Попов Максим
Machine learning, task 3

Для данного задания было необходимо предсказать результаты выборов используя метод опорных векторов.
Скрипт состоит из двух частей - подбор параметров и непосредственно предсказание.
Для подбора парметров:

```
fitStats <- function(y,y.pred)
{
  accuracy <- sum(y==y.pred)/length(y);
  precision <- sum(y==1&y.pred==1)/sum(y==1);
  recall <- sum(y==1&y.pred==1)/sum(y.pred==1);
  if (precision + recall == 0)
  {
    f1.score <- 0;
  }else{
    f1.score <- 2*precision*recall/(precision+recall);
  }
  stat <- c(accuracy,precision,recall,f1.score);
  names(stat) <- c("accuracy","precision","recall","f1.score");
  stat;
}

raw <- read.csv("elections_usa96_train.csv", header = TRUE);

y <- as.numeric(raw$vote);
X <- data.frame();
X <- cbind(raw$popul, raw$TVnews, as.numeric(raw$ClinLR), as.numeric(raw$DoleLR), raw$age, as.numeric(raw
$educ), as.numeric(raw$income));

m <- nrow(X);
m.train <- round(0.8*m);
m.cv <- m - m.train;
train.obs <- sample(1:m,size=m.train,replace=FALSE);
cv.obs <- (1:m)[-train.obs];
X.train <- X[train.obs,];
X.cv <- X[cv.obs,];
y.train <- y[train.obs];
y.cv <- y[cv.obs];

par <- c(0.01,0.05,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0,1,5,10,50,100,500,1000);
par <- expand.grid(par,par);
dimnames(par)[[2]] <- c("C","sigma");

res <- NULL
library(kernlab)
for (i in 1:nrow(par))
{
  model <- ksvm(X.train, y.train, type="C-svc", C = par$C[i], kern = "rbfdot", kpar = list(sigma=par$sigma[i]));
  y.pred <- predict(model, newdata = X.cv, type = "response");

  res <- rbind(res, c(par$C[i],par$sigma[i],fitStats(y.cv,y.pred)) );
}
dimnames(res)[[2]][1:2] <- c("C","sigma");

j <- which.max(res[,"f1.score"]);
res[j,]
```

В результате исполнения этого скрипта получаем следующее:

C	sigma	accuracy	precision	recall	f1.score
1.0000000	0.4000000	0.7605634	0.8800000	0.7252747	0.7951807

Далее, используя эти параметры мы можем произвести прогнозирование:

```
raw_tst <- read.csv("elections_usa96_test.csv", header = TRUE);

X_TST <- data.frame();
X_TST <- cbind(raw_tst$popul, raw_tst$TVnews, as.numeric(raw_tst$ClinLR), as.numeric(raw_tst$DoleLR), raw_tst
$age, as.numeric(raw_tst$educ), as.numeric(raw_tst$income));

model <- ksvm(X.train, y.train, type="C-svc", C = 1.0, kern = "rbfdot", kpar = list(sigma=0.4));
y_tst <- predict(model, newdata = X_TST, type = "response");

write.csv(y_tst,"M_L_3.csv")
```

В конечном файле с результатом следующее обозначение (согласно выводу as.numeric() со входным аргументом в виде исходных данных):

2 - Dole
1 - Klinton