

```

> library(dslabs)
> library(caret)
> data(tissue_gene_expression)
> set.seed(1993, sample.kind = 'Rounding')
Warning message:
In set.seed(1993, sample.kind = "Rounding") :
  non-uniform 'Rounding' sampler used
> class(tissue_gene_expression)
[1] "list"
> 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
>
> ind <- which(tissue_gene_expression$y %in% c("cerebellum", "hippocampus"))
> y <- droplevels(tissue_gene_expression$y[ind])
> length(y)
[1] 69
> unique(y)
[1] cerebellum hippocampus
Levels: cerebellum hippocampus
> x <- tissue_gene_expression$x[ind, ]
> dim(x)
[1] 69 500
> x <- x[, sample(ncol(x), 10)]
> dim(x)
[1] 69 10
>
>
> fit_lda <- train(x, y, method = "lda")
> fit_lda
Linear Discriminant Analysis

69 samples
10 predictors
2 classes: 'cerebellum', 'hippocampus'

```

```

No pre-processing
Resampling: Bootstrapped (25 reps)
Summary of sample sizes: 69, 69, 69, 69, 69, ...
Resampling results:

```

```

Accuracy  Kappa
0.8727591 0.7386465

```

```

>
>
> fit_lda$finalModel
Call:
lda(x, grouping = y)

```

```

Prior probabilities of groups:
cerebellum hippocampus
0.5507246 0.4492754

```

```

Group means:
      SAPCD1  HEMK1  PLCB1  SPI1  RAB1B  MSH4  IL18R1
cerebellum 6.986817 7.438612 5.040267 6.048843 10.26148 5.325845 5.421491
hippocampus 7.027106 7.350232 5.023790 6.490261 10.24046 5.247354 5.539534
      OAZ2  FOXE3  C21orf62
cerebellum 10.00821 6.022749 5.911605
hippocampus 9.65903 6.062418 5.721539

```

Coefficients of linear discriminants:

LD1

SAPCD1 -1.0380541
HEMK1 -1.3300757
PLCB1 0.2590667
SPI1 3.1261507
RAB1B 0.1836489
MSH4 -2.8846563
IL18R1 3.6575231
OAZ2 -1.1365484
FOX3 -0.8198930
C21orf62 -3.4085635

> fit_lda\$finalModel\$means

| | SAPCD1 | HEMK1 | PLCB1 | SPI1 | RAB1B | MSH4 | IL18R1 |
|-------------|----------|----------|----------|----------|----------|----------|----------|
| cerebellum | 6.986817 | 7.438612 | 5.040267 | 6.048843 | 10.26148 | 5.325845 | 5.421491 |
| hippocampus | 7.027106 | 7.350232 | 5.023790 | 6.490261 | 10.24046 | 5.247354 | 5.539534 |

| | OAZ2 | FOX3 | C21orf62 |
|-------------|----------|----------|----------|
| cerebellum | 10.00821 | 6.022749 | 5.911605 |
| hippocampus | 9.65903 | 6.062418 | 5.721539 |

> t(fit_lda\$finalModel\$means)

| | cerebellum | hippocampus |
|----------|------------|-------------|
| SAPCD1 | 6.986817 | 7.027106 |
| HEMK1 | 7.438612 | 7.350232 |
| PLCB1 | 5.040267 | 5.023790 |
| SPI1 | 6.048843 | 6.490261 |
| RAB1B | 10.261483 | 10.240458 |
| MSH4 | 5.325845 | 5.247354 |
| IL18R1 | 5.421491 | 5.539534 |
| OAZ2 | 10.008207 | 9.659030 |
| FOX3 | 6.022749 | 6.062418 |
| C21orf62 | 5.911605 | 5.721539 |

> t(fit_lda\$finalModel\$means) %>% data.frame() %>% mutate(predictor_name = rownames(.))

| | cerebellum | hippocampus | predictor_name |
|----|------------|-------------|----------------|
| 1 | 6.986817 | 7.027106 | SAPCD1 |
| 2 | 7.438612 | 7.350232 | HEMK1 |
| 3 | 5.040267 | 5.023790 | PLCB1 |
| 4 | 6.048843 | 6.490261 | SPI1 |
| 5 | 10.261483 | 10.240458 | RAB1B |
| 6 | 5.325845 | 5.247354 | MSH4 |
| 7 | 5.421491 | 5.539534 | IL18R1 |
| 8 | 10.008207 | 9.659030 | OAZ2 |
| 9 | 6.022749 | 6.062418 | FOX3 |
| 10 | 5.911605 | 5.721539 | C21orf62 |

> ds_theme_set()

> t(fit_lda\$finalModel\$means) %>% data.frame() %>%

+ mutate(predictor_name = rownames(.)) %>%

+ ggplot(aes(cerebellum, hippocampus, label = predictor_name)) +

+ geom_point() +

+ geom_text() +

+ geom_abline()

>