ x + y, data = binary_data)
```

a. What is the accuracy of the logistic regression classifier?

```
log_confusion_matrix <- table(binary_data$label, round(predict.glm(log_mod_binary, newdata = binary_data)))
log_correct <- log_confusion_matrix["0", "0"] + log_confusion_matrix["1", "1"]
log_total <- sum(log_confusion_matrix)

log_accuracy <- log_correct / log_total
log_accuracy
```

```
## [1] 0.5834446
```

The logistic regression classifier is about 58.3% accurate.

b. How does the accuracy of the logistic regression classifier compare to the nearest neighbors algorithm?

```
knn_results <- knn(binary_data, binary_data, cl = binary_data$label, k = round(sqrt(nrow(binary_data))))
knn_confusion_matrix <- table(binary_data$label, knn_results)

knn_correct <- knn_confusion_matrix["0", "0"] + knn_confusion_matrix["1", "1"]
knn_total <- sum(knn_confusion_matrix)

knn_accuracy <- knn_correct / knn_total
knn_accuracy
```

```
## [1] 0.9739653
```

The nearest neighbors algorithm is much more accurate, at 97.3% accuracy.

c. Why is the accuracy of the logistic regression classifier different from that of the nearest neighbors?

The accuracy of the logistic regression classifier is different from that of the nearest neighbors accuracy because they are completely different methods. They work in completely different ways. While logistic regression is probability based, nearest neighbors is deterministic. Nearest neighbors makes a guess, based on the surrounding context, but does not tell how confident that guess is. So depending on the data set, one technique may produce more accurate results than the other.