20191002

kernel

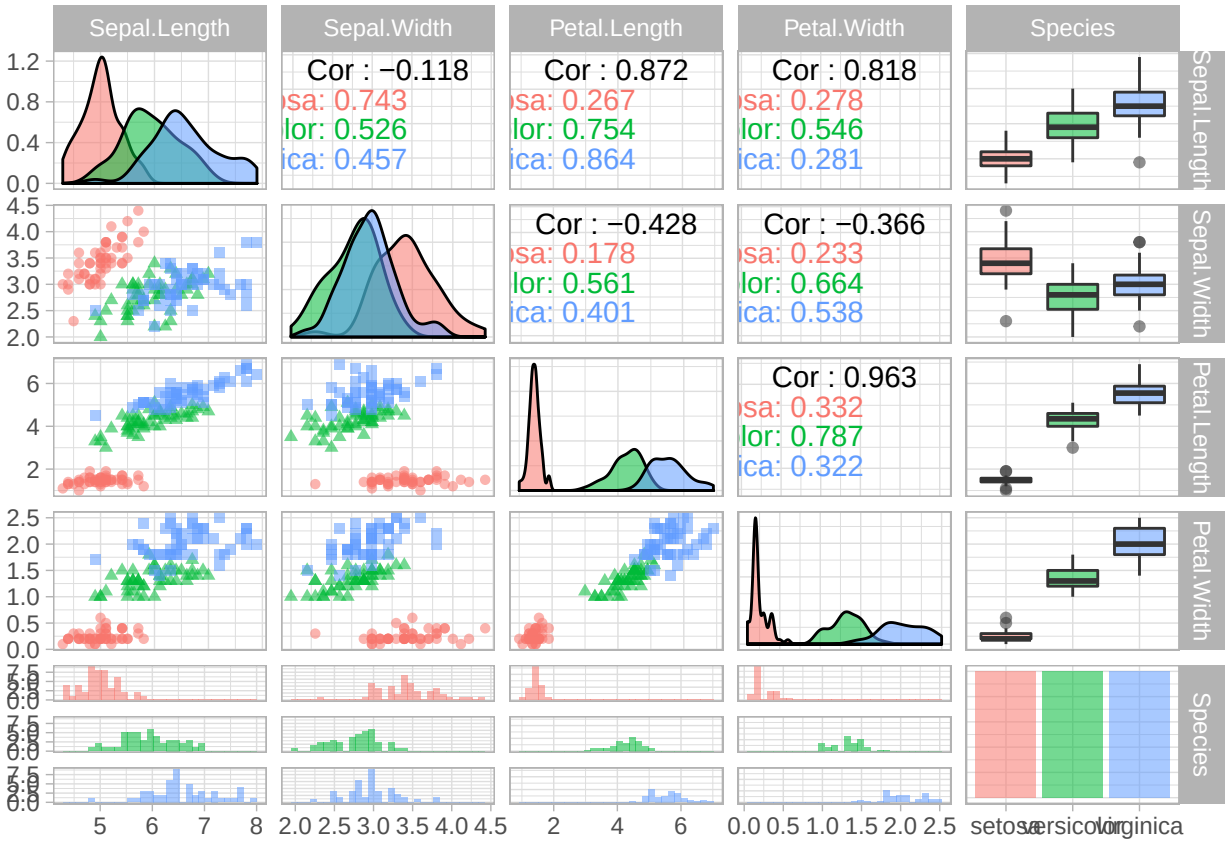
- Cat and Dog problem
- A simple geometric solution
- A more general solution

Dot product $\vec{a}\vec{b} = a_x b_x + a_y b_y = |\vec{a}||\vec{b}|\cos(\theta)$
Exercise:
$g(x) =< C_+ - C_-, X - C >=< C_+, X > - < C_-, X > - < C_+, C > + < C_-, C >;$
$< C_+, X >=< \frac{1}{n_+} \sum_{l \in I_+}^{n} x_i, x >;$
$< C_-, X >=< \frac{1}{n_-} \sum_{l \in I_-}^{n} x_i, x >;$
$< C_+, C >=< C_+, \frac{1}{2}C_+ > + < C_+, \frac{1}{2}C_- >= \frac{1}{2n_+^2} \sum_{(i,j) \in I_+} < x_i, x_j > + \frac{1}{2} < C_+, C_- >$
$< C_-, C >=< C_-, \frac{1}{2}C_+ > + < C_-, \frac{1}{2}C_- >= \frac{1}{2} < C_+, C_- > + \frac{1}{2n_-^2} \sum_{(i,j) \in I_-} < x_i, x_j >$
$g(x) = \sum_{l=1}^{n} \alpha_i < x_i, x > + b,$
$b = \frac{1}{2}\left[\frac{1}{n_-^2} \sum_{(i,j) \in I_-} < x_i, x_j > - \frac{1}{n_+^2} \sum_{(i,j) \in I_+} < x_i, x_j >\right]$
$\alpha_i = \begin{cases} \frac{1}{n_+} & y_i = +1 \\ -\frac{1}{n_-} & y_i = -1 \end{cases}$

- Import the iris data

```
rm(list=ls())
library(datasets)
data(iris)
```

```
library(ggplot2)
GGally::ggpairs(iris,mapping=aes(color =Species,shape=Species,alpha=0.3),
        columns=c("Sepal.Length","Sepal.Width","Petal.Length","Petal.Width","Species"))+theme_light()
```



- Define the train and test sets

```
iris$class <- NA
iris_setosa <- iris[iris$Species=="setosa",]
iris_versicolor <- iris[iris$Species=="versicolor",]
iris_virginica <- iris[iris$Species=="virginica",]

iris_train_se<- iris_setosa[1:40,]
iris_train_ve<- iris_versicolor[1:40,]
iris_train_vi<- iris_virginica[1:40,]

iris_test_se<- iris_setosa[41:50,]
iris_test_ve<- iris_versicolor[41:50,]
iris_test_vi<- iris_virginica[41:50,]

iris_train_se.ve<- rbind(iris_train_se,iris_train_ve)
iris_train_ve.vi<- rbind(iris_train_ve,iris_train_vi)
iris_test_se.ve<- rbind(iris_test_se,iris_test_ve)
iris_test_ve.vi<- rbind(iris_test_ve,iris_test_vi)
```

- Define the kernel function and Computing the classifier

```
k = function(x,y)  return(sum(x*y)+1)
k.pp=outer(1:40,1:40,Vectorize(function(i,j) k(iris_train_se[i,1:4],iris_train_se[j,1:4])))
k.mm=outer(1:40,1:40,Vectorize(function(i,j) k(iris_train_ve[i,1:4],iris_train_ve[j,1:4])))
b=(sum(k.mm)/(40^2)-sum(k.pp)/(40^2))/2
alpha=ifelse(iris_train_se.ve$Species=="setosa",1/40,-1/40)

k.x=outer(1:80,1:20,Vectorize(function(i,j) k(iris_train_se.ve[i,1:4],iris_test_se.ve[j,1:4])))
iris_test_se.ve[,6]=(t(k.x)%*%alpha+b)
```

```
k = function(x,y)  return(sum(x*y)+1)
k.pp=outer(1:40,1:40,Vectorize(function(i,j) k(iris_train_vi[i,1:4],iris_train_vi[j,1:4])))
k.mm=outer(1:40,1:40,Vectorize(function(i,j) k(iris_train_ve[i,1:4],iris_train_ve[j,1:4])))
b=(sum(k.mm)/(40^2)-sum(k.pp)/(40^2))/2
alpha=ifelse(iris_train_ve.vi$Species=="virginica",1/40,-1/40)

k.x=outer(1:80,1:20,Vectorize(function(i,j) k(iris_train_ve.vi[i,1:4],iris_test_ve.vi[j,1:4])))
iris_test_ve.vi[,6]=(t(k.x)%*%alpha+b)
```

- Evaluate the classifier

```
iris_test_se.ve$evaluate=ifelse(iris_test_se.ve$class>0,"setosa","versicolor")
1-length(which(iris_test_se.ve$Species==iris_test_se.ve$evaluate))/20
```

```
## [1] 0
```

```
iris_test_ve.vi$evaluate=ifelse(iris_test_ve.vi$class>0,"virginica","versicolor")
1-length(which(iris_test_ve.vi$Species==iris_test_ve.vi$evaluate))/20
```

```
## [1] 0.05
```

Table 1: Confusion matrix

	Actural Species					
	test 1	Setosa	Versicolor	test 2	Virginica	Versicolor
Test Species	Setosa	10	0	Virginica	9	0
	Versicolor	0	10	Versicolor	1	10

Error rate = 0% and 5% in two tests respectively.

setosa v.s.versicolor			
	Species	class	evaluate
41	setosa	5.8564094	setosa
42	setosa	5.5356594	setosa
43	setosa	6.3496594	setosa
44	setosa	4.6646594	setosa
45	setosa	4.1349094	setosa
46	setosa	5.4289094	setosa
47	setosa	5.2151594	setosa
48	setosa	5.8694094	setosa
49	setosa	5.2391594	setosa
50	setosa	5.5476594	setosa
91	versicolor	-5.0968406	versicolor
92	versicolor	-6.2058406	versicolor
93	versicolor	-4.2455906	versicolor
94	versicolor	-1.4460906	versicolor
95	versicolor	-4.6668406	versicolor
96	versicolor	-4.4508406	versicolor
97	versicolor	-4.6295906	versicolor
98	versicolor	-5.4015906	versicolor
99	versicolor	-0.6630906	versicolor
100	versicolor	-4.4110906	versicolor

virginica v.s.versicolor			
	Species	class	evaluate
91	versicolor	-1.57497812	versicolor
92	versicolor	-0.74947812	versicolor
93	versicolor	-1.90722813	versicolor
94	versicolor	-3.48222812	versicolor
95	versicolor	-1.68972812	versicolor
96	versicolor	-1.63847812	versicolor
97	versicolor	-1.59247813	versicolor
98	versicolor	-1.15722812	versicolor
99	versicolor	-3.70797812	versicolor
100	versicolor	-1.73947812	versicolor
141	virginica	1.56602187	virginica
142	virginica	0.97952188	virginica
143	virginica	-0.02222812	versicolor
144	virginica	1.96827188	virginica
145	virginica	1.79502188	virginica
146	virginica	0.96802188	virginica
147	virginica	0.11902188	virginica
148	virginica	0.65352188	virginica
149	virginica	0.99177188	virginica
150	virginica	0.02902187	virginica