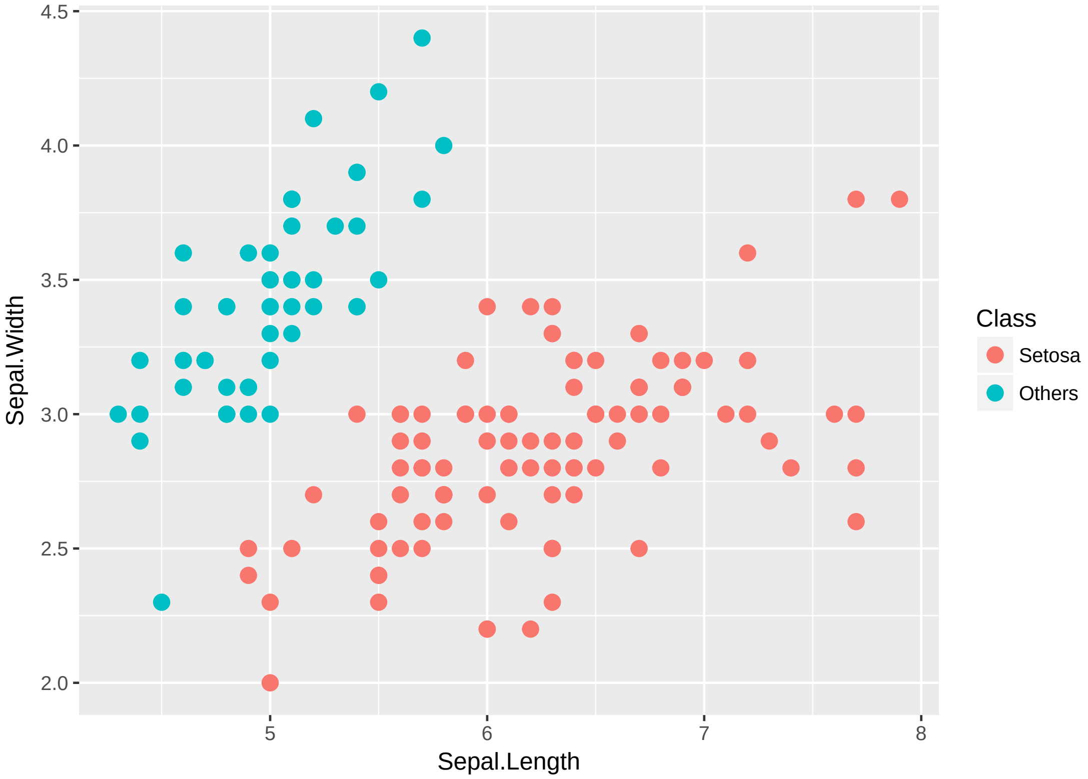
0.1 Fisher's iris data

We will look at Fisher's iris data set as a toy example. There are originally three categories, but we combined the last two categories, so it is now only Setosa flowers and the rest. The four covariates give various measurements of the flowers.

```
## 'data.frame': 150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 ..
```



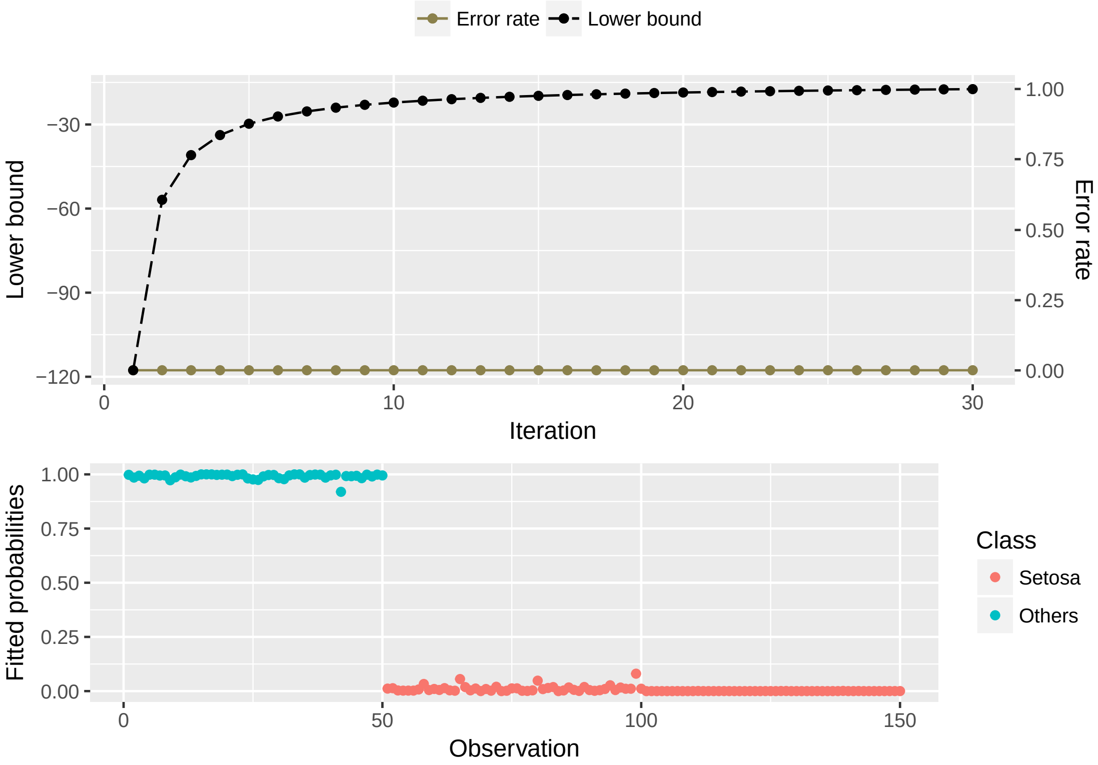
Fitting the probit I-prior model doesn't long at all (almost instantly). The VB-EM algorithm was run for 100 iterations. However, it is seen that it converges after just 30 iterations.

```
system.time(mod <- iprobit(y, X, silent = TRUE, maxit = 100))
```

```
##    user  system elapsed
##   0.987   0.028   1.018
```

```
print(mod)
```

```
## Training error rate: 0.00 %
```



0.2 Ionosphere data

Data was obtained from `https://archive.ics.uci.edu/ml/datasets/Ionosphere`. This radar data was collected by a system in Goose Bay, Labrador. This system consists of a phased array of 16 high-frequency antennas with a total transmitted power on the order of 6.4 kW. The targets were free electrons in the ionosphere. “Good” radar returns are those showing evidence of some type of structure in the ionosphere. “Bad” returns are those that do not; their signals pass through the ionosphere. Received signals were processed using an autocorrelation function whose arguments are the time of a pulse and the pulse number. There were 17 pulse numbers for the Goose Bay system. Instances in this databse are described by 2 attributes per pulse number, corresponding to the complex values returned by the function resulting from the complex electromagnetic signal.

In total, there are 350 observations, and 34 covariates. Of the 350 observations, 126 were bad signals and the rest were good. Convergence was achieved pretty quickly – only after about 12 iterations.

```
##   b   g
## 126 224
```

We randomly split the data into a training set of size 200, and the remaining is used for validation. We obtain a training error rate of zero, but also a test error rate of 7.33%. The FBM kernel was used.

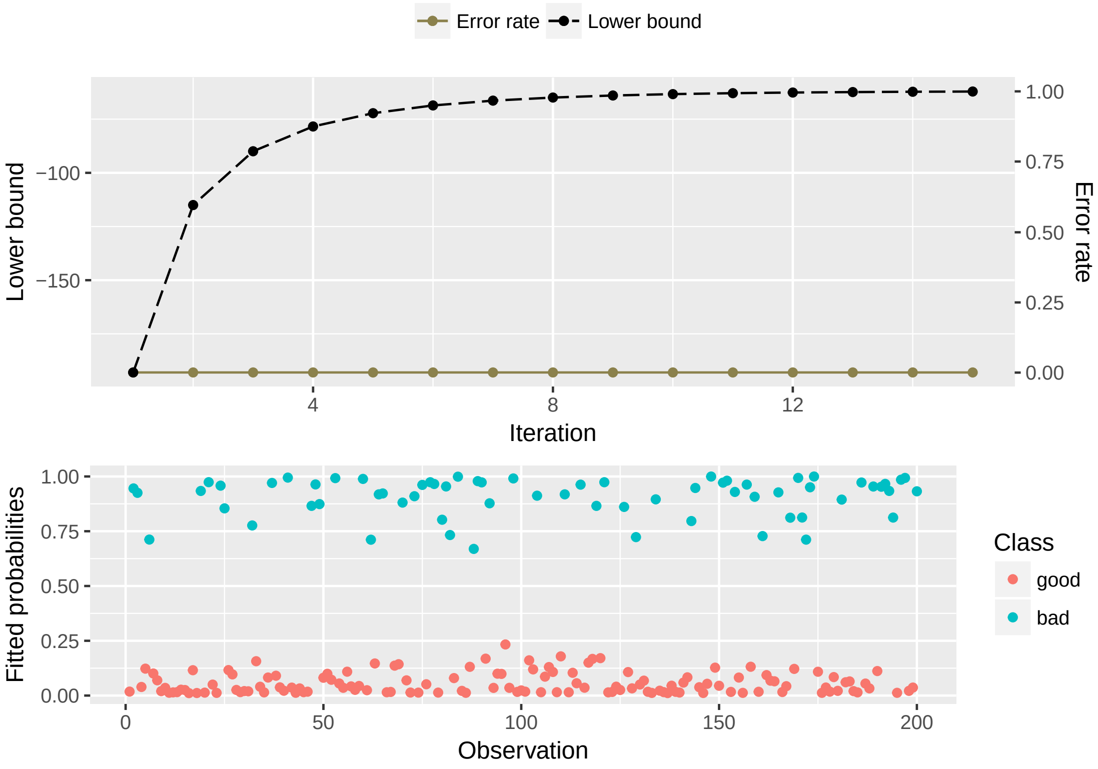
```
mod <- iprobit(y.train, X.train, kernel = "FBM", silent = TRUE)
print(mod)
```

```
## Training error rate: 0.00 %
```

```
print(mod, X.test, y.test)  # Test error rate
```

```
## Test error rate: 7.33 %
```

```
plot(mod, 15, levels = c("good", "bad"))
```



0.3 Cardiac data

Data was obtained from `https://archive.ics.uci.edu/ml/datasets/Arrhythmia`. This database contains 279 attributes, 206 of which are linear valued and the rest are nominal. Concerning the study of H. Altay Guvenir: "The aim is to distinguish between the presence and absence of cardiac arrhythmia and to classify it in one of the 16 groups. Class 1 refers to normal ECG, whereas classes 2 to 15 refers to different classes of arrhythmia, and class 16 refers to the rest of unclassified ones. For the time being, there exists a computer program that makes such a classification. However there are differences between the cardiolog's and the programs classification. Taking the cardiolog's as a gold standard we aim to minimise this difference by means of machine learning tools."

We combined the 15 of the arrhythmia classes into a single class, and normal being a class on its own. Out of the 451 ECG observations, 245 were classed as normal, and 206 as having some form of arrhythmia. To test predictions, we split the data set into a training set of size 300, and the remaining was left for validation. We fit a probit I-prior model with an FBM kernel.

```
##     Normal Arrhythmia
##        245        206
```

```
mod <- iprobit(y.train, X.train, kernel = "FBM", silent = F, maxit = 100)
```

```
##
  |
  |                                                                |   0%
  |
  |=                                                               |   1%
  |
  |=                                                               |   2%
  |
  |==                                                              |   3%
  |
  |===                                                             |   4%
  |
  |===                                                             |   5%
  |
  |====                                                            |   6%
  |
  |=====                                                           |   7%
  |
  |=====                                                           |   8%
  |
  |======                                                          |   9%
  |
  |=======                                                         |  10%
  |
  |=======                                                         |  11%
  |
  |========                                                        |  12%
  |
  |=========                                                       |  13%
  |
  |=========                                                       |  14%
  |
  |==========                                                      |  15%
  |
  |===========                                                     |  16%
  |
  |===========                                                     |  17%
  |
  |============                                                    |  18%
  |
  |============                                                    |  19%
```

```
|
|=============                                                    |  20%
|
|==============                                                   |  21%
|
|==============                                                   |  22%
|
|===============                                                  |  23%
|
|================                                                 |  24%
|
|================                                                 |  25%
|
|=================                                                |  26%
|
|==================                                               |  27%
|
|==================                                               |  28%
|
|===================                                              |  29%
|
|====================                                             |  30%
|
|====================                                             |  31%
|
|=====================                                            |  32%
|
|======================                                           |  33%
|
|======================                                           |  34%
|
|=======================                                          |  35%
|
|========================                                         |  36%
|
|========================                                         |  37%
|
|=========================                                        |  38%
|
|==========================                                       |  39%
|
|==========================                                       |  40%
|
|===========================                                      |  41%
|
|============================                                     |  42%
|
|============================                                     |  43%
```

```
  |
  |=============================                                     |  44%
  |
  |==============================                                    |  45%
  |
  |==============================                                    |  46%
  |
  |===============================                                   |  47%
  |
  |================================                                  |  48%
  |
  |================================                                  |  49%
  |
  |=================================                                 |  51%
  |
  |=================================                                 |  52%
  |
  |==================================                                |  53%
  |
  |===================================                               |  54%
  |
  |===================================                               |  55%
  |
  |====================================                              |  56%
## Converged after 56  iterations
```

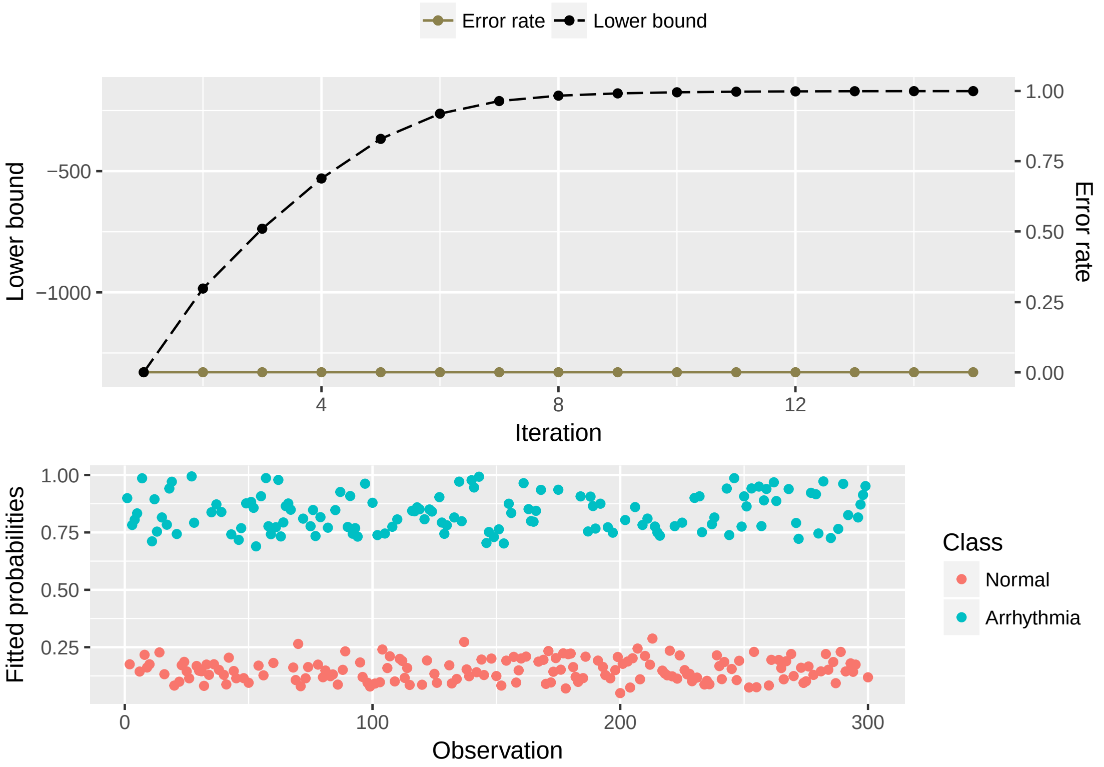
```
print(mod)
```

```
## Training error rate: 0.00 %
```

```
print(mod, X.test, y.test)  # Test error rate
```

```
## Test error rate: 28.67 %
```

```
plot(mod, 15, levels = c("Normal", "Arrhythmia"))
```


Error rate
Lower bound
Lower bound
−500
−1000
1.00
0.75
0.50
0.25
0.00
Error rate
4
8
12
Iteration
Fitted probabilities
1.00
0.75
0.50
0.25
0
100
200
300
Observation
Class
Normal
Arrhythmia