STATS 7022 - Data Science PG Assignment 1

Dang Thinh Nguyen

2024-06-13

Question 3: ROC function

```r
get_ROC <- function(obs, pred) {
  # Create a dataframe with 2 variables: 'obs' and 'pred'
  data <- tibble(
    obs = factor(obs),
    pred = pred
  )

  # Remove duplicate rows
  data <- data[!duplicated(data),]

  # Sort predicted probabilities
  data <- data %>% arrange(pred)

  # Initialize vectors for thresholds, specificity, and sensitivity
  thresholds <- c(-Inf, data$pred, Inf)
  sens <- numeric(length(thresholds))
  spec <- numeric(length(thresholds))

  # Calculate sensitivity and specificity at each threshold
  for (i in seq_along(thresholds)) {
    threshold <- thresholds[i]
    tp <- sum(data$pred >= threshold & data$obs == "A")
    fp <- sum(data$pred >= threshold & data$obs == "B")
    fn <- sum(data$pred < threshold & data$obs == "A")
    tn <- sum(data$pred < threshold & data$obs == "B")

    sens[i] <- tp / (tp + fn)
    spec[i] <- tn / (tn + fp)
  }

  # Create a tibble with the results
  roc_tibble <- tibble(
    threshold = thresholds,
    specificity = spec,
    sensitivity = sens
  )
```

```
  return(roc_tibble)
}
```

Test function

```
df <- tibble(
  obs = rep(factor(c("A", "B")), each = 2),
  A = rep(c(0.8, 0.2), each = 2)
)

get_ROC(df$obs, df$A)
```

```
## # A tibble: 4 x 3
##   threshold specificity sensitivity
##       <dbl>       <dbl>       <dbl>
## 1    -Inf             0           1
## 2       0.2           0           1
## 3       0.8           1           1
## 4     Inf             1           0
```