ST340 Lab 7: Support vector machines

2020–21

Load package e1071:

```
if (!require("e1071")){
  install.packages("e1071")
  library("e1071")
}
```

(a) Run the following code, and then construct and plot an SVM with a quadratic kernel. What is its test set accuracy? (The code to answer this is on the lecture slides).

```
set.seed(1)
th=runif(100,-pi,pi)
y=sign(th)
x1=sin(th)-y/3+rnorm(100,sd=0.1)
x2=cos(th)+y/2+rnorm(100,sd=0.1)
plot(x2,x1,asp=1,col=(y+3)/2)
d <- data.frame(x1=x1,x2=x2,y=factor(y))
```

(b) Use leave-one-out-cross-validation to compare polynomial kernels. Write some code to perform a grid search over possible choices of parameters. For simplicity, fix $c = 1$ (`coef0 = 1`) and $C = 1$ (`cost=1`) and search over the remaining parameters γ and p and to select the model which is optimal in terms of cross-validation error. A suggested grid is:

```
p.range <- 1:10
gamma.range <- c(0.001,0.01,0.1,0.5,1,2,5,10)
```

(Hint: You do not need to write a function to compute the cross-validation error: look at the `cross` flag in `?svm`. You can extract the accuracies of the cross-validation sets from an svm object `s` using `s$accuracies` and `s$tot.accuracy`.)

(c) Repeat part (b) to choose a value of γ for the radial basis function.

(d) Compare the performance of your chosen polynomial kernel and RBF kernel by simulating an independent test dataset and evaluating their accuracies:

```
s.poly <- svm(y~.,d,type="C-classification",kernel="polynomial",degree=chosen.p,
              gamma=chosen.poly.gamma,coef0=1,cost=1)
s.rbf <- svm(y~.,d,type="C-classification",kernel="radial",gamma=chosen.rbf.gamma,cost=1)
set.seed(2)
th=runif(100,-pi,pi)
y.test=sign(th)
x1.test=sin(th)-y.test/3+rnorm(100,sd=0.1)
x2.test=cos(th)+y.test/2+rnorm(100,sd=0.1)
plot(x2.test,x1.test,asp=1,col=(y.test+3)/2)
d.test <- data.frame(x1=x1.test,x2=x2.test,y=factor(y.test))

mean(predict(s.poly,d.test)==y.test)
mean(predict(s.rbf,d.test)==y.test)
plot(s.poly,d.test,grid=200,asp=1)
```

```
plot(s.rbf,d.test,grid=200,asp=1)
```