ROC for logistic regression

There are numerous R functions that can be used to plot an ROC and find the associated AUC.

Here we will look at one method that works well for logistic regression. This method is essentially an extension of the ggplot package. The `geom_roc` function that we are going to use can be found in the `plotROC` package.

First we load the necessary packages.

```
library(tidyverse) # to manipulate the generated data + ggplot
library(plotROC) # has geom_roc
```

And to keep things simple we will generate our own data that has one explanatory variable `Obs` and the class is either `0` or `1`.

```
# simulate data
df <- data.frame(`Yes`=rnorm(1000,-1,1),`No`=rnorm(1000,1,1)) %>%
  gather(key='Class',value='Obs') %>%
  mutate(Class = as.numeric(case_when(Class == 'Yes' ~ '1',Class == 'No' ~ '0')))
```

Next we fit the logistic regression model to the data.

```
# Fit glm
fit <- glm(Class~Obs,data=df,family='binomial')
```

Recall that we can use the fitted model to get the probability that a given observation belongs to class `1` using the `predict` function.

```
# predict(fit,type='response')
```

In order to use the `geom_roc` function we have to create a dataframe that contains the estimated probabilities and the actual class labels.

```
# Create a dataframe that has the true labels,
#  and the probabilities that the glm fit gives for belonging to class 1.
df_class <- data.frame(a=df$Class,b=predict(fit,type='response'))
```

Finally we use `ggplot` with the `geom_roc` option to get the ROC curve. The asthetics of `geom_roc` are the true class labels, `d`, and the associated probabilities of belonging to class `1`, `m`.

```
p <- ggplot(df_class,aes(d=a,m=b)) +
       geom_roc() +
       ylab('sensitivity') +
       xlab('1-specificity')

p
```

Finally, we can get the are under the receiver operating curveby using the `calc_auc` function on the previous `ggplot` object.

```
calc_auc(p)$AUC
```