

Multivariate Statistical Analysis - Statistical Depth

P. Ilmonen and N. Vesselinova, March 2023

Statistical depth answers the question of how typical a given observation is with respect to a given sample. Formally, statistical depth measures the centrality of any p -variate random vector y with respect to a multivariate sample (data set) or a cumulative distribution function. It is always between 0 and 1, where 0 indicates atypical—relative to the given data set—data point and 1 denotes central—to the given sample—observation. As the statistical depth can be considered an indicator of the typicality of the data, the larger its value is, the more typical the observation is for the considered data set.

Statistical depth finds practical application in classification problems among many other. Different statistical depth functions have been developed along the years (with different applications in mind). The Mahalanobis depth and half-space depth are two examples.

The R package `ddalpha` contains functions that allow us to compute these statistical depths, namely

- `depth.Mahalanobis()`
- `depth.halfspace()`

Demo problem 1: Half-space depth

To illustrate some of the concepts we learnt about statistical depth, we will focus on half-space depth.

Let $S_n = \{x_1, \dots, x_n\}$ denote a set of p variate observations from a distribution F_x . The half space depth $D_H(y, S_n)$ is defined as follows.

$$D_H(y, S_n) = \min_{u \in U} \frac{1}{n} |\{x_i \in S_n \mid u^T(x_i - y) \geq 0\}|,$$

where U denotes the unit sphere in R^p .

Note that, as the name suggest, the half-space depth is enclosed between 0 and 0.5.

For the practical calculation of half-space, please have a look at the video available on [Mycourses MS-E2112 Multivariate Statistical Analysis](#) course web site in the ‘Lectures’ folder.

Let us install the package `ddalpha` first.

```
#install.packages("ddalpha")
```

Next, we will load this package and the `MASS` library as we will make use of the `Iris` data set.

```
library(MASS)
library(ddalpha)
```

Load the dataset.

```
data(iris)
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1           5.1         3.5          1.4          0.2  setosa
## 2           4.9         3.0          1.4          0.2  setosa
```

```
## 3      4.7      3.2      1.3      0.2 setosa
## 4      4.6      3.1      1.5      0.2 setosa
## 5      5.0      3.6      1.4      0.2 setosa
## 6      5.4      3.9      1.7      0.4 setosa
```

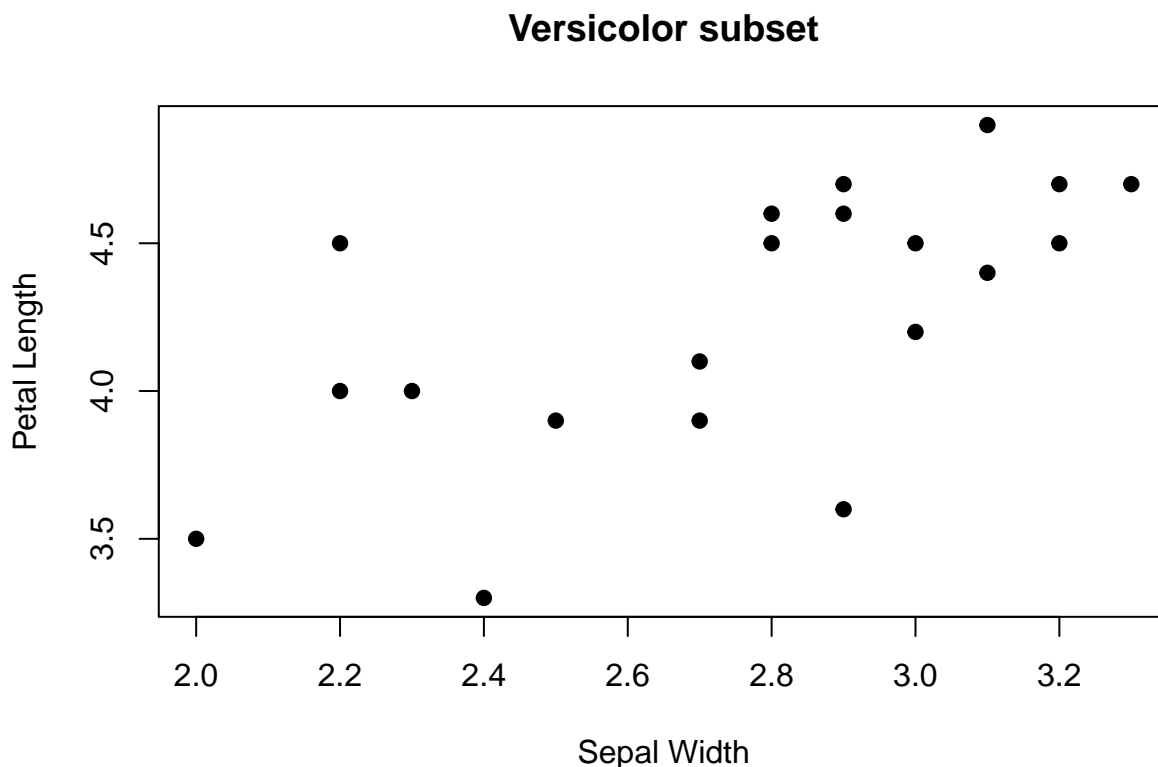
We will use a two dimensional subset of the data set `Iris`, consisting of the variables `Sepal.Width` and `Petal.Length` measured on `versicolour`.

```
# Slicing the iris data set
features <- c(2:3)
versicolor <- iris[51:70, features]
dim(versicolor)[1]
```

```
## [1] 20
```

We have 20 data points in our new sample. Let us plot the data.

```
plot(versicolor, main="Versicolor subset",
     xlab="Sepal Width", ylab="Petal Length", pch=19)
```



a) Visualization of the data and measuring the half-space depth

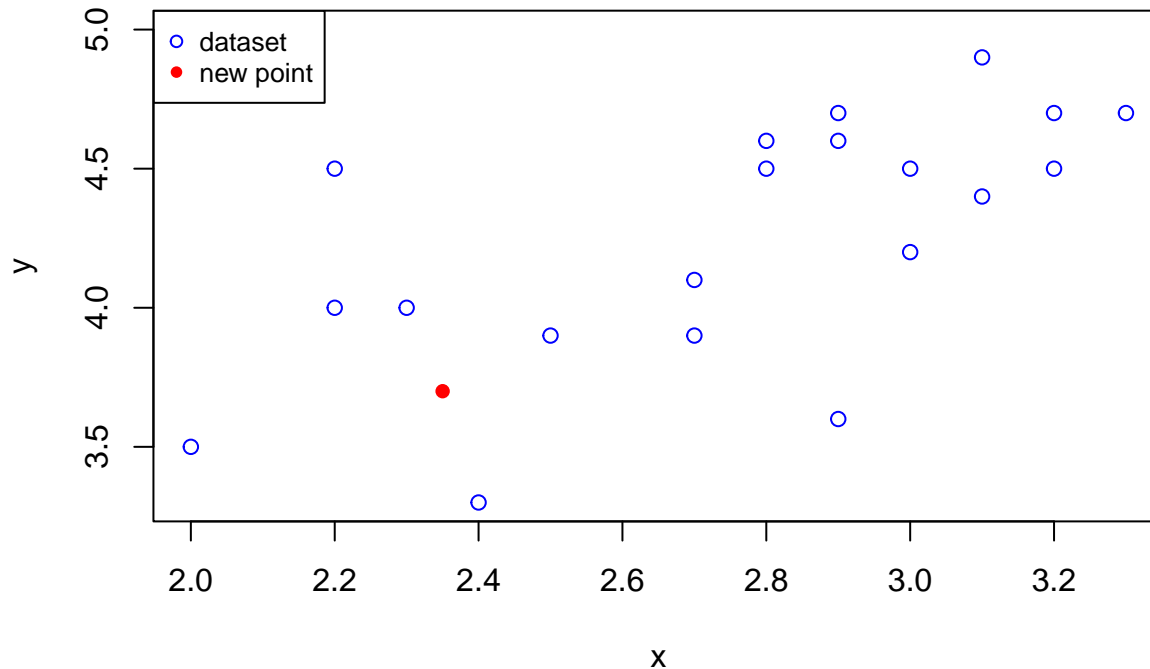
Create a new observation with parameters `Sepal.Width = 2.35` and `Petal.Length = 3.7`. Plot the sample together with the new observation using different colours. Determine the half-space depth for the new observation relevant to the sample by visually inspecting the plot. Calculate the half-space depth using R.

Add a new observation

```
newobs <- c(2.35, 3.7)
```

Visualise the data together with the new observation in different colours

```
plot(c(2, 3.3), c(3.3, 5), type="n",
     xlab="x", ylab="y")
points(versicolor, col="blue")
points(newobs[1], newobs[2], col="red", pch=16)
legend("topleft", col=c("blue","red"), cex=0.8, legend=c("dataset", "new point"),
      pch=c(1,16))
```



Can you visually determine the half-space depth of the red point? How much is it? (Recall that the sample has 20 data points.)

```
help(depth.halfspace)
```

Let us now use R to verify our answer.

```
depth <- depth.halfspace(newobs ,versicolor, exact=T)
depth
```

```
## [1] 0.1
```

b) Atypical observation?

We will look at a data set constructed by a mixture of points sampled from a gamma and a Weibull distributions.

First of all, to be able to reproduce the results, let us set the seed to a specific number

```
set.seed(123)
n <- 20
x1 <- rweibull(n,shape=1,scale=2)
x2 <- rgamma(n,shape=2,scale=1)

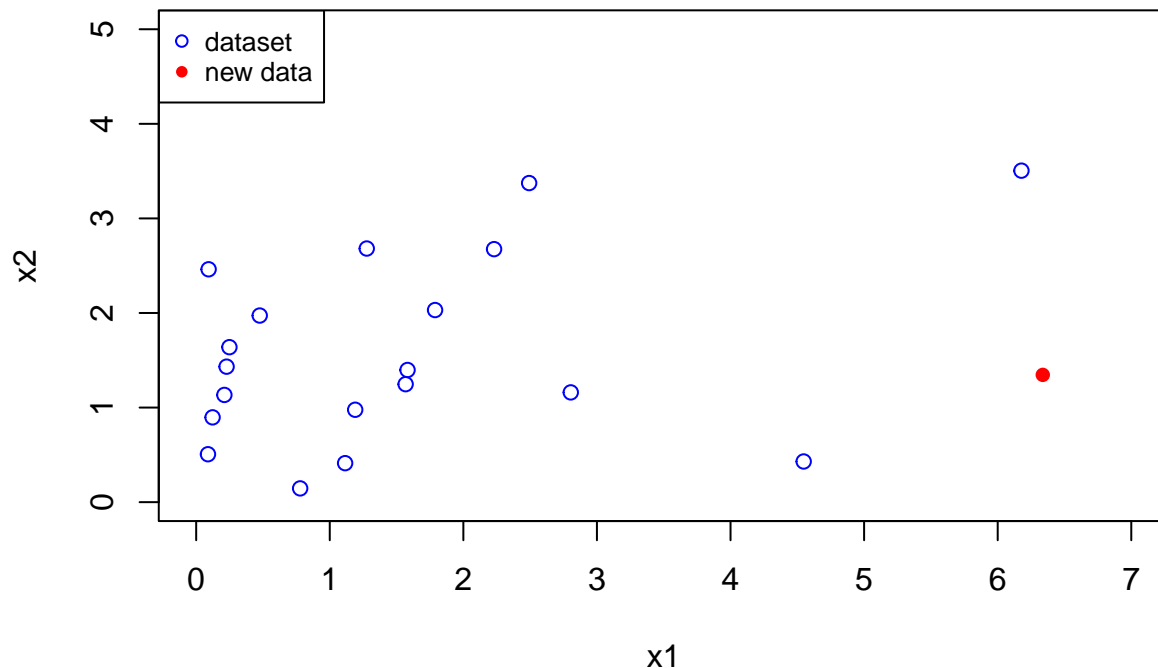
# The sample
sample2 <- cbind(x1,x2)
dim(sample2)
```

```
## [1] 20 2
sample3 <- sample2[-18,]
dim(sample3)
```

```
## [1] 19 2
# and the new point
newpoint <- sample2[18,]
```

Let us plot `sample3` together with the new observation

```
plot(c(0, 7), c(0, 5), type="n",
     xlab="x1", ylab="x2")
points(sample3,col="blue")
points(newpoint[1], newpoint[2], col="red",pch=16)
legend("topleft", col=c("blue","red"), cex=0.8, legend=c("dataset", "new data"),
      pch=c(1,16))
```



How much is the depth of the red point? Can you determine it just by inspecting the plot?

Verify the visual calculation

```
depth <- depth.halfspace(newpoint, sample3, exact=T)
depth
```

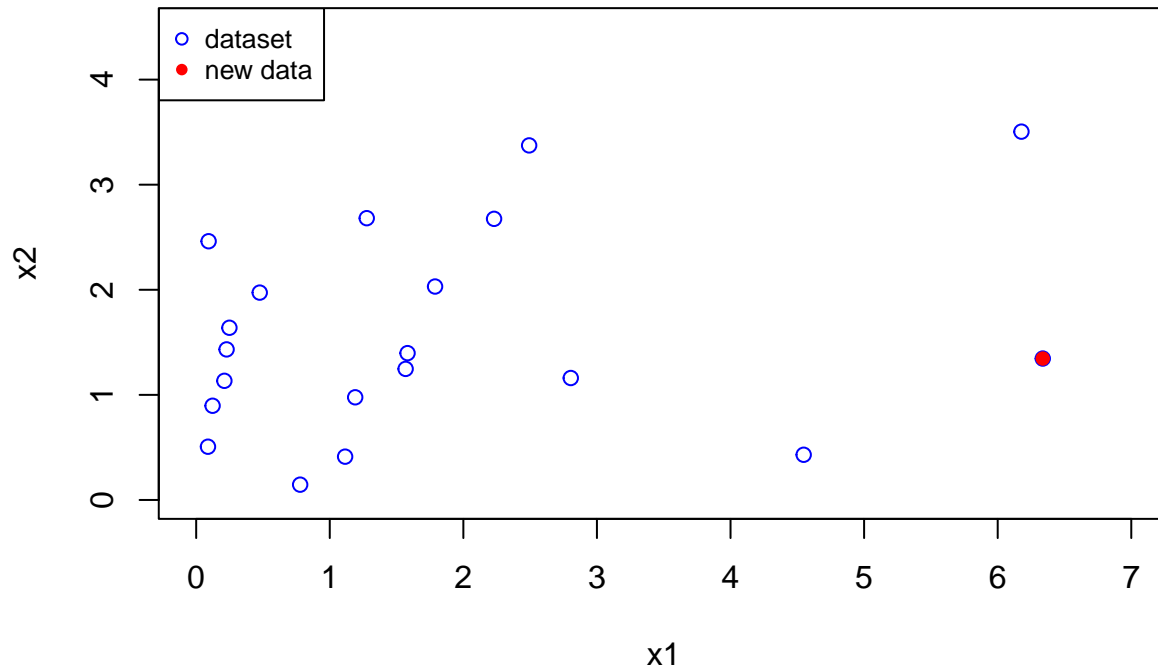
```
## [1] 0
```

What if the red point were not a new observation but a data point from the sample. Would the half-space depth change in any way?

Let us first plot the original `sample2` and this same data point, which is part of `sample2` dataset

```
plot(c(0, 7), c(0, 4.5), type="n",
     xlab="x1", ylab="x2")
points(sample2,col="blue")
points(newpoint[1], newpoint[2], col="red",pch=16)
```

```
legend("topleft", col=c("blue","red"), cex=0.8, legend=c("dataset","new data"),
      pch=c(1,16))
```

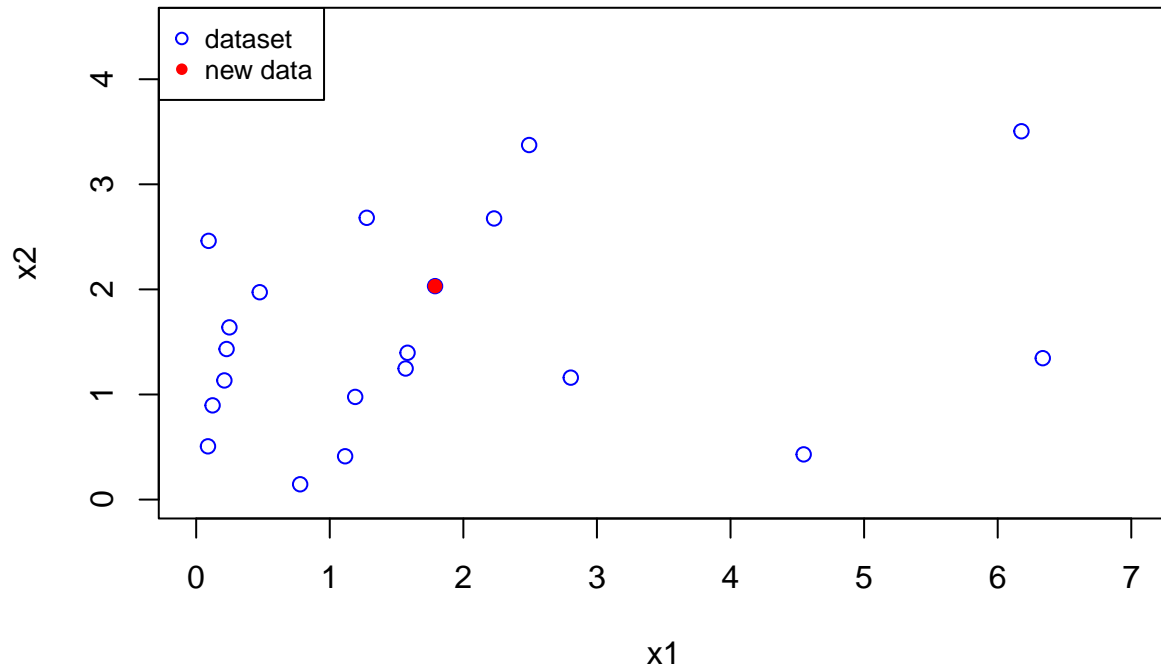


```
depth <- depth.halfspace(newpoint, sample2, exact=T)
depth
```

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

c) How typical (deep) is an observation from a dataset?

```
pointsample <- c(1.78819310, 2.0309972)
plot(c(0, 7), c(0, 4.5),type="n",
     xlab="x1",ylab="x2")
points(sample2,col="blue")
points(pointsample[1], pointsample[2], col="red",pch=16)
legend("topleft",col=c("blue","red"),cex=0.8,legend=c("dataset","new data"),
      pch=c(1,16))
```



```
depth <- depth.halfspace(pointsample, sample2, exact=T)
depth
```

```
## [1] 0.3
```