

Программирование в среде R

Занятие 21

Все комментарии направляются преподавателю по электронной почте.
В теме или теле письма обязательно указать номер группы.

1. Модель логистической регрессии в R

- 1.1. `#загружаем данные`
- 1.2. `train <- nbsp="" span="">read.csv('train.csv')`
- 1.3. `#создаем обучающую и тестовую выборки из наших данных`
- 1.4. `install.packages('caTools')`
- 1.5. `library(caTools)`
- 1.6. `set.seed(88)`
- 1.7. `split <- nbsp="" span="">sample.split(train$Recommended, SplitRatio = 0.75)`
- 1.8. `#создаем обучающую и тестовую выборки`
- 1.9. `dresstrain <- nbsp="" span="">subset(train, split == TRUE)`
- 1.10. `drestest <- nbsp="" span="">subset(train, split == FALSE)`
- 1.11. `#модель логистической регрессии`
- 1.12. `model <- nbsp="" span="">glm (Recommended ~ .-ID, data = dresstrain, family = binomial)`
- 1.13. `summary(model)`
- 1.14. `predict <- nbsp="" span="">predict(model, type = 'response')`
- 1.15. `#матрица ошибок`
- 1.16. `table(dresstrain$Recommended, predict > 0.5)`
- 1.17. `#ROC-кривая`
- 1.18. `library(ROCR)`
- 1.19. `ROCRpred <- nbsp="" span="">prediction(predict, dresstrain$Recommended)`
- 1.20. `ROCRperf <- nbsp="" span="">performance(ROCRpred, 'tpr','fpr')`
- 1.21. `plot(ROCRperf, colorize = TRUE, text.adj = c(-0.2,1.7))`
- 1.22. `#строим график glm`
- 1.23. `library(ggplot2)`
- 1.24. `ggplot(dresstrain, aes(x=Rating, y=Recommended)) + geom_point() +`
- 1.25. `stat_smooth(method="glm", family="binomial", se=FALSE)`

2. ROC-кривая с одной выборкой

- 2.1. `Инсталлируйте пакет ROCR`
- 2.2. пример 1**
- 2.3. `data(ROCR.simple)`
- 2.4. `pred <- prediction(ROCR.simple$predictions, ROCR.simple$labels)`
- 2.5. `perf <- performance(pred,"tpr","fpr")`
- 2.6. `plot(perf)`
- 2.7. пример 2**
- 2.8. `perf <- performance(pred,"acc")`
- 2.9. `plot(perf)`
- 2.10. пример 3**
- 2.11. `perf <- performance(pred,"ecost")`
- 2.12. `plot(perf)`
- 2.13. пример 4**
- 2.14. `perf <- performance(pred,"auc")`
- 2.15. пример 5**

- 2.16. `perf <- performance(pred,"tpr","fpr")`
- 2.17. `par(bg="lightblue", mai=c(1.2,1.5,1,1))`
- 2.18. `plot(perf, main="ROCR fingerprinting toolkit", colorize=TRUE, xlab="fpr", ylab="tpr",
+box.lty=7, box.lwd=5, xaxis.col="blue", xaxis.col.axis="blue", yaxis.col="blue",
+yaxis.at=c(0,0.5,0.8,0.85,0.9,1), yaxis.las=1, xaxis.lwd=2, yaxis.lwd=2, yaxis.col.axis="blue",
+cex.lab=2, cex.main=2)`
- 3. ROCR Пример с несколькими выборками
 - 3.1. пример 1**
 - 3.2. `data(ROCR.xval)`
 - 3.3. `pred <- prediction(ROCR.xval$predictions, ROCR.xval$labels)`
 - 3.4. `perf <- performance(pred,"tpr","fpr")`
 - 3.5. `plot(perf,col="black",lty=3)`
 - 3.6. пример 2**
 - 3.7. `plot(perf,lwd=3,avg="vertical",spread.estimate="boxplot",add=TRUE)`
- 4. Сохранить историю команд в файл. Результат выслать В.В.Шевцову vvshevtsov@fa.ru
- 5. Завершить работу