

```

> library(tidyverse)
> library(caret)
> library(dslabs)
> library(rpart)
> data(tissue_gene_expression)
> names(tissue_gene_expression)
[1] "x" "y"
> dim(tissue_gene_expression$x)
[1] 189 500
> dim(as.matrix(tissue_gene_expression$y))
[1] 189 1
>
> set.seed(1991, sample.kind = 'Rounding')
Warning message:
In set.seed(1991, sample.kind = "Rounding") :
  non-uniform 'Rounding' sampler used
> fit <- train(tissue_gene_expression$x, tissue_gene_expression$y, method = 'rpart', tuneGrid = data.frame
(cp = seq(0, 0.1, 0.01)))
> plot(fit)
>
>
> set.seed(1991, sample.kind = 'Rounding')
Warning message:
In set.seed(1991, sample.kind = "Rounding") :
  non-uniform 'Rounding' sampler used
> fit <- train(tissue_gene_expression$x, tissue_gene_expression$y, method = 'rpart', tuneGrid = data.frame
(cp = seq(0, 0.1, 0.01)), control = rpart.control(minsplit = 0))
> plot(fit)
> confusionMatrix(fit)
Bootstrapped (25 reps) Confusion Matrix

```

(entries are percentual average cell counts across resamples)

	Reference						
Prediction	cerebellum	colon	endometrium	hippocampus	kidney	liver	placenta
cerebellum	19.5	0.0	0.2	0.9	0.4	0.0	0.1
colon	0.3	16.5	0.1	0.0	0.1	0.0	0.1
endometrium	0.1	0.2	6.4	0.1	0.9	0.1	0.5
hippocampus	0.2	0.0	0.0	15.6	0.1	0.0	0.0
kidney	0.3	0.3	0.9	0.1	19.1	0.5	0.3
liver	0.0	0.0	0.3	0.0	0.3	12.6	0.2
placenta	0.1	0.1	0.5	0.0	0.6	0.1	1.8

Accuracy (average) : 0.9141

```

>
>
> plot(fit$finalModel)
> text(fit$finalModel)
>
>
> set.seed(1991, sample.kind = 'Rounding')
Warning message:
In set.seed(1991, sample.kind = "Rounding") :
  non-uniform 'Rounding' sampler used
> fit <- train(tissue_gene_expression$x, tissue_gene_expression$y, method = 'rf', tuneGrid = data.frame(mt
ry = seq(50, 200, 25)), nodesize = 1)
> plot(fit)
>
>
> imp <- varImp(fit)
> imp
rf variable importance

```

only 20 most important variables shown (out of 500)

```

Overall
GPA33    100.00
BIN1     64.65
GPM6B    62.35
KIF2C    62.15
CLIP3     52.09
COLGALT2 46.48
CFHR4     35.03
SHANK2    34.90
TFR2     33.61
GALNT11   30.70
CEP55     30.49
TCN2     27.96
CAPN3     27.52
CYP4F11   25.74
GTF2IRD1 24.89
KCTD2     24.34
FCN3      22.68
SUSD6     22.24
DOCK4     22.02
RARRES2   21.53
>
>
>
> set.seed(1991, sample.kind = 'Rounding')
Warning message:
In set.seed(1991, sample.kind = "Rounding") :
  non-uniform 'Rounding' sampler used
> fit_rpart <- train(tissue_gene_expression$x, tissue_gene_expression$y, method = 'rpart', tuneGrid = data
.frame(cp = seq(0, 0.1, 0.01)), control = rpart.control(minsplit = 0))
> plot(fit_rpart)
> tree_terms <- as.character(unique(fit_rpart$finalModel$frame$var[!(fit_rpart$finalModel$frame$var == "<l
eaf>")]))
> tree_terms
[1] "GPA33" "CLIP3" "CAPN3" "CFHR4" "CES2" "HRH1" "B3GNT4"
> data_frame(term = rownames(imp$importance), importance = imp$importance$Overall, rank = rank(-importance
)) %>% filter(term %in% tree_terms) %>% arrange(rank)
]# A tibble: 7 x 3
  term      importance rank
<chr>      <dbl> <dbl>
1 GPA33      100      1
2 CLIP3      52.1     5
3 CFHR4      35.0     7
4 CAPN3      27.5    13
5 CES2       20.0    22
6 HRH1        2.35   97
7 B3GNT4      0.136  343
>
>

```