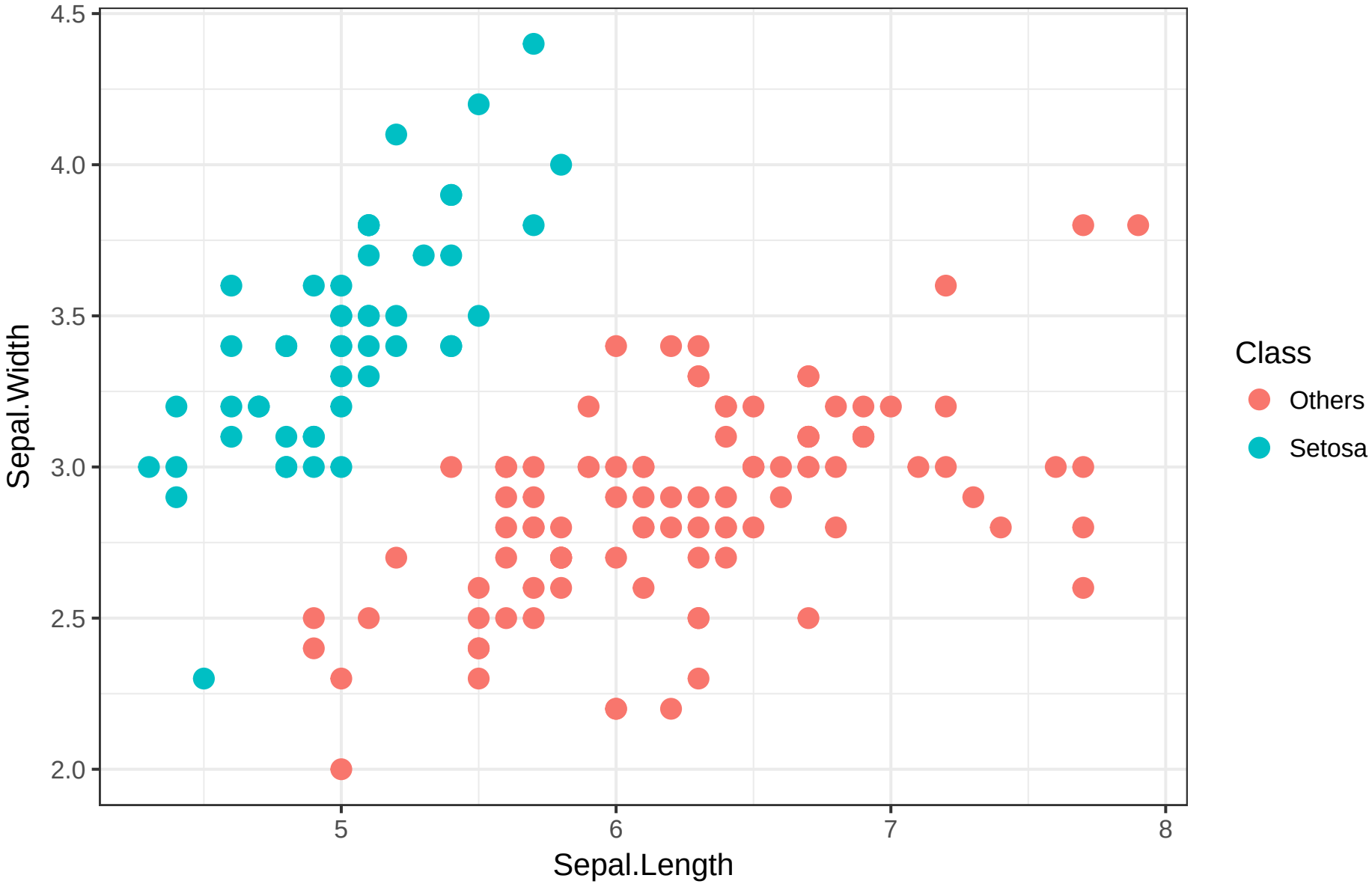
Fisher's Iris data set



Fisher's Iris data set - Model fitting

```
R> system.time(
+   (mod <- iprobit(y, X))
+ )
```

```
##
## |==================                                    |  61%
##  Converged after 6141 iterations.
##  Training error rate: 0 %
##     user  system elapsed
##   67.857   6.396  74.277
```

Fisher's Iris data set - Model summary

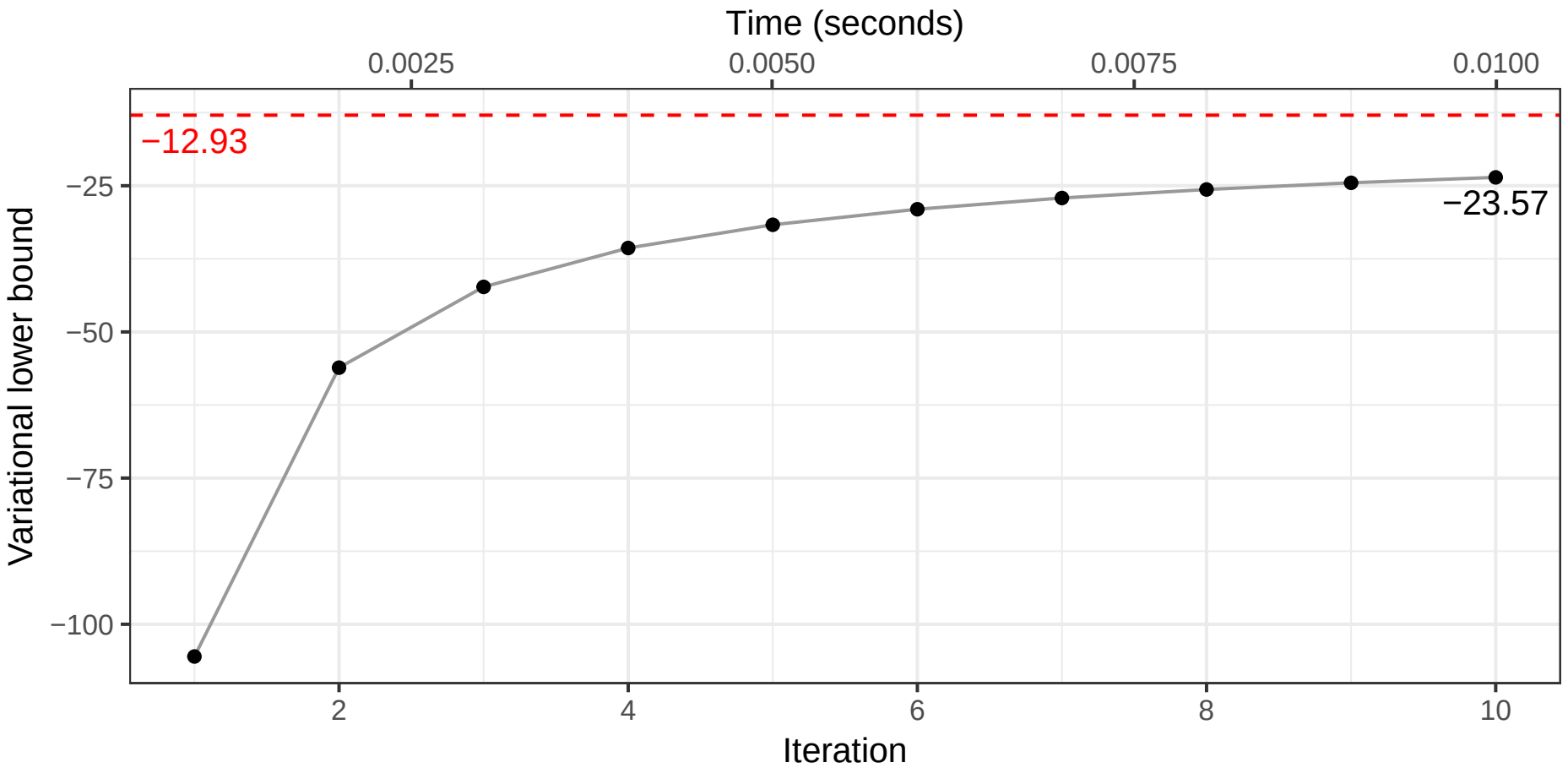
```
R> summary(mod)

##
## Call:
## iprobit(y = y, X, maxit = 10000)
##
## RKHS used: Canonical
##
##            Mean   S.E.     2.5%   97.5%
## alpha  -4.1730 0.0816 -4.3330 -4.0129
## lambda  1.2896 0.0142  1.2618  1.3175
##
## Converged to within 1e-05 tolerance. No. of iterations: 6141
## Model classification error rate (%): 0
## Variational lower bound: -12.93486
```

Fisher's Iris data set - Lower bound

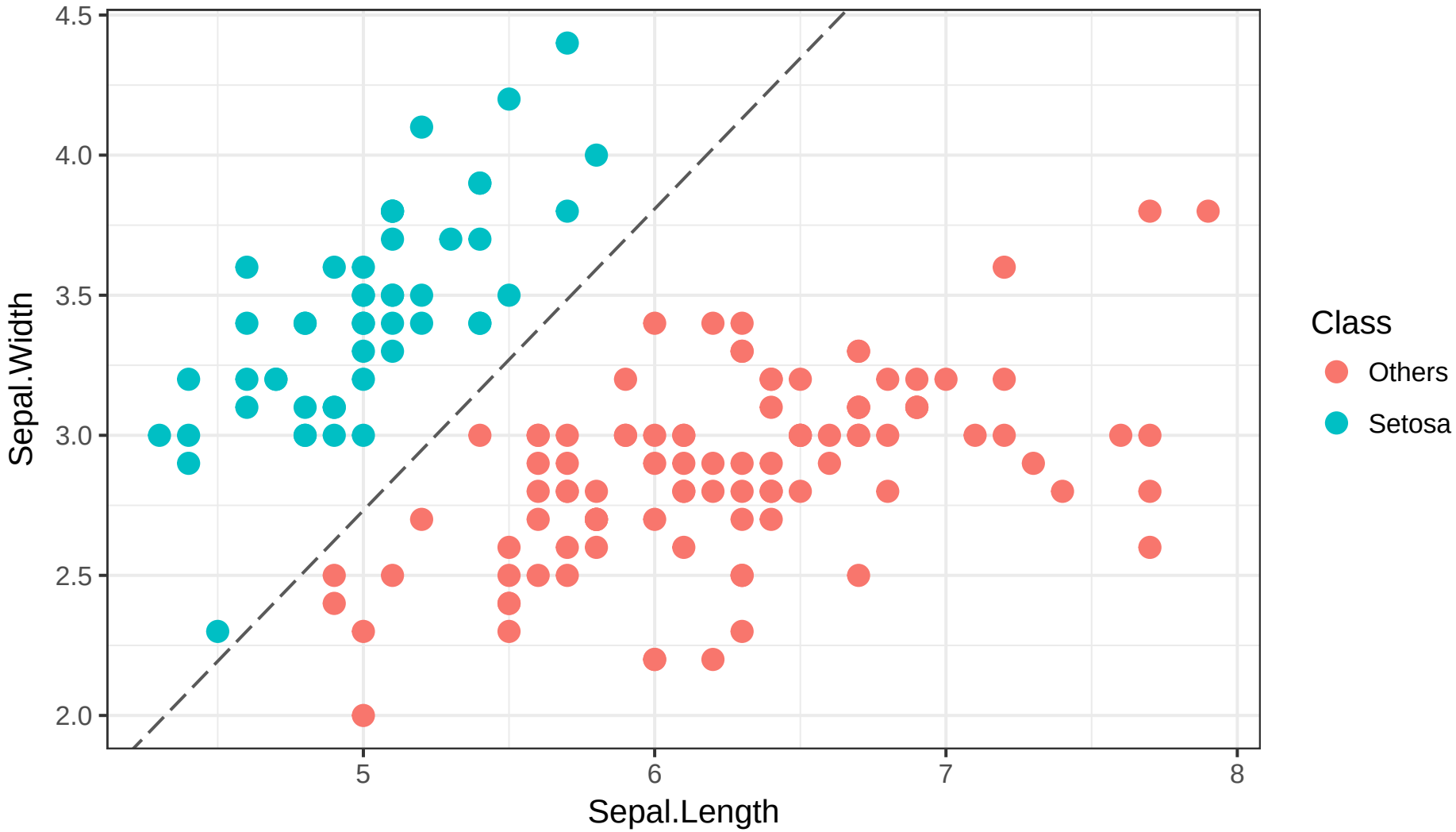
- Monitoring the lower bound (first 10 iterations)

```
R> iplot_lb(mod, niter.plot = 10)
```



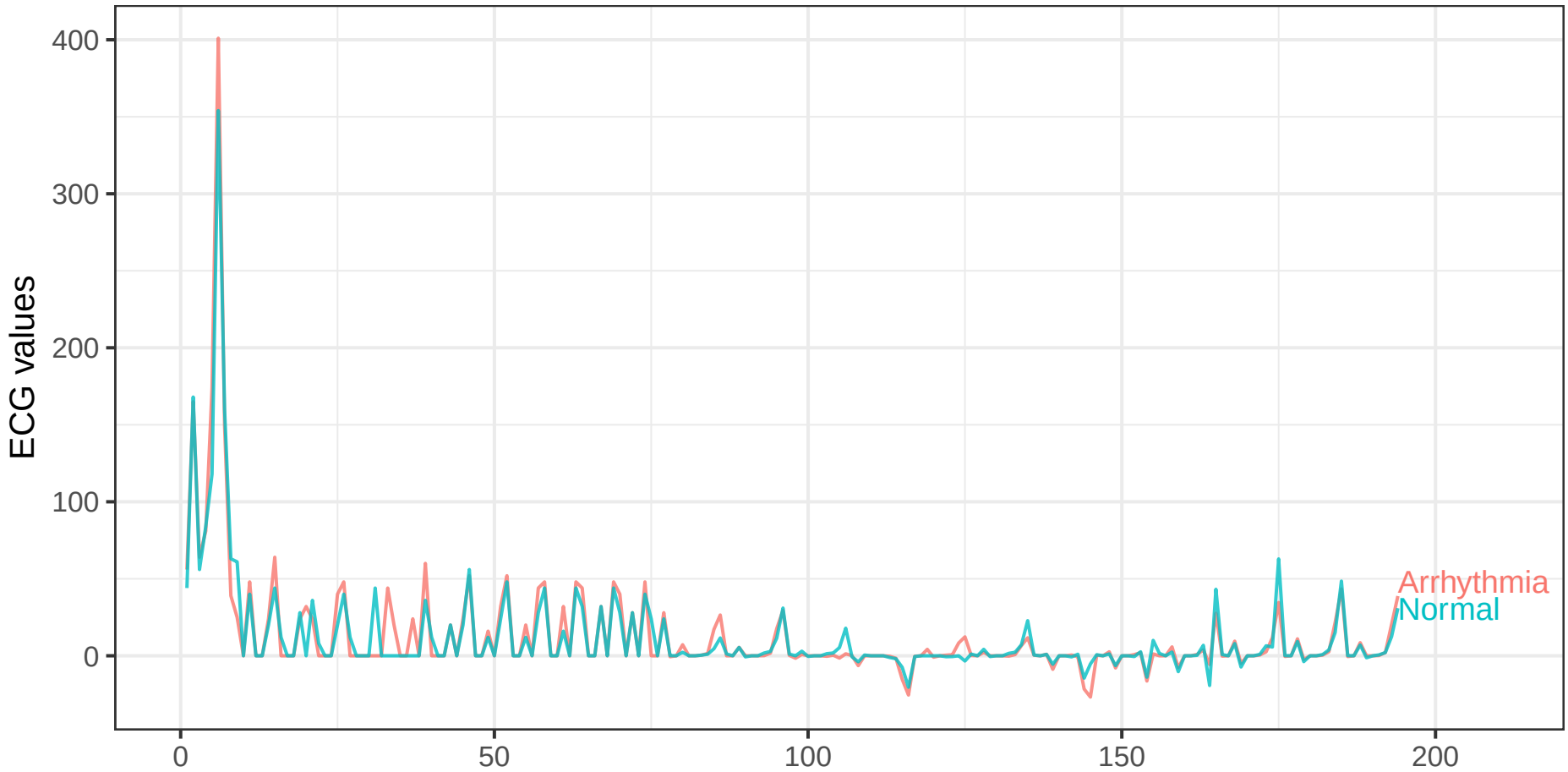
Fisher's Iris data set - Decision boundary

```
R> iplot_decbound(mod)
```



Cardiac arrhythmia data set

- Distinguish between the presence and absence of cardiac arrhythmia based on ECG data ($n = 451, p = 194$).



https://archive.ics.uci.edu/ml/datasets/Arrhythmia

Cardiac arrhythmia data set - Model fit

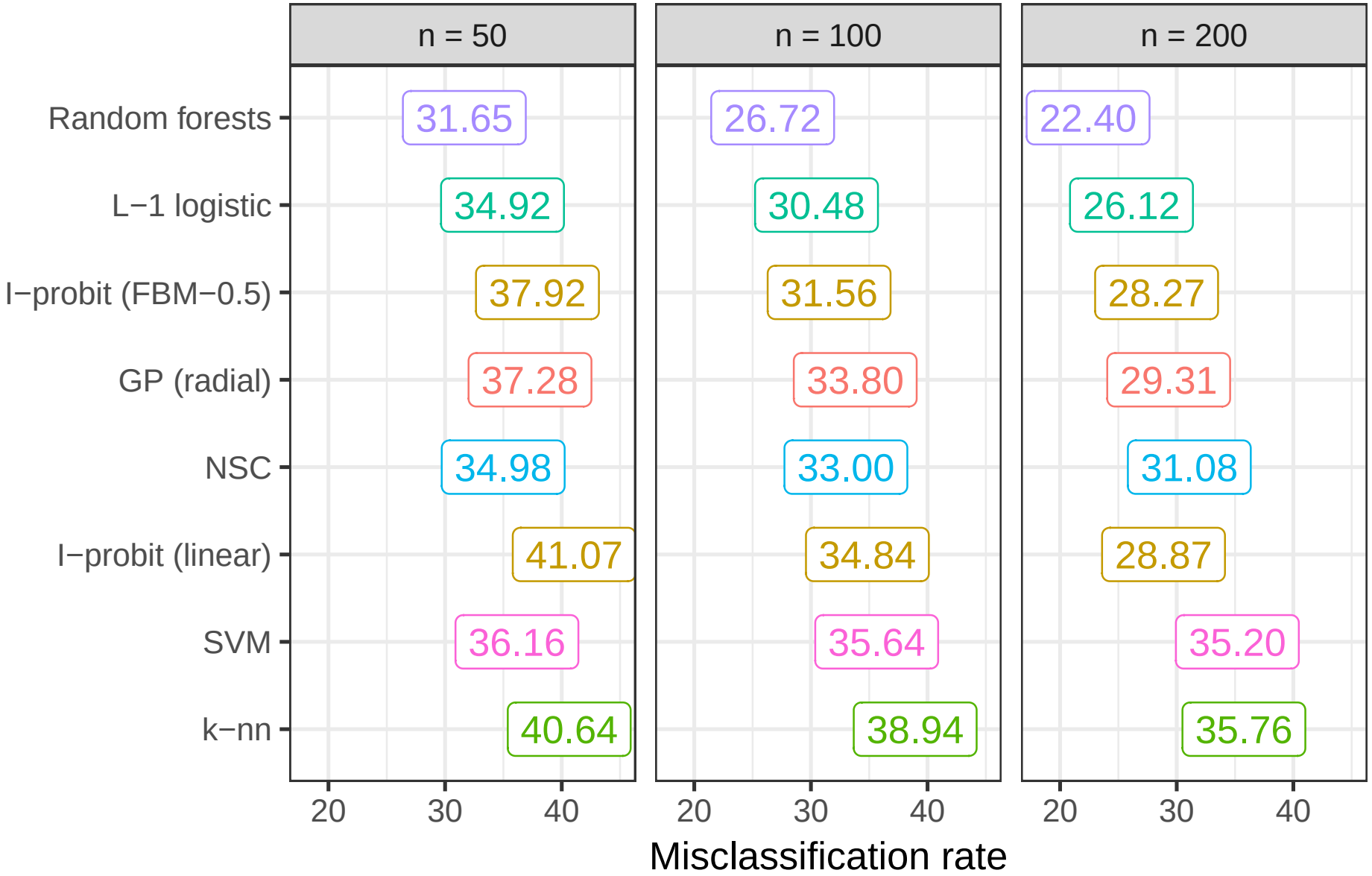
- Fit an I-prior probit model using Canonical and FBM kernel. The full data set takes about 35 seconds.

```
R> mod <- iprior(y, X, kernel = "FBM")
```

- Compare against popular classifiers: 1) k-nearest n3eighbours; 2) support vector machine; 3) Gaussian process classification; 4) random forests; 5) nearest shrunken centroids (Tibshirani et. al., 2003); and 6) L-1 penalised logistic regression.
- Experiment set-up:
  - Form training set by sub-sampling $n_{\text{sub}} \in \{50, 100, 200\}$ data points.
  - Use remaining data as test set.
  - Fit model on training set and obtain test error rates.
  - Repeat 100 times to obtain standard errors.

Cannings cite

Cardiac arrhythmia data set - Results



Meta-analysis of smoking cessation

- Data from 27 separate smoking cessation studies, where participants subjected to nicotine gum treatment or placed in control group.
- Some summary statistics:

	Min	Average	Max	Prop. quit	Odds quit
Control	20	101	617	0.207	0.261
Treated	21	117	600	0.320	0.470

cite Rabe Hasketh Skrondall

Meta-analysis of smoking cessation - model

- Let $i = 1, \ldots, n_j$ index the patients in study group $j \in 1, \ldots, 27$.
- Denote y_{ij} as the binary response variable indicating `Quit` (1) or `Remain` (0), and x_{ij} indicating which treatment group the patient is in.
- Model binary data using I-probit model

$$\begin{aligned}\Phi^{-1}(p_{ij}) &= f(x_{ij}, j) \\ &= f_1(x_{ij}) + f_2(j) + f_{12}(x_{ij}, j)\end{aligned}$$

with $f_1, f_2 \in$ Pearson RKHS, and $f_{12} \in$ ANOVA RKHS.

Meta-analysis of smoking cessation - results