STAT 671

Statistical Learning I

Fall 2019

Homework 1

Due October 14^{th} at the beginning of class

«echo=T,fig.height=5,fig.width=5,fig.align='center'»=
@

1 A simple classifier

1. Finish the derivation of the simple classifier provided in class.

   We know

   $g(x) =< C_+ - C_-, X - C >=< C_+, X > - < C_-, X > - < C_+, C > + < C_-, C >$

   For

   $< C_+, X >=< \frac{1}{n_+} \sum_{l \in I_+}^{n} x_i, x >$

   $< C_-, X >=< \frac{1}{n_-} \sum_{l \in I_-}^{n} x_i, x >$

   $< C_+, C >=< C_+, \frac{1}{2}C_+ > + < C_+, \frac{1}{2}C_- >= \frac{1}{2n_+^2} \sum_{(i,j) \in I_+} < x_i, x_j > + \frac{1}{2} < C_+, C_- >$

   $< C_-, C >=< C_-, \frac{1}{2}C_+ > + < C_-, \frac{1}{2}C_- >= \frac{1}{2} < C_+, C_- > + \frac{1}{2n_-^2} \sum_{(i,j) \in I_-} < x_i, x_j >$

   $\implies g(x) = \sum_{l=1}^{n} \alpha_i < x_i, x > +b$

   Where

   $b = \frac{1}{2}\left[\frac{1}{n_-^2} \sum_{(i,j) \in I_-} < x_i, x_j > - \frac{1}{n_+^2} \sum_{(i,j) \in I_+} < x_i, x_j > \right]$

   $\alpha_i = \frac{1}{n_+}$ when $y_i = +1$; $\alpha_i = -\frac{1}{n_-}$ when $y_i = -1$

2. A code in R for this classifier is provided in D2L. Modify this code, or write your own in the language of your choice such that you can compute a classifier for the Iris data. The Iris dataset is described and is also available at `https://en.wikipedia.org/wiki/Iris_flower_data_set`. Create a classifier for the labels "I. setosa" versus "I. versicolor" using 80% of the data. compute the classification error using the 20% remaining. Then, repeat the same thing for the labels "I. virginica" versus "I. versicolor". Report your results in a clear and concise form.

«echo=T,fig.height=5,fig.width=5,fig.align='center'»= rm(list=ls()) set.seed(0.1) @

A few kernel functions

«echo=T,fig.height=5,fig.width=5,fig.align='center'»= k1 <- function(x,y) sum(x*y)

k2 <- function(x,y) sum(x*y)+1 k3 <- function(x,y) (1+sum(x*y))$^2 d < -4k4 < -function(x,y)(1 + sum($
$-1k5 < -function(x, y)exp(-sum((x - y)^2)/(2 * sigma^2))kappa < -1theta < -1k6 < -function(x, y)tanh$
$-function(x, y)k1(x, y)$@

generate some data in 2d

«echo=T,fig.height=5,fig.width=5,fig.align='center'»= n.p=10 n.m=10 n=n.p+n.m library(mvtnorm) x.p=rmvnorm(n=n.p,mean=c(2,2),sigma=diag(rep(1,2))) x.m=rmvnorm(n=n.m,mean=c(1,1),sigma=diag(re y = c(rep(1,n.p),rep(-1,n.m)) x=rbind(x.p,x.m) k.mm=outer(1:n.m,1:n.m,Vectorize(function(i,j) k(x.m[i,],x.m[k.pp=outer(1:n.p,1:n.p,Vectorize(function(i,j) k(x.p[i,],x.p[j,]))) b=(sum(k.mm)/(n.m*n.m)-sum(k.pp)/(n.p*n alpha=c(rep(1/n.p,n.p),rep(-1/n.m,n.m)) @

«echo=F, message=F, warning=F, fig.width=9, fig.height=4, fig.align='center'»= g.n=50 x.min=min(x) x.max=max(x) y.hat=matrix(NA,nrow=g.n,ncol=g.n) g=seq(from=x.min,to=x.max,length.out=g.n) for (i in (1:g.n)) for (j in (1:g.n)) u=c(g[i],g[j]) k.x=outer(1:n,1,Vectorize(function(i,j) k(x[i,],u))) y.hat[i,j]=sum(k.x*alpha)+b contour(x=g,y=g,z=y.hat,asp=1) points(x.p,col=4,pch=16) points(x.m,col=2,pch=16) @

Figure 1: evaluate the classifier over a grid

2 Perceptron

Consider a training set $\{(x_1, y_1), \dots, (x_n, y_n)\}$, with $x_i \in \mathbb{R}^d$ and $y_i \in -1, 1$. The perceptron is one of the oldest algorithm in machine learning. Historical notes are provide at `https://en.wikipedia.org/wiki/Perceptron`. The perceptron is a linear classifier $f(x) = w^T x$ where $w \in \mathbb{R}^d$. The algorithm for computing w is as follows:

1. Write the kernalized perceptron algorithm. Hint: assume that the kernalized perceptron classifier can be written as
$$f(x) = \sum_{i=1}^{n} \alpha_i < \phi(x_i), \phi(x) >$$
for some function ϕ and that the algorithm above corresponds to the situation when ϕ is the identity function and $< .,. >$ is the usual inner product in $\mathbb{R}^d$. Initialize with $\alpha_1 = \dots = \alpha_n = 0$. Provide a pseudo-code.

2. Write the code for data in 2 dimensions, similarly than for the simple classifier. Show 3 examples using 3 different kernels.

3 Kernels over $\mathcal{X} = \mathbb{R}^2$

Let $x = (x_1, x_2) \in \mathbb{R}^2$ and $y = (y_1, y_2) \in \mathbb{R}^2$,

1. Let
$$\phi(x) = (x_1^2, \sqrt{2} x_1 x_2, x_2^2)$$
Verify that $\phi(x)^T \phi(y) = (x^T y)^2$

2. Find a function $\phi(x)$:$\mathbb{R}^2 \mapsto \mathbb{R}^6$ such that for any (x, y), $\phi(x)^T \phi(y) = (x^T y + 1)^2$

3. Find a function $\phi(x)$:$\mathbb{R}^2 \mapsto \mathbb{R}^9$ such that for any (x, y), $\phi(x)^T \phi(y) = (x^T y + 1)^2$

4. Verify that
$$K(x, y) = (1 + x^T y)^d$$
for $d = 1, 2 \dots$ is a positive definite kernel

5. Can you find a function ϕ:$\mathbb{R}^2 \mapsto H$, where H is an inner product space such that for any (x, y), $< \phi(x), \phi(y) >_H = x^T y - 1$?