setwd("C:\\Users\\Administrator\\Desktop\\NBA\\NBA")

#载入包

library(dplyr)

library(readxl)

library(ggplot2)

library(e1071)

library(gridExtra)

library(DMwR)

library(rpart)

Player\_sta <- read\_xlsx("2011-12.xlsx",sheet="Player\_sta",col\_names=T,na="NA")

Team\_Name <- read\_xlsx("2011-12.xlsx",sheet="Team\_Name",col\_names=T,na="NA")

Team\_Result <- read\_xlsx("2011-12.xlsx",sheet="Team\_Result",col\_names=T,na="NA")

Team\_sta <- read\_xlsx("2011-12.xlsx",sheet="Team\_sta",col\_names=T,na="NA")

Opponent\_sta <- read\_xlsx("2011-12.xlsx",sheet="Opponent\_sta",col\_names=T,na="NA")

Team\_Result <- right\_join(Team\_Name,Team\_Result,by=c("Full Name"="Full Name"))

Team\_sta <- right\_join(Team\_Name,Team\_sta,by=c("Full Name"="Full Name"))

Opponent\_sta <- right\_join(Team\_Name,Opponent\_sta,by=c("Full Name"="Full Name"))

Player\_sta <- full\_join(Team\_sta,Player\_sta,by=c("Short Name"="Tm"),keep=F)

Player\_sta <- full\_join(Team\_Result,Player\_sta,by=c("Short Name"="Short Name"),keep=F)

Player\_sta <- full\_join(Opponent\_sta,Player\_sta,by=c("Short Name"="Short Name"),keep=F)

#.x是player的数据，.y是所在team的数据，什么都不加的是对手队伍的数据

#无用的列Player\_sta[,c("Full Name.x","Rk.x.x","Full Name.y","Rk.x")]

#构造数据变量

VOP = Player\_sta$PTS.y / (Player\_sta$FGA.y - Player\_sta$ORB.y + Player\_sta$TOV.y + 0.44 \* Player\_sta$FTA.y)

DRB\_percentage = (Player\_sta$TRB.y - Player\_sta$ORB.y) / Player\_sta$TRB.y

factor = (2/3)-(0.5\*(Player\_sta$AST.y/Player\_sta$FG.y))/(2\*(Player\_sta$FG.y/Player\_sta$FT.y))

#未调整的PER

uPER = (1/Player\_sta$MP.y)\*(Player\_sta$`3P.y`+(2/3)\*Player\_sta$AST.y+

(2-factor\*(Player\_sta$AST.x/Player\_sta$FG.x))\*Player\_sta$FG.y+

(0.5\*Player\_sta$FT.y\*(1+(1-(Player\_sta$AST.x/Player\_sta$FG.x))+(2/3)\*

(Player\_sta$AST.x/Player\_sta$FG.x)))-VOP\*Player\_sta$TOV.y-VOP\*DRB\_percentage\*

(Player\_sta$FGA.y-Player\_sta$FG.y)-VOP\*0.44\*(0.44+(0.56\*DRB\_percentage))\*

(Player\_sta$FTA.y-Player\_sta$FT.y)+VOP\*(1-DRB\_percentage)\*(Player\_sta$TRB.y-

Player\_sta$ORB.y)+VOP\*DRB\_percentage\*Player\_sta$ORB.y+VOP\*Player\_sta$STL.y+

VOP\*(1-DRB\_percentage)\*(Player\_sta$TRB.y-Player\_sta$ORB.y)+VOP\*DRB\_percentage\*

Player\_sta$ORB.y+VOP\*Player\_sta$STL.y+VOP\*DRB\_percentage\*Player\_sta$BLK.y-

(Player\_sta$FT.y-Player\_sta$PF.y)-0.44\*VOP\*(Player\_sta$FTA.y/Player\_sta$PF.y))

Play\_time<-aggregate(Player\_sta$MP.y,by=list(Player\_sta$`Short Name`,Player\_sta$Player),sort,decreasing=T)

colnames(Play\_time) <- c("Tm","Player","Minutes\_Play")

#构建teamPER

Player\_PER <- as.data.frame(cbind(Player\_sta$Player,uPER))

colnames(Player\_PER) <- c("Player","uPER")

Player\_PER <- right\_join(Play\_time,Player\_PER,by=c("Player"="Player"))

Player\_PER$uPER <- as.numeric(as.character(Player\_PER$uPER))

Player\_PER$multi\_min <- Player\_PER$Minutes\_Play\*Player\_PER$uPER

Team\_PER <- aggregate(Player\_PER$multi\_min,by=list(Player\_PER$Tm),sum)

Team\_PER <- Team\_PER[-which(is.na(Team\_PER$x)),]

colnames(Team\_PER) <- c("Tm","Team\_PER")

Team\_Conclusion <- right\_join(Team\_PER,Team\_Result,by=c("Tm"="Short Name"))

#summary(Team\_Conclusion)

Team\_Conclusion$Conf <- as.factor(Team\_Conclusion$Conf)

Team\_Conclusion$Div <- as.factor(Team\_Conclusion$Div)

colnames(Team\_Conclusion)<-c("Team","Team\_PER","Full Name","Rk","Conference","Div","W", "L","Wining\_Rate","MoV","ORtg","DRtg","NRtg","A\_MoV",

"A\_ORtg","A\_DRtg","A\_NRtg")

#model2 svm

svm1 <- svm(Wining\_Rate~Team\_PER+Rk+A\_MoV+A\_ORtg+A\_DRtg+A\_NRtg,data=Team\_Conclusion,kernal="sigmoid")

pre\_result <- round(predict(svm1,Team\_Conclusion[,c("Team\_PER","Rk","A\_MoV","A\_ORtg","A\_DRtg","A\_NRtg")]),3)

Compare <- as.data.frame(cbind(Team\_Conclusion$Team,as.character(Team\_Conclusion$Conference),pre\_result,Team\_Conclusion$Wining\_Rate,Team\_Conclusion$Rk))

colnames(Compare)<-c("Team","Conf","Predict","Real","Rank")

Compare$Rank <- as.numeric(as.character(Compare$Rank))

Compare$Predict <- as.numeric(as.character(Compare$Predict))

Compare$Real <- as.numeric(as.character(Compare$Real))

mse <- sqrt(mean((Compare$Predict-Compare$Real)^2))

p1 <- ggplot(Compare,aes(x=Rank,y=Real))+

geom\_point(aes(color=Conf))+geom\_text(aes(label=Team),size=2.5,nudge\_y=-0.01)+

labs(title="Real Wining Rate")+

xlab("Rank")+ylab("Wining Rate")+

theme(plot.title = element\_text(size = 8))+

theme\_bw()

p2 <- ggplot(Compare,aes(x=Rank,y=Predict))+

geom\_point(aes(color=Conf))+geom\_text(aes(label=Team),size=2.5,nudge\_y=-0.01)+

labs(title="Predict Wining Rate")+

xlab("Rank")+ylab("Wining Rate")+

theme(plot.title = element\_text(size = 8))+

theme\_bw()

grid.arrange(p1, p2,ncol=2)

#model3 glm

lm1 <- lm(Wining\_Rate~Team\_PER+Rk+A\_MoV+A\_ORtg+A\_DRtg+A\_NRtg,data=Team\_Conclusion)

summary(lm1)

pre\_result <- round(predict(lm1,Team\_Conclusion[,c("Team\_PER","Rk","A\_MoV","A\_ORtg","A\_DRtg","A\_NRtg")]),3)

Compare <- as.data.frame(cbind(Team\_Conclusion$Team,as.character(Team\_Conclusion$Conference),pre\_result,Team\_Conclusion$Wining\_Rate,Team\_Conclusion$Rk))

colnames(Compare)<-c("Team","Conf","Predict","Real","Rank")

Compare$Rank <- as.numeric(as.character(Compare$Rank))

Compare$Predict <- as.numeric(as.character(Compare$Predict))

Compare$Real <- as.numeric(as.character(Compare$Real))

mse <- sqrt(mean((Compare$Predict-Compare$Real)^2,na.rm = T))

p1 <- ggplot(Compare,aes(x=Rank,y=Real))+

geom\_point(aes(color=Conf))+geom\_text(aes(label=Team),size=2.5,nudge\_y=-0.01)+

labs(title="Real Wining Rate")+

xlab("Rank")+ylab("Wining Rate")+

theme(plot.title = element\_text(size = 8))+

theme\_bw()

p2 <- ggplot(Compare,aes(x=Rank,y=Predict))+

geom\_point(aes(color=Conf))+geom\_text(aes(label=Team),size=2.5,nudge\_y=-0.01)+

labs(title="Predict Wining Rate")+

xlab("Rank")+ylab("Wining Rate")+

theme(plot.title = element\_text(size = 8))+

theme\_bw()

grid.arrange(p1, p2,ncol=2)