Logistic

2025-12-22

```r
data(iris)
y =(iris$Species =="versicolor")
x = iris[,-5]
n=nrow(iris)
set.seed(123)
ind_train = sample(1:n, size = 0.8*n)

y_train = y[ind_train]
y_test = y[-ind_train]
x_train = x[ind_train,]
x_test = x[-ind_train,]
#model fitting
fit = glm(y_train~., family = "binomial", data = x_train)
coef = fit$coef
#prediction
pred_prob = predict(fit, x_test, "response")
pred_class = (pred_prob > 1/2)
mean(pred_class == y_test)
```

```
## [1] 0.6333333
```

Take home exercise: 1. Use logistic regression to do classification task for the GALA dataset.

2. Since the iris dataset has three classes, we would like to do multi-class classification. Use the R-package nnet, and the function multinom to achieve multi-class classification. For more documentation, you can refer to https://cran.r-project.org/web/packages/nnet/nnet.pdf.

3. Show me theoretically how prediction was done. As a reference, please refer to ESL page 119-122.