# Classification: Predicting Heart Disease

Sujay Talanki

8/1/2022

---

This project uses several different classification methods to predict whether or not a patient has heart disease.

Loading the dataset:

```
heart <- read.csv("C:/Users/talan/Downloads/heart.csv")
```

I will split the data into a training set and a test set:

```
set.seed(1)
n = nrow(heart)

#Creating the training set
train = sample(1:n, size = n/2)
train.set = heart[train, ]

#Creating the test set
test = (-train)
test.set = heart[test, ]
```

- **(a)**: Using **Logistic Regression**:

  ```
  #Creating the logistic regression model
  logistic.fit = glm(target ~ ., data = train.set, family = "binomial")

  #Using the training model on the test set
  probability = predict(logistic.fit, test.set, type = "response");
  prediction = rep(0, 1)
  prediction[probability > 0.5] = 1
  test_pred <- ifelse(predict(logistic.fit, newdata = test.set,
  type = "response") > 0.5, 1, 0)
  test.accuracy = (mean(test_pred == test.set$target))*100
  ```

  The test accuracy is 84.87%.

- **(b)**: Using **Linear Discriminant Analysis (LDA)**:

  ```
  lda.fit = lda(target ~ ., data = train.set)
  pred <- predict(lda.fit, test.set)
  test.accuracy = (mean(pred$class == test.set$target))*100
  ```

  The test accuracy of LDA is ~ 84.21%.

- **(c)**: Using **Quadratic Discriminant Analysis (QDA)**:

```
qda.fit = qda(target ~ ., data = train.set)
pred <- predict(qda.fit, test.set)
test.accuracy = (mean(pred$class == test.set$target))*100
```

Thus, the test accuracy of QDA is ~ 79.61%.

- **(d)**: Using a **Support Vector Machine (SVM)**. This will be a linear SVM and I will tune the model in order to find the best cost.

```
#Finding the optimal cost
tune.fit = tune(svm, target ~ ., data = train.set,
kernel = "linear", ranges = list(cost = seq(0.01, 10, 0.5)))
tune.fit
```

```
##
## Parameter tuning of 'svm':
##
## - sampling method: 10-fold cross validation
##
## - best parameters:
##  cost
##  0.01
##
## - best performance: 0.1362613
```

From the output, the test accuracy is 86.73%.

- **(e)**: Using a **Support Vector Machine (SVM)**. This will be a radial SVM and I will tune the model in order to find the best cost.

```
#Finding the optimal cost
tune.fit = tune(svm, target ~ ., data = train.set,
kernel = "radial", ranges = list(cost = seq(0.01, 10, 0.5)))
tune.fit
```

```
##
## Parameter tuning of 'svm':
##
## - sampling method: 10-fold cross validation
##
## - best parameters:
##  cost
##  0.51
##
## - best performance: 0.143667
```

From the output, the test accuracy is 85.63%.

**Conclusion**: Outside of the QDA, most of the classification methods result in a test accuracy of around 84% - 86%.