Outputs inspection CIFAR10

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.5

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

```
library("ggpubr")
```

```
## Warning: package 'ggpubr' was built under R version 4.0.5
```

```
library(LDATS)
```

```
## Warning: package 'LDATS' was built under R version 4.0.5
```

```
library(stringr)
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 4.0.3

##
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':
##
##     smiths
```

```r
library(reticulate)
```

```
## Warning: package 'reticulate' was built under R version 4.0.5
```

```r
library(abind)
```

```
## Warning: package 'abind' was built under R version 4.0.3
```

```r
library(ggVennDiagram)
```

```
## Warning: package 'ggVennDiagram' was built under R version 4.0.5
```

```r
np <- import("numpy")

source("utils.R")
```

```
## Warning: package 'hash' was built under R version 4.0.5

## hash-2.2.6.1 provided by Decision Patterns

## Warning: package 'berryFunctions' was built under R version 4.0.5

##
## Attaching package: 'berryFunctions'

## The following object is masked from 'package:ggVennDiagram':
##
##     circle

## The following object is masked from 'package:dplyr':
##
##     between

## Warning: package 'purrr' was built under R version 4.0.3
```

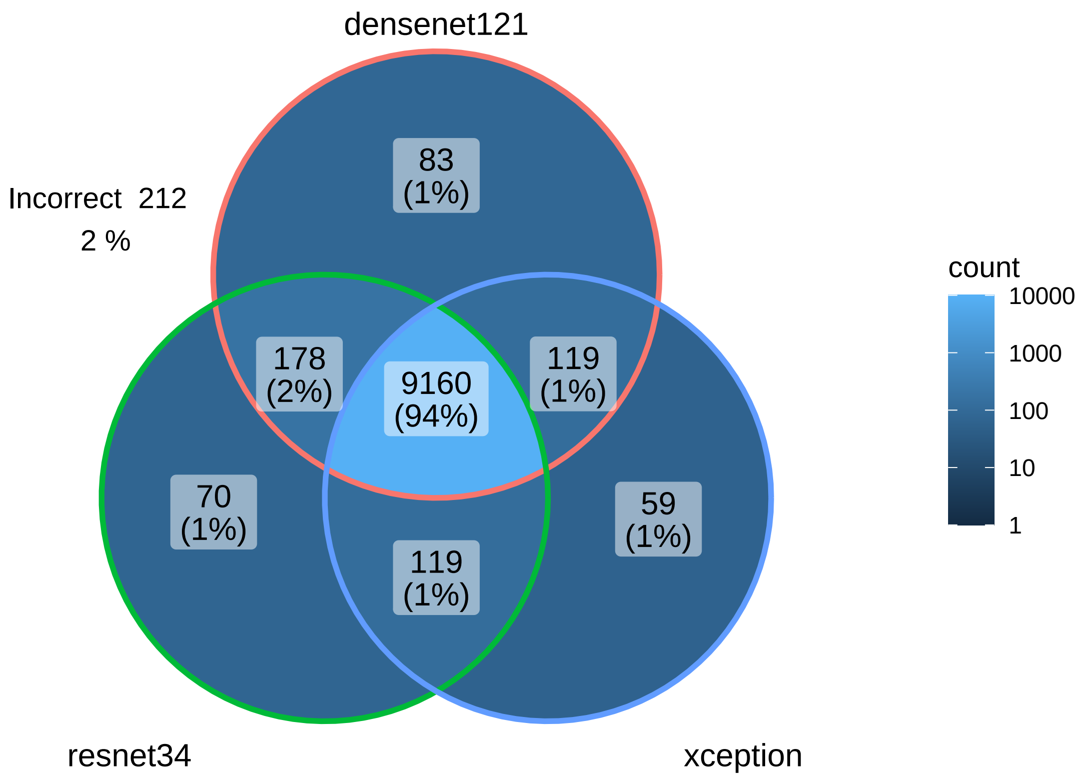
Visualization on CIFAR10. We are using data of three neural networks trained on reduced CIFAR10 training set. These networks were trained in 30 replications. In each replication, 500 samples from the training set were randomly extracted and formed validation set. In each replication, we trained two ensembles on the outputs of neural networks. First one was trained on randomly chosen subset, of size 500, of nn training set, second on the extracted validation set. In this visualization, we are trying to inspect the outputs deeper, mainly to make sense of strange behavior of nll metric for ensemble outputs.

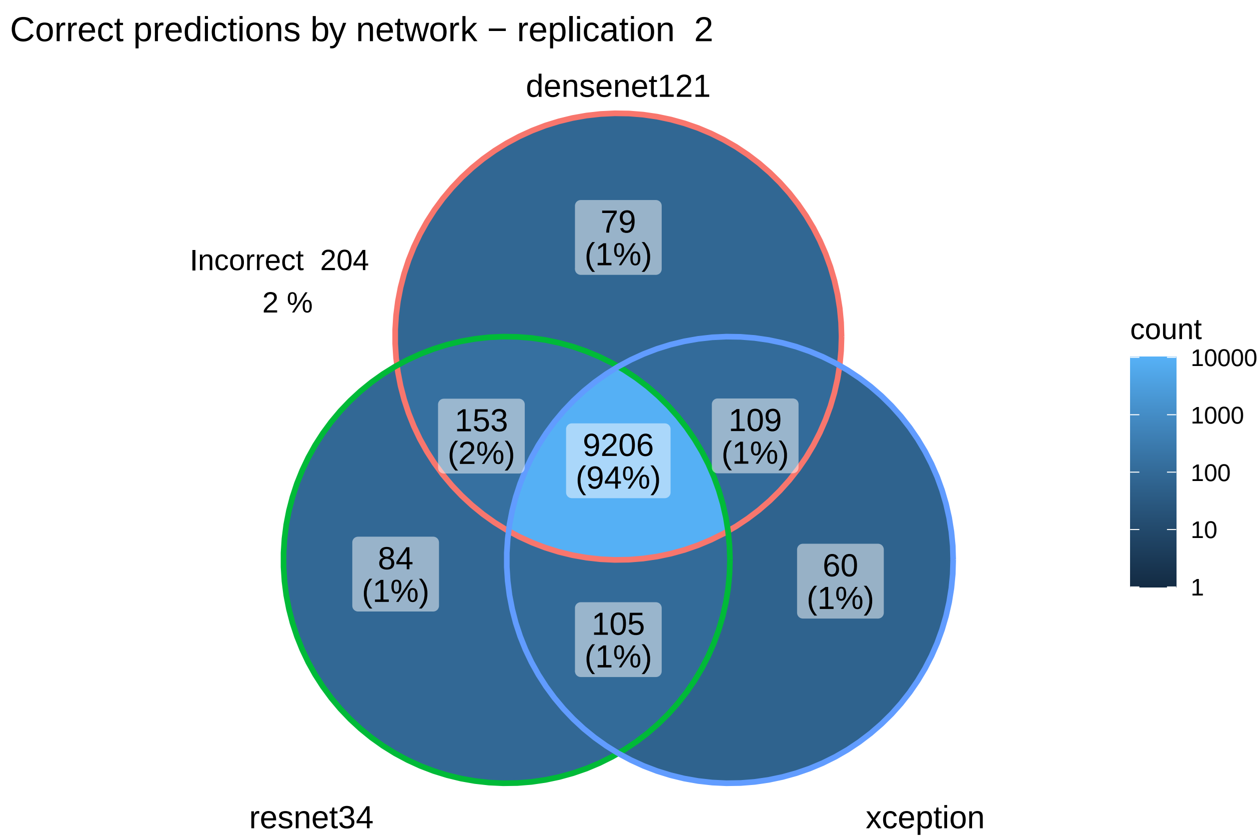
```r
base_dir <- "../data/data_train_val_c10"
repls <- 0:29
classes <- 10

nets_outputs <- load_network_outputs(base_dir, repls)
ens_outputs <- load_ensemble_outputs(base_dir, repls)
net_results <- read.csv(file.path(base_dir, "net_accuracies.csv"))
ens_results <- read.csv(file.path(base_dir, "ensemble_accuracies.csv"))
```

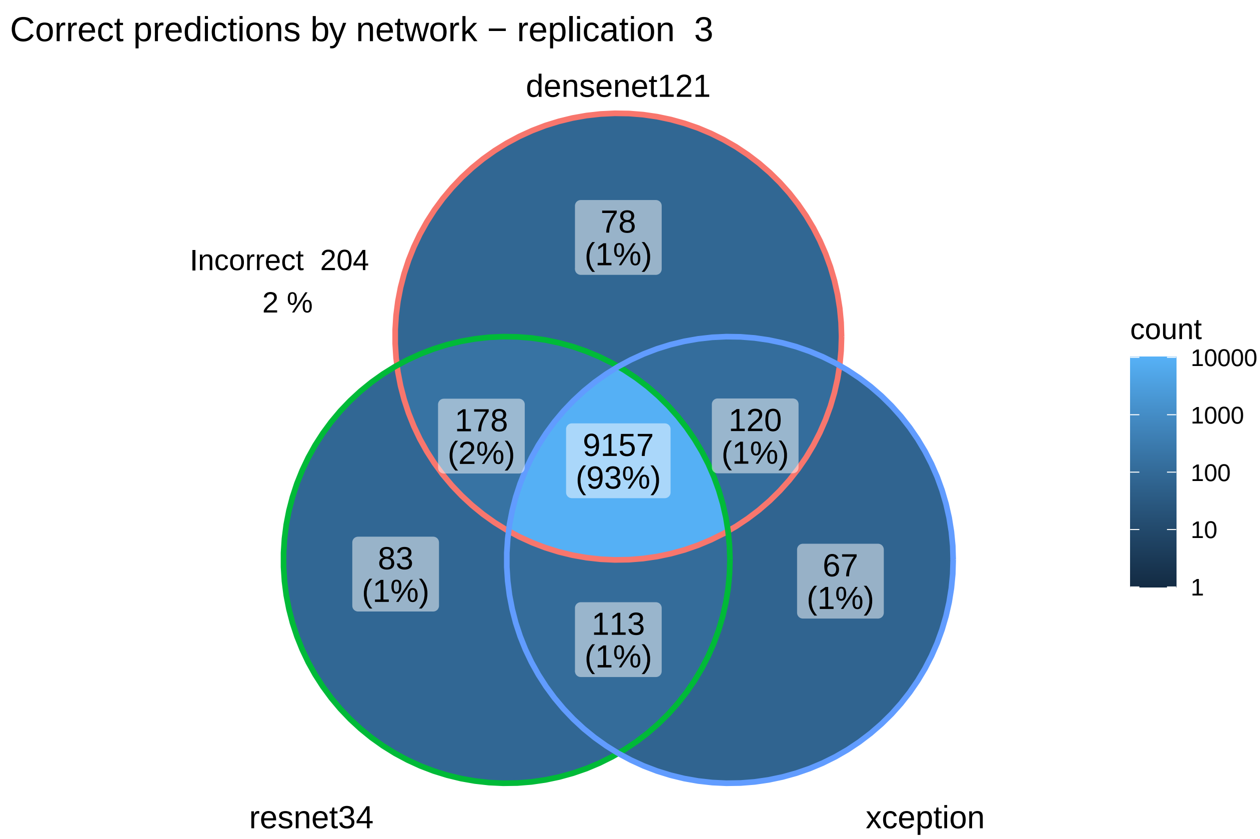
```r
sort_ind <- function(lst)
{
  return(sort(lst, index.return=TRUE, decreasing=TRUE)$ix)
}
nets_test_top_indices <- apply(X=nets_outputs$test_outputs, MARGIN=c(1, 2, 3), FUN=sort_ind)[1, , , ]
r_n <- length(repls)
samples_n <- dim(nets_outputs$test_labels)[2]
nets_n <- length(nets_outputs$networks)
test_labs <- nets_outputs$test_labels + 1
dim(test_labs) <- c(r_n, 1, samples_n)
test_labs <- aperm(abind(array(rep(aperm(test_labs, perm=c(2, 1, 3)), nets_n), c(r_n, samples_n, nets_n
nets_test_cor_preds <- test_labs == nets_test_top_indices
```

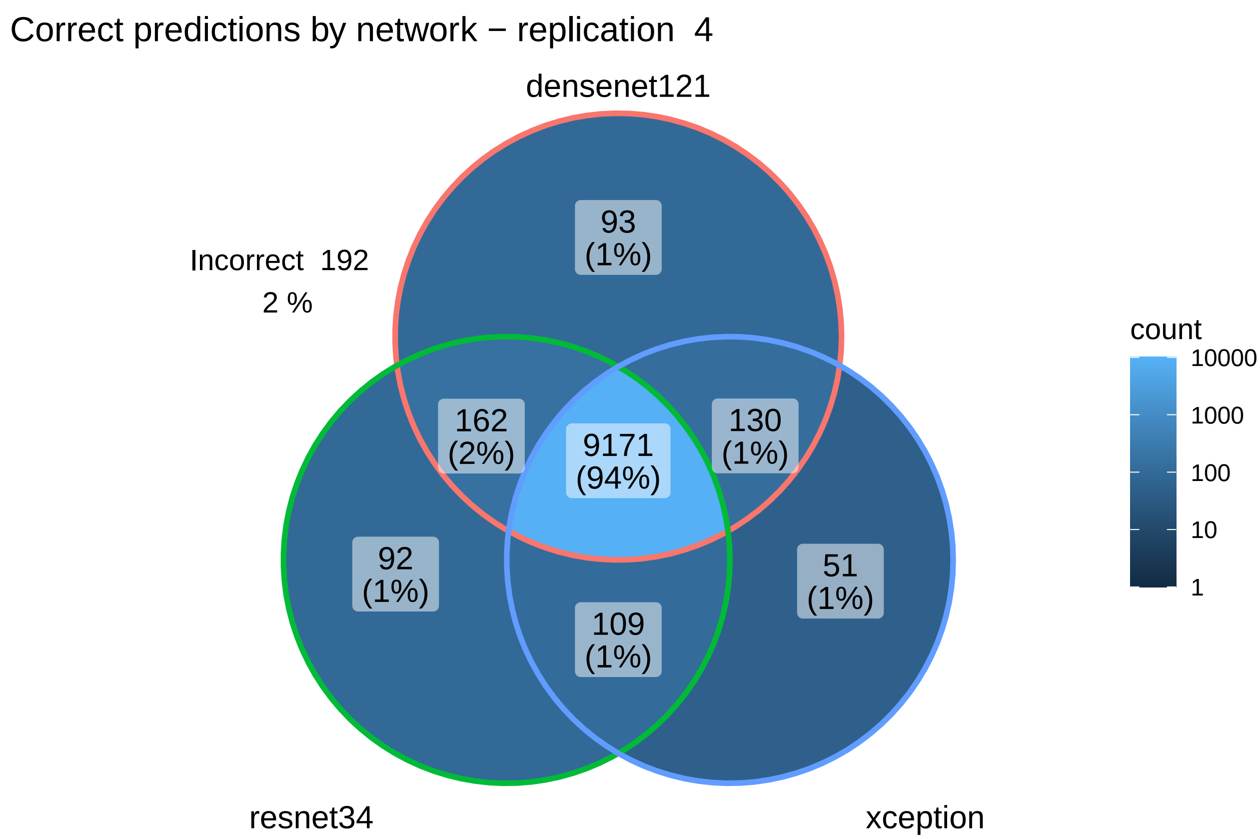
```r
for (ri in 1:r_n)
{
  nets_cor_list <- list()
  incor <- 1:samples_n
  for (ni in 1:nets_n)
  {
    cor_list <- which(nets_test_cor_preds[ri, ni, ])
    nets_cor_list[[nets_outputs$networks[ni]]] = cor_list
    incor <- setdiff(incor, cor_list)
  }
  incor_n <- length(incor)
  venn_diag <- ggVennDiagram(nets_cor_list) + scale_fill_gradient(trans="log10", name="count", limits=c
    annotate(geom="text", x=-4, y=5, label=paste("Incorrect ", incor_n, "\n", round(incor_n / samples_n
    ggtitle(paste("Correct predictions by network - replication ", ri)) +
    scale_x_continuous(limits=c(-8, 10))
  print(venn_diag)
}
```

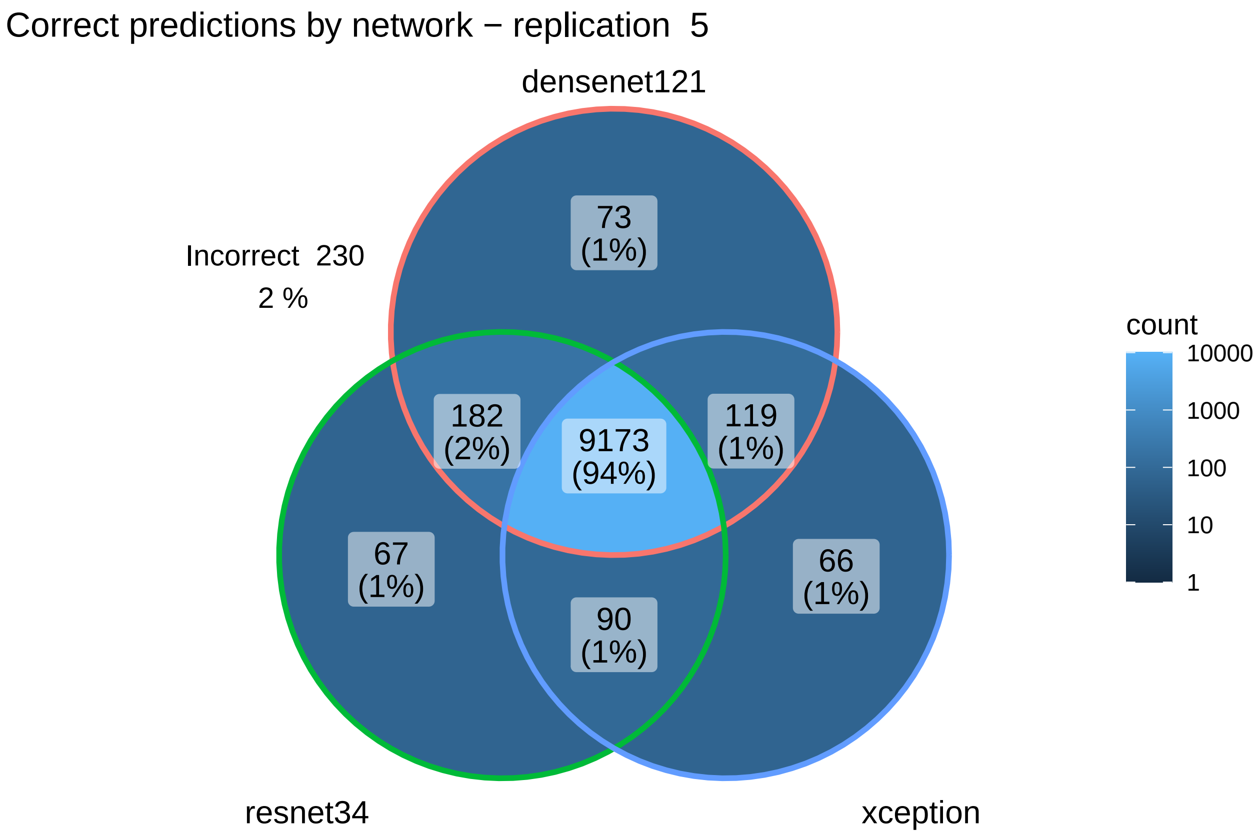
Correct predictions by network − replication 1

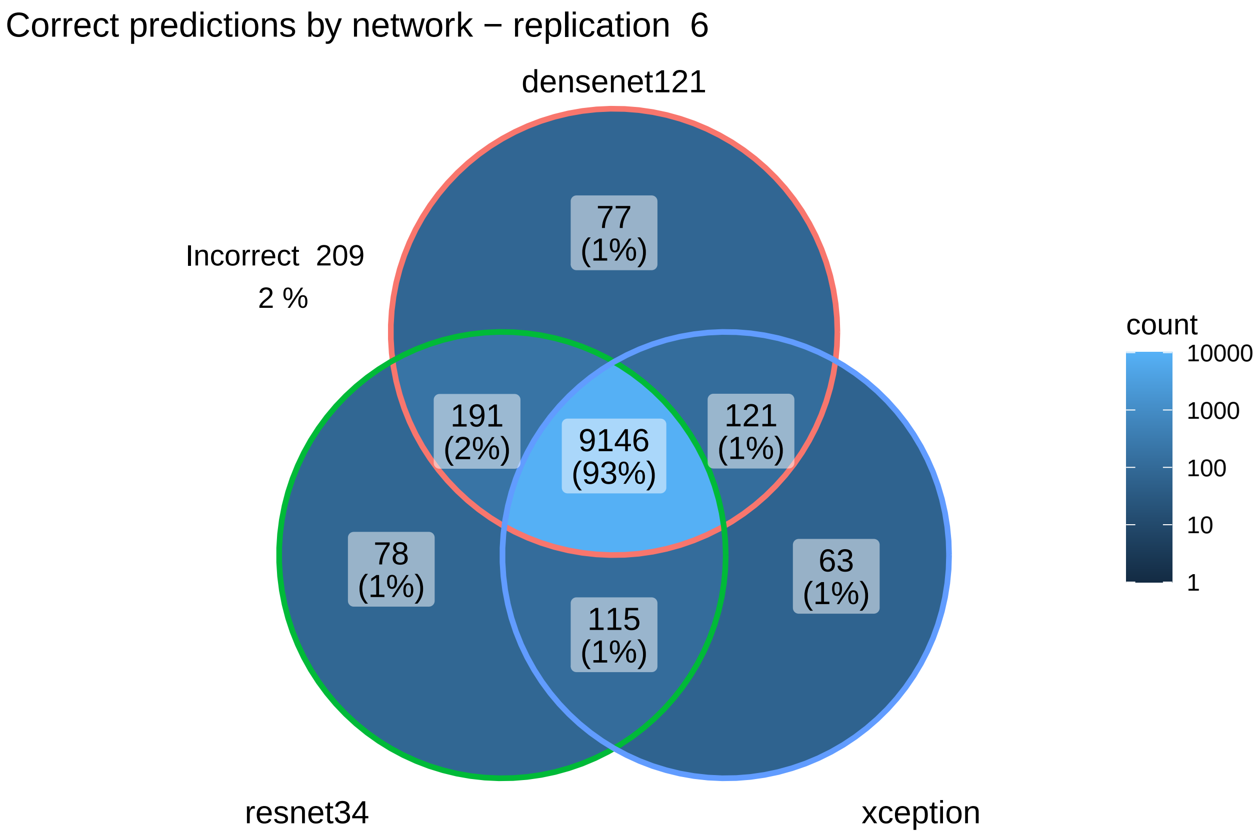


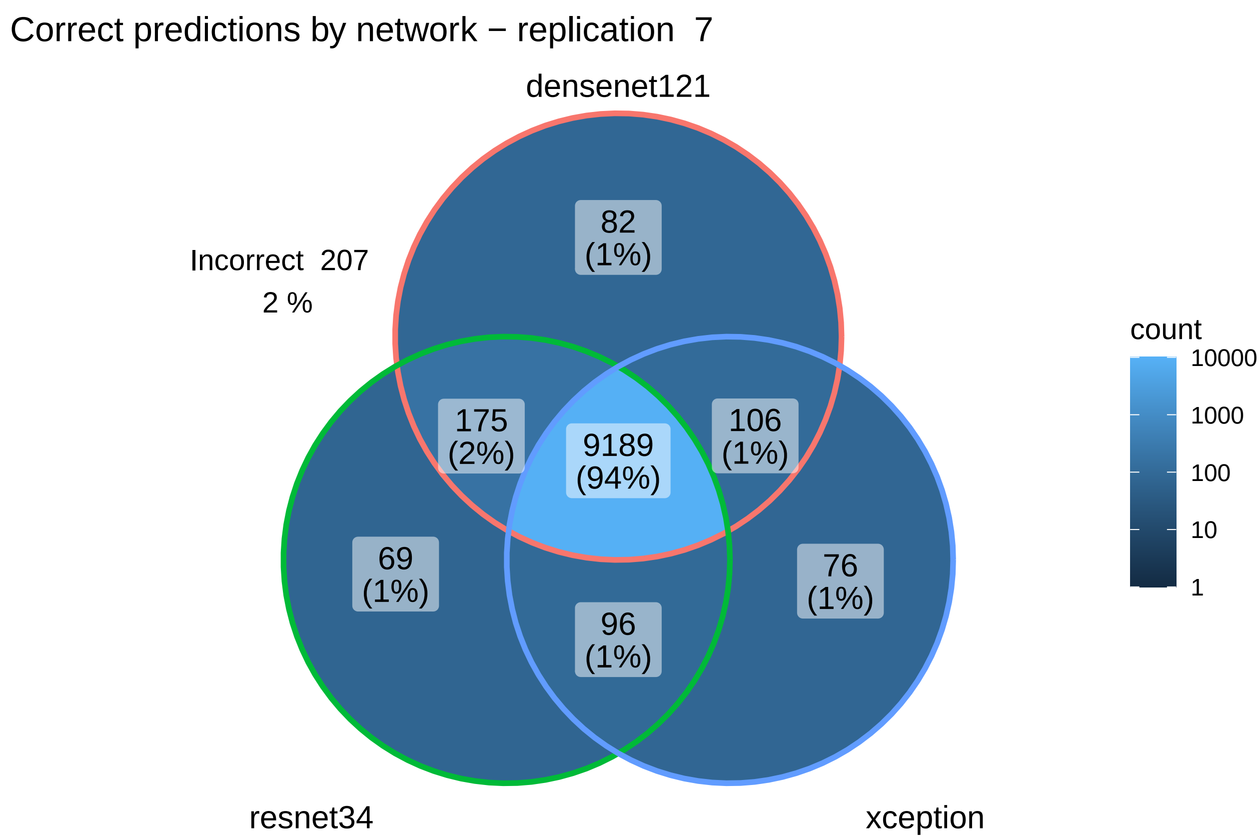

Correct predictions by network − replication 2
densenet121
79
(1%)
Incorrect 204
2 %
153
(2%)
9206
(94%)
109
(1%)
84
(1%)
105
(1%)
60
(1%)
resnet34
xception
count
10000
1000
100
10
1

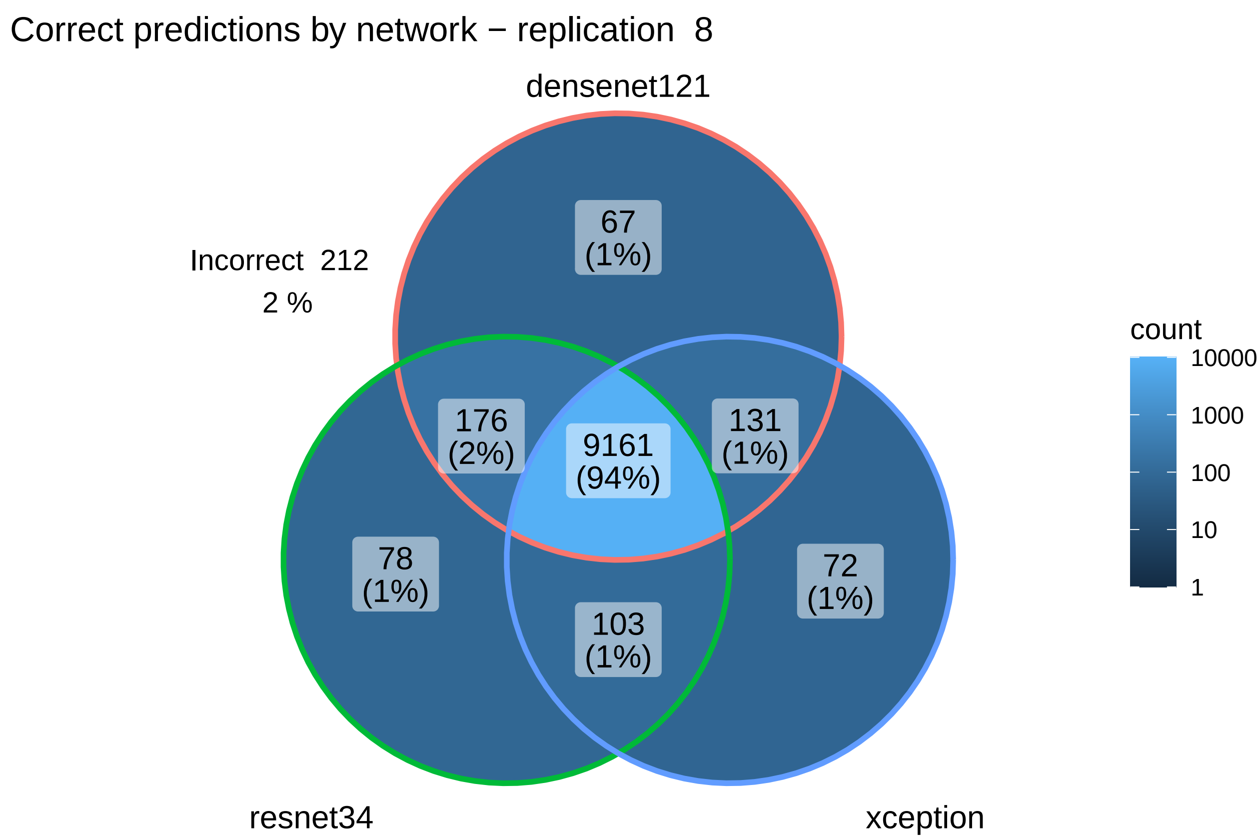

Correct predictions by network − replication 3
densenet121
Incorrect 204
2 %
78
(1%)
178
(2%)
9157
(93%)
120
(1%)
83
(1%)
67
(1%)
113
(1%)
resnet34
xception
count
10000
1000
100
10
1

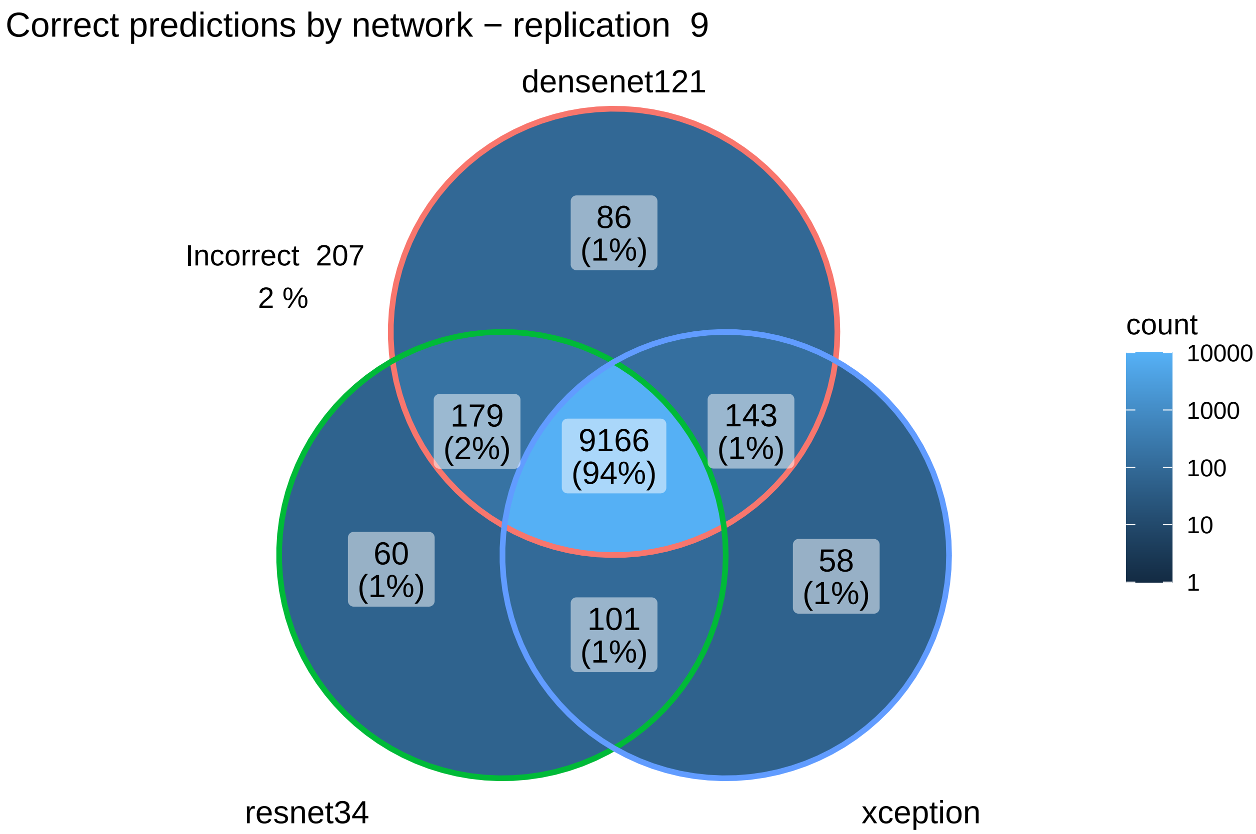

Correct predictions by network − replication 4
densenet121
Incorrect 192
2 %
93
(1%)
162
(2%)
9171
(94%)
130
(1%)
92
(1%)
109
(1%)
51
(1%)
resnet34
xception
count
10000
1000
100
10
1

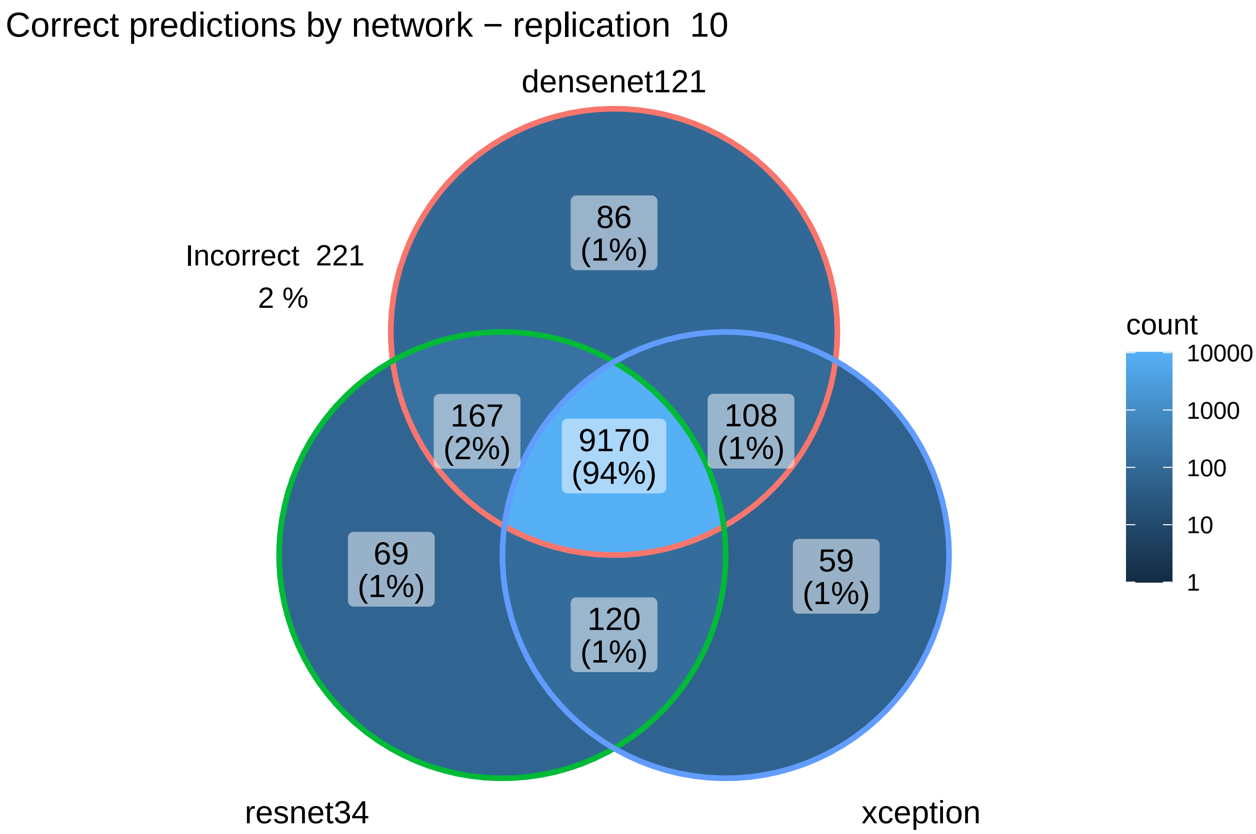

Correct predictions by network − replication 5
densenet121
Incorrect 230
2 %
73
(1%)
182
(2%)
9173
(94%)
119
(1%)
67
(1%)
90
(1%)
66
(1%)
resnet34
xception
count
10000
1000
100
10
1

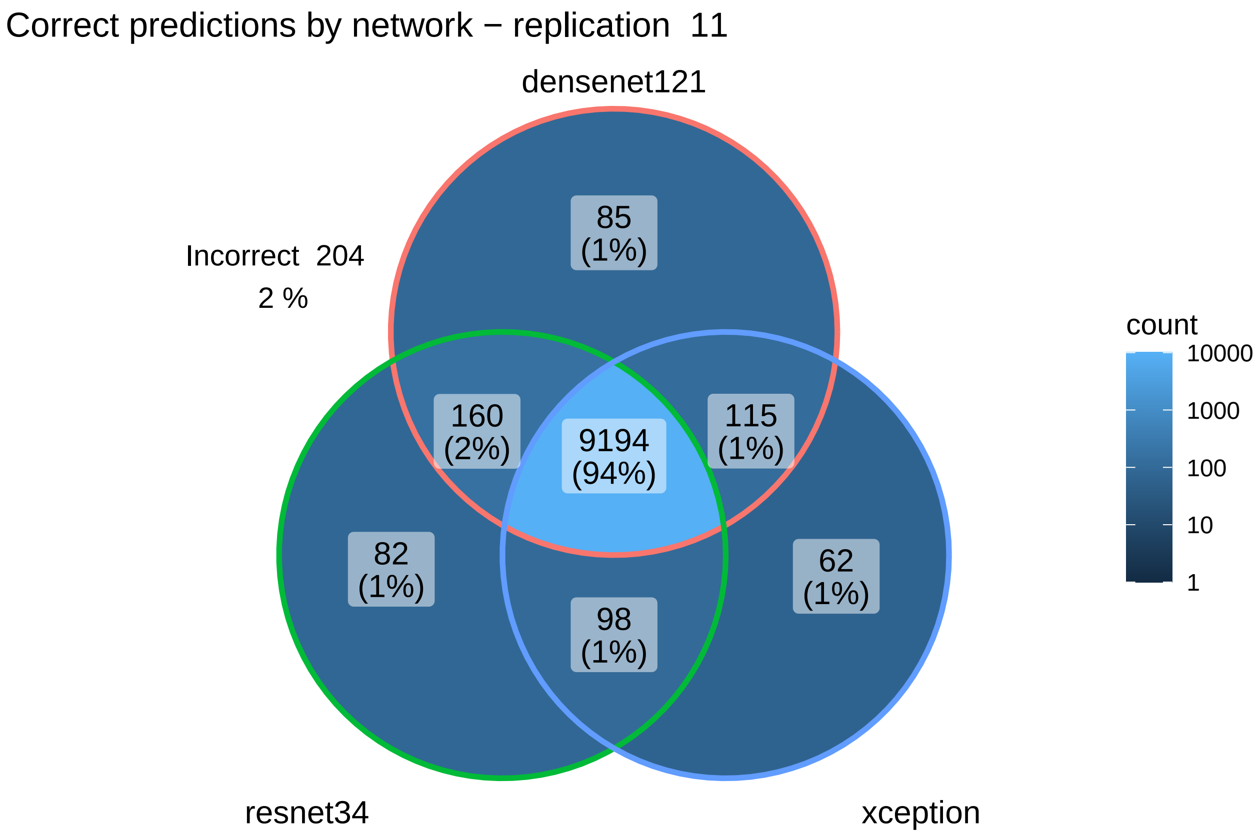

Correct predictions by network − replication 6
densenet121
77
(1%)
Incorrect 209
2 %
191
(2%)
9146
(93%)
121
(1%)
78
(1%)
115
(1%)
63
(1%)
resnet34
xception
count
10000
1000
100
10
1

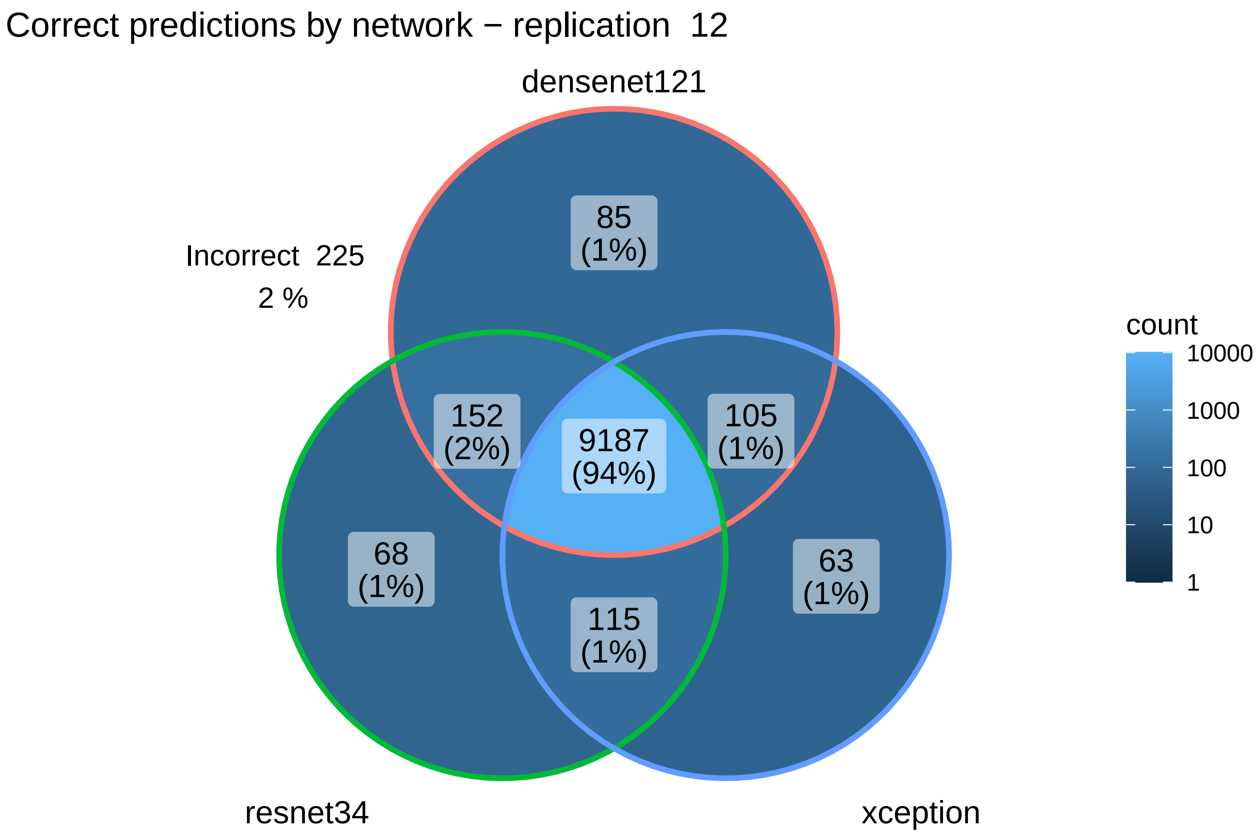

Correct predictions by network − replication 7
densenet121
Incorrect 207
2 %
82
(1%)
175
(2%)
9189
(94%)
106
(1%)
69
(1%)
96
(1%)
76
(1%)
resnet34
xception
count
10000
1000
100
10
1

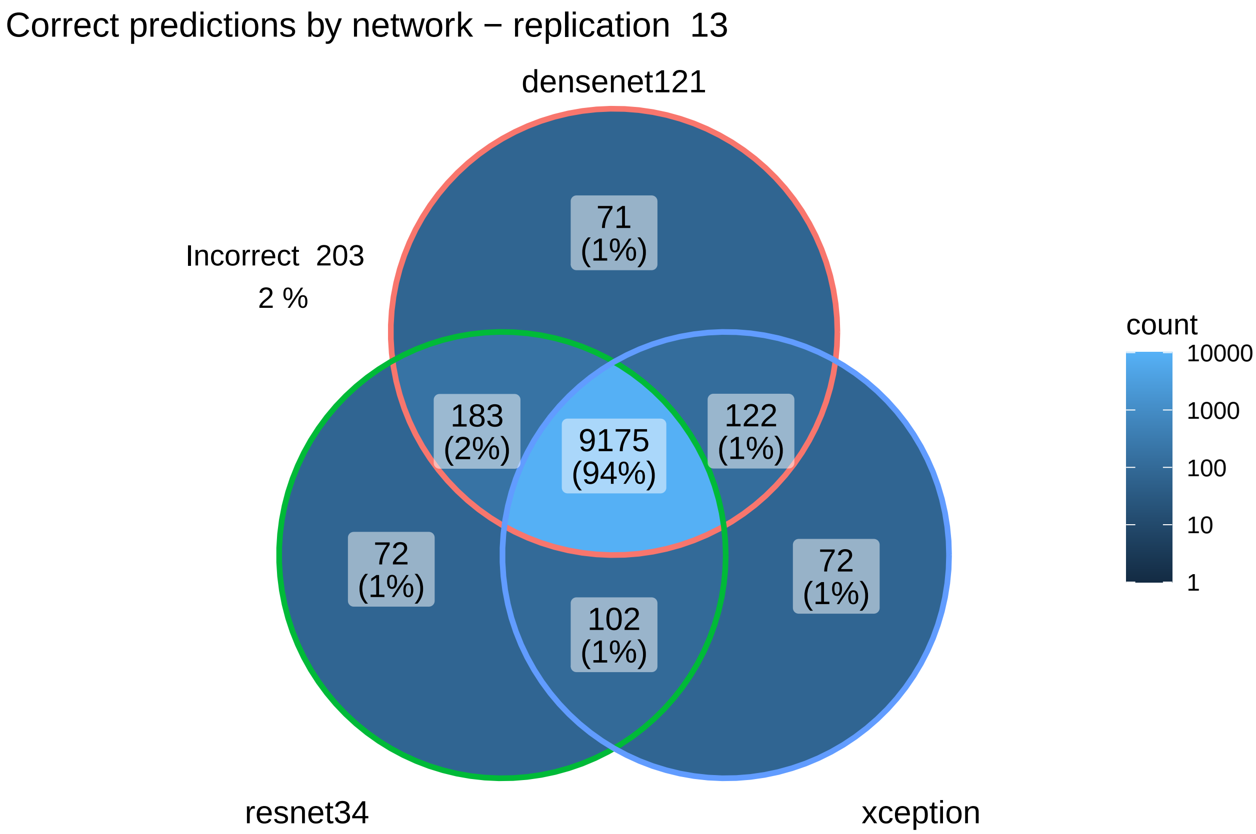

Correct predictions by network − replication 8
densenet121
Incorrect 212
2 %
67
(1%)
176
(2%)
9161
(94%)
131
(1%)
78
(1%)
103
(1%)
72
(1%)
resnet34
xception
count
10000
1000
100
10
1

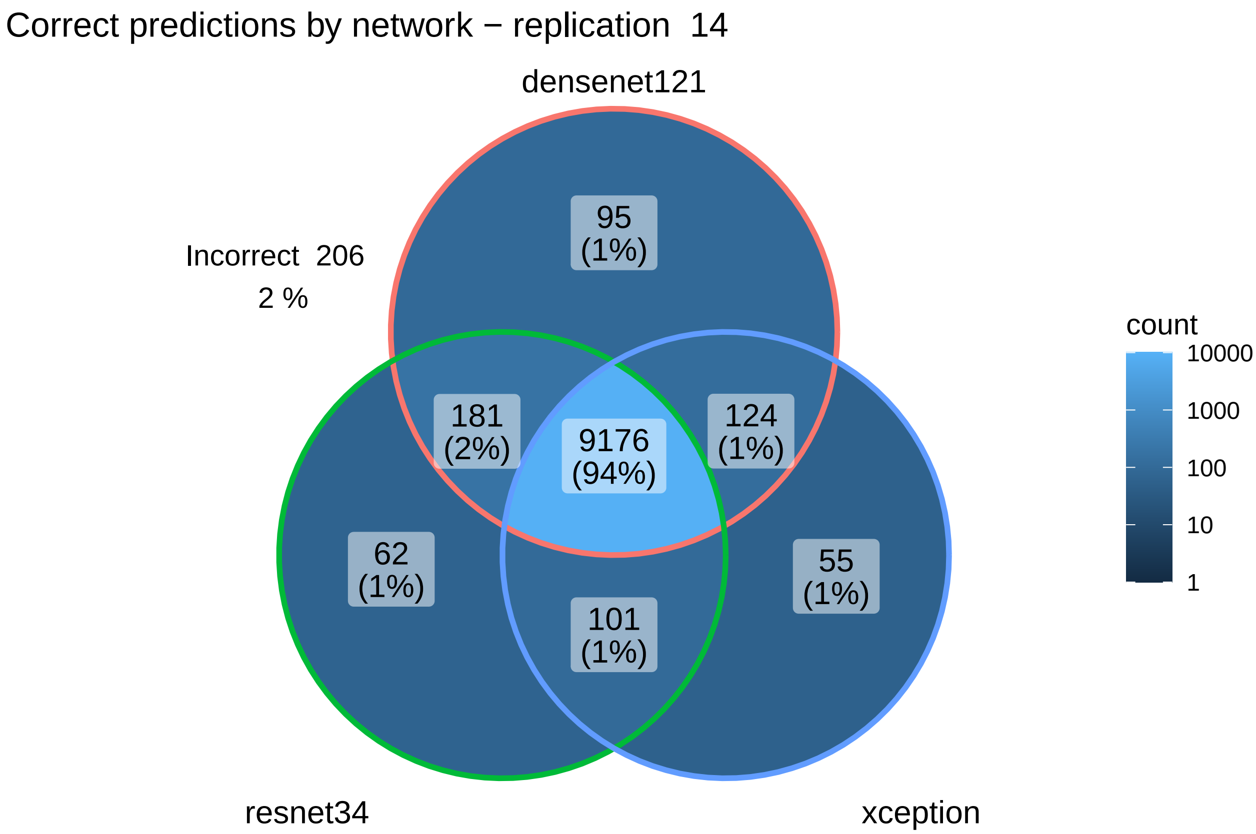

Correct predictions by network − replication 9
densenet121
86
(1%)
Incorrect 207
2 %
179
(2%)
9166
(94%)
143
(1%)
60
(1%)
101
(1%)
58
(1%)
resnet34
xception
count
10000
1000
100
10
1

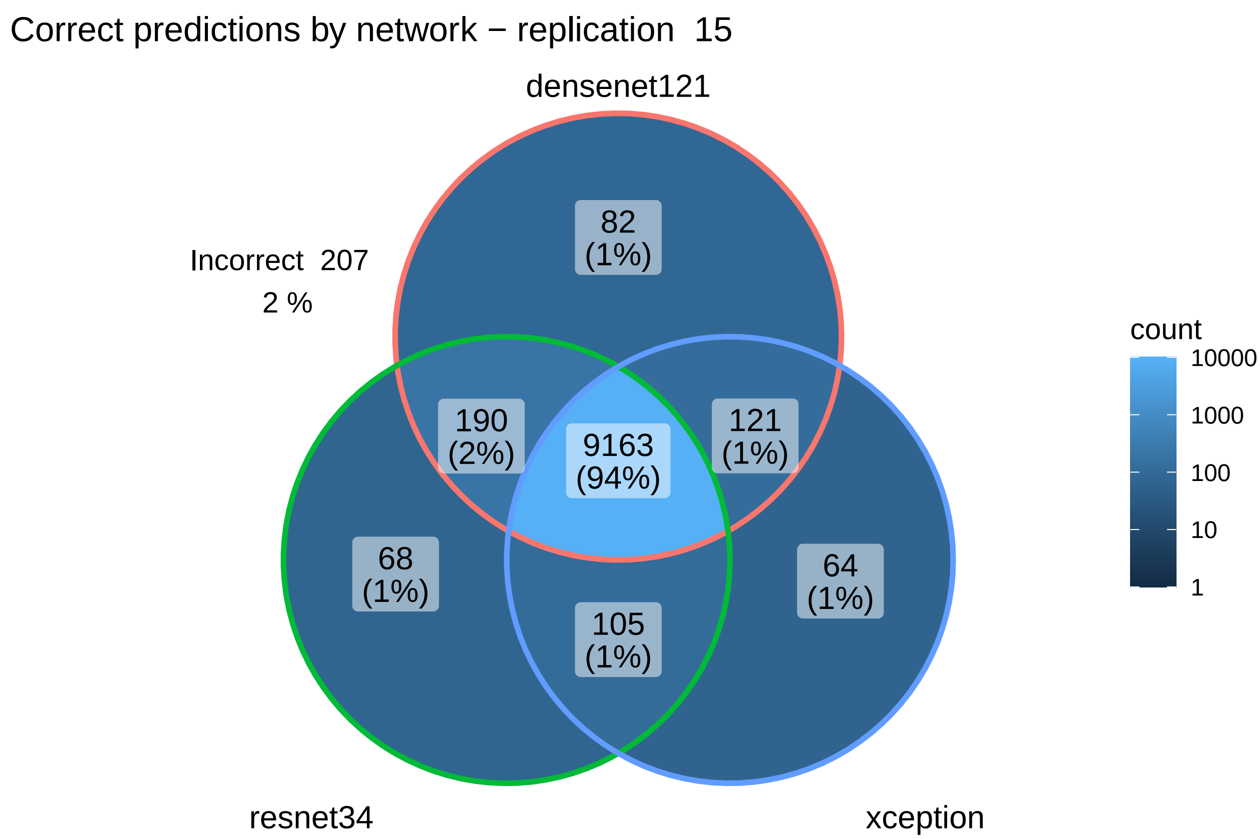

Correct predictions by network − replication 10
densenet121
Incorrect 221
2 %
86
(1%)
167
(2%)
9170
(94%)
108
(1%)
69
(1%)
120
(1%)
59
(1%)
resnet34
xception
count
10000
1000
100
10
1

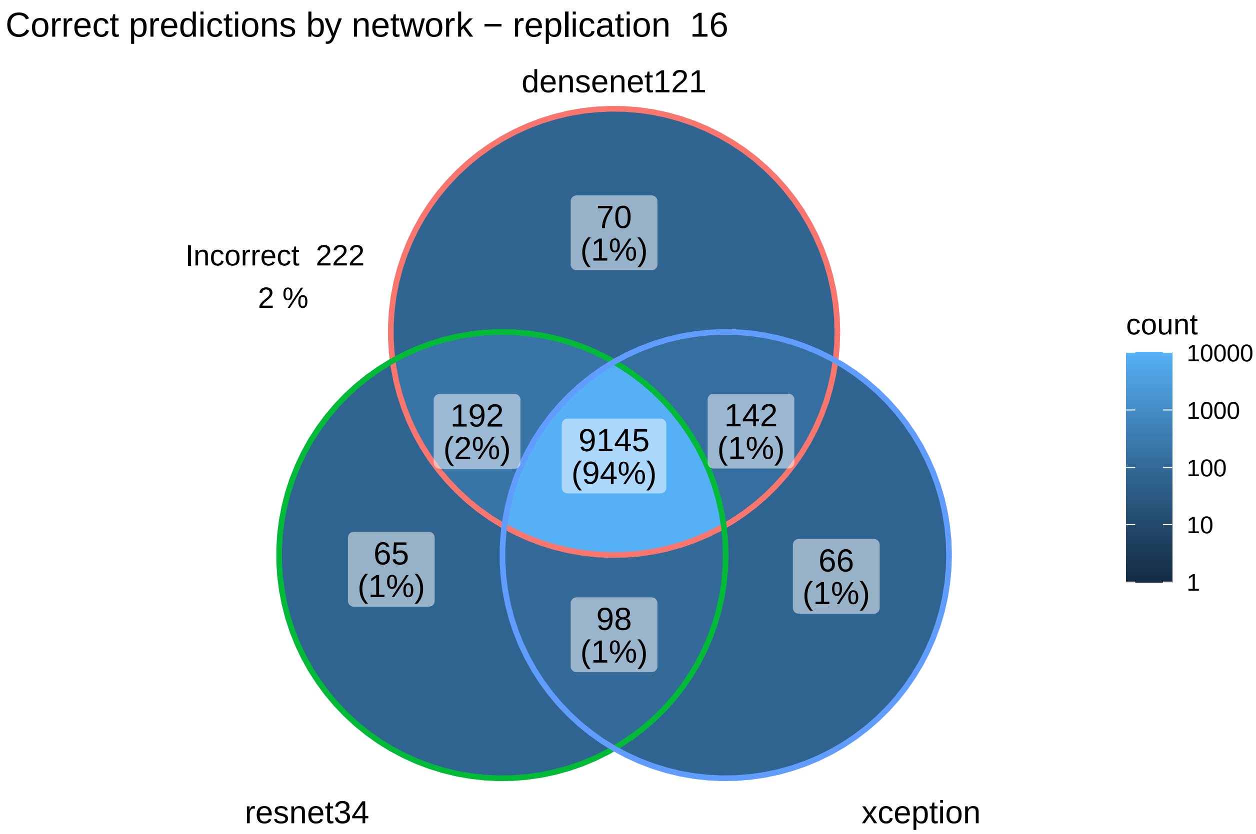

Correct predictions by network − replication 11
densenet121
85
(1%)
Incorrect 204
2 %
160
(2%)
9194
(94%)
115
(1%)
82
(1%)
98
(1%)
62
(1%)
resnet34
xception
count
10000
1000
100
10
1

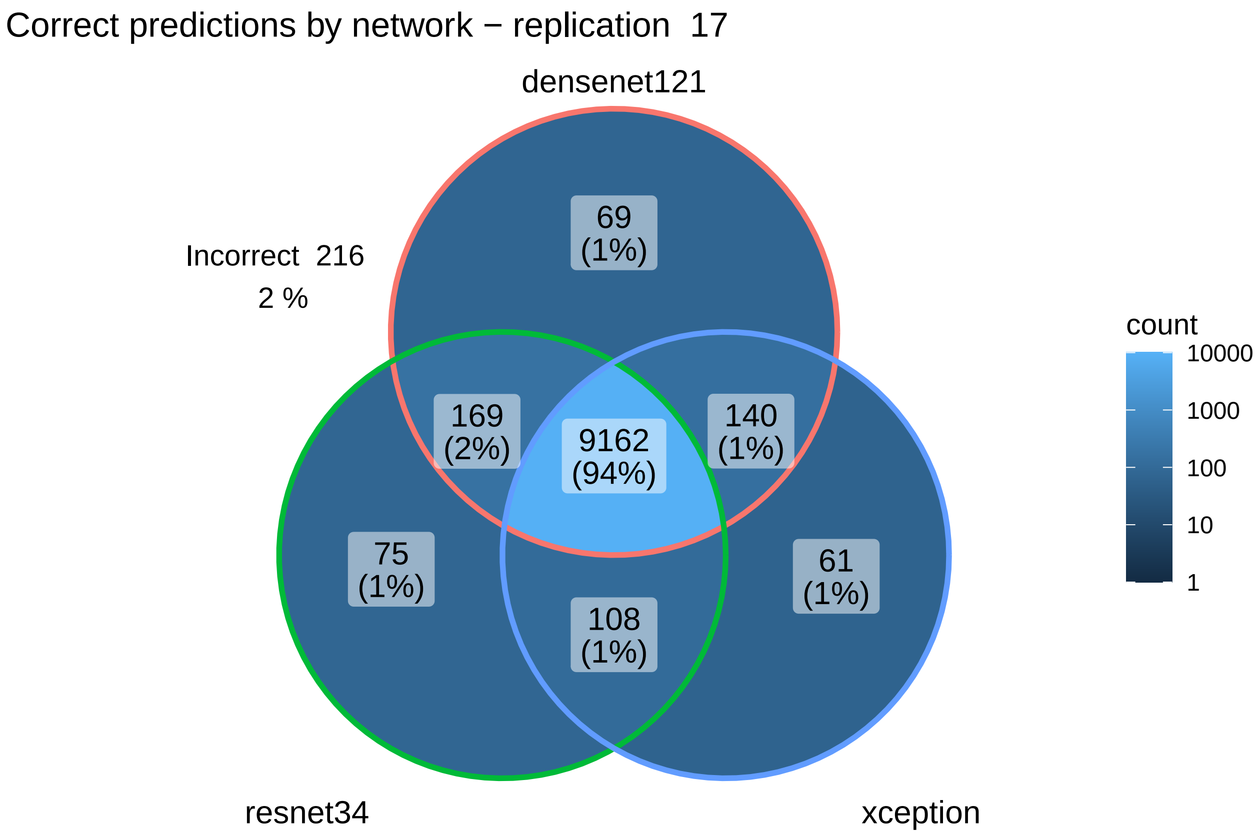

Correct predictions by network − replication 12
densenet121
Incorrect 225
2 %
85
(1%)
152
(2%)
9187
(94%)
105
(1%)
68
(1%)
63
(1%)
115
(1%)
resnet34
xception
count
10000
1000
100
10
1

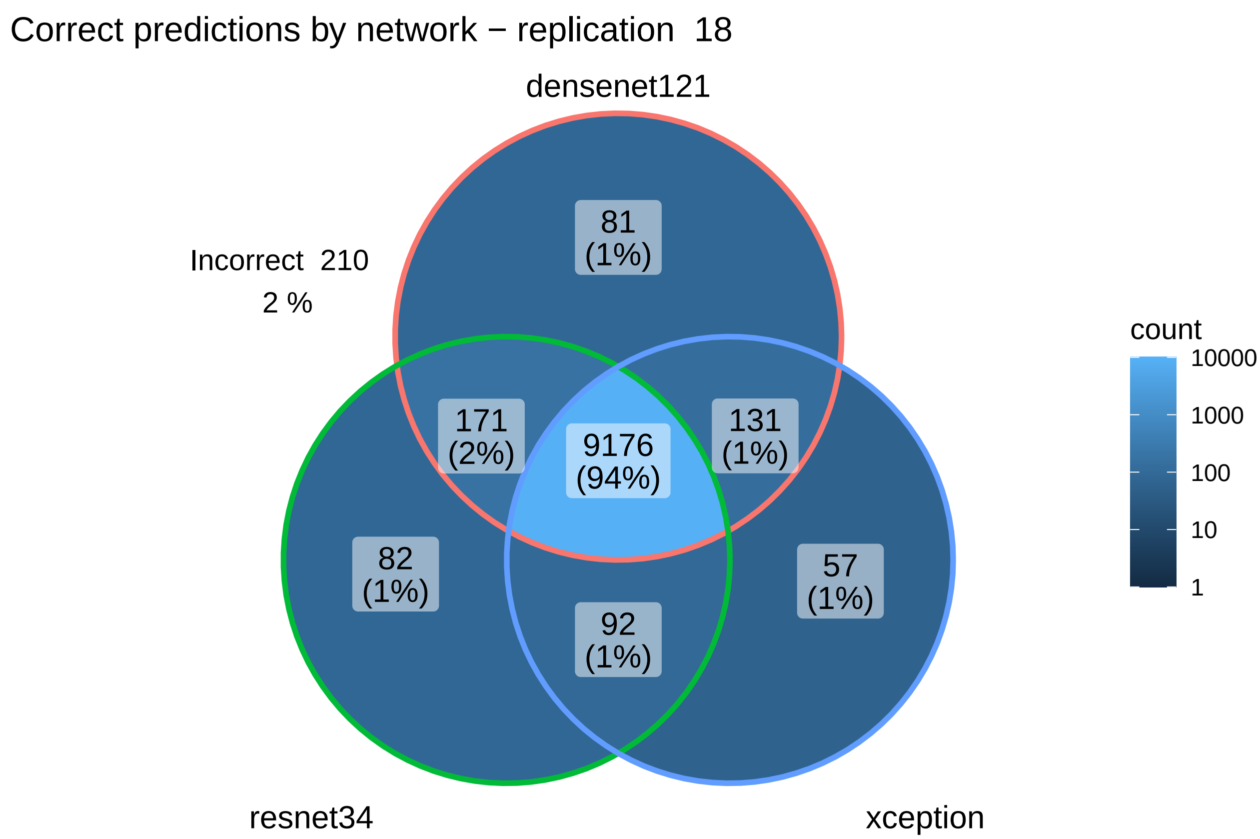

Correct predictions by network − replication 13
densenet121
Incorrect 203
2 %
71
(1%)
183
(2%)
9175
(94%)
122
(1%)
72
(1%)
102
(1%)
72
(1%)
resnet34
xception
count
10000
1000
100
10
1

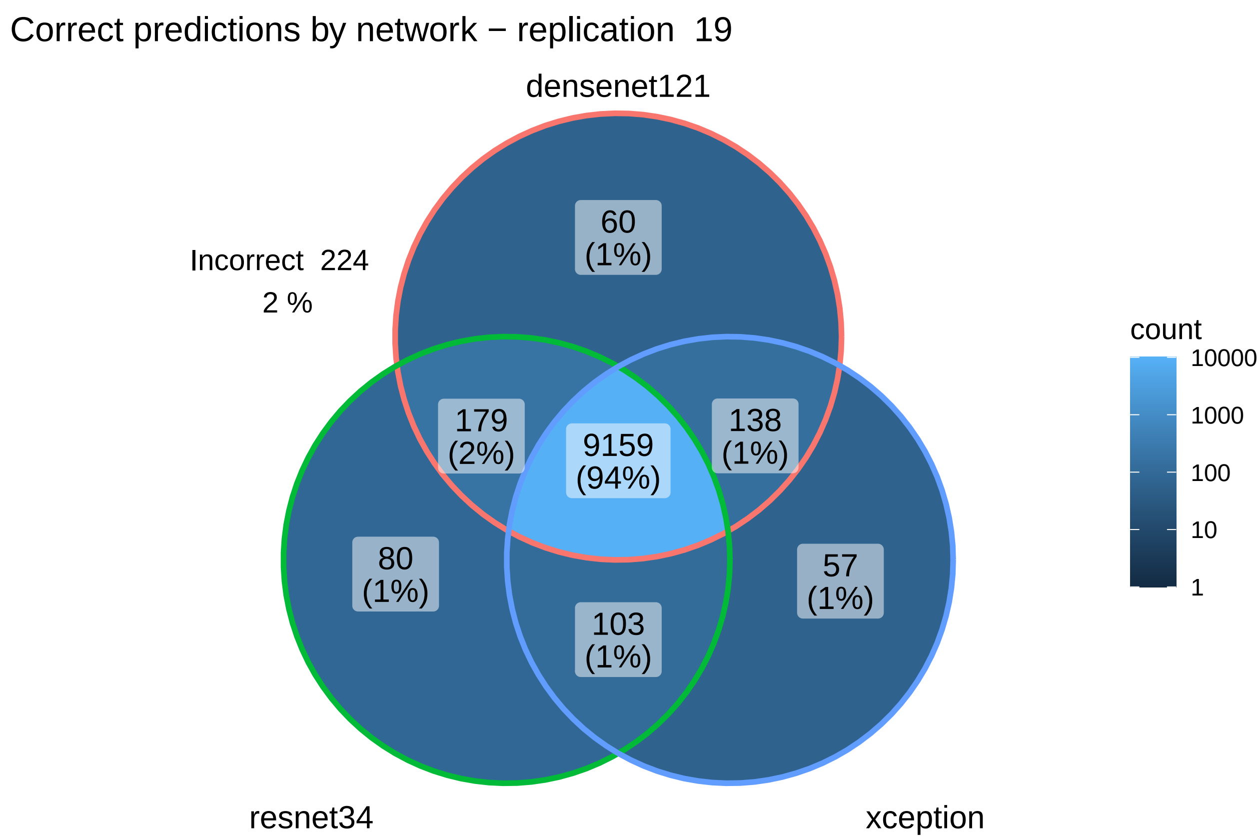

Correct predictions by network − replication 14
densenet121
Incorrect 206
2 %
95
(1%)
181
(2%)
9176
(94%)
124
(1%)
62
(1%)
101
(1%)
55
(1%)
resnet34
xception
count
10000
1000
100
10
1

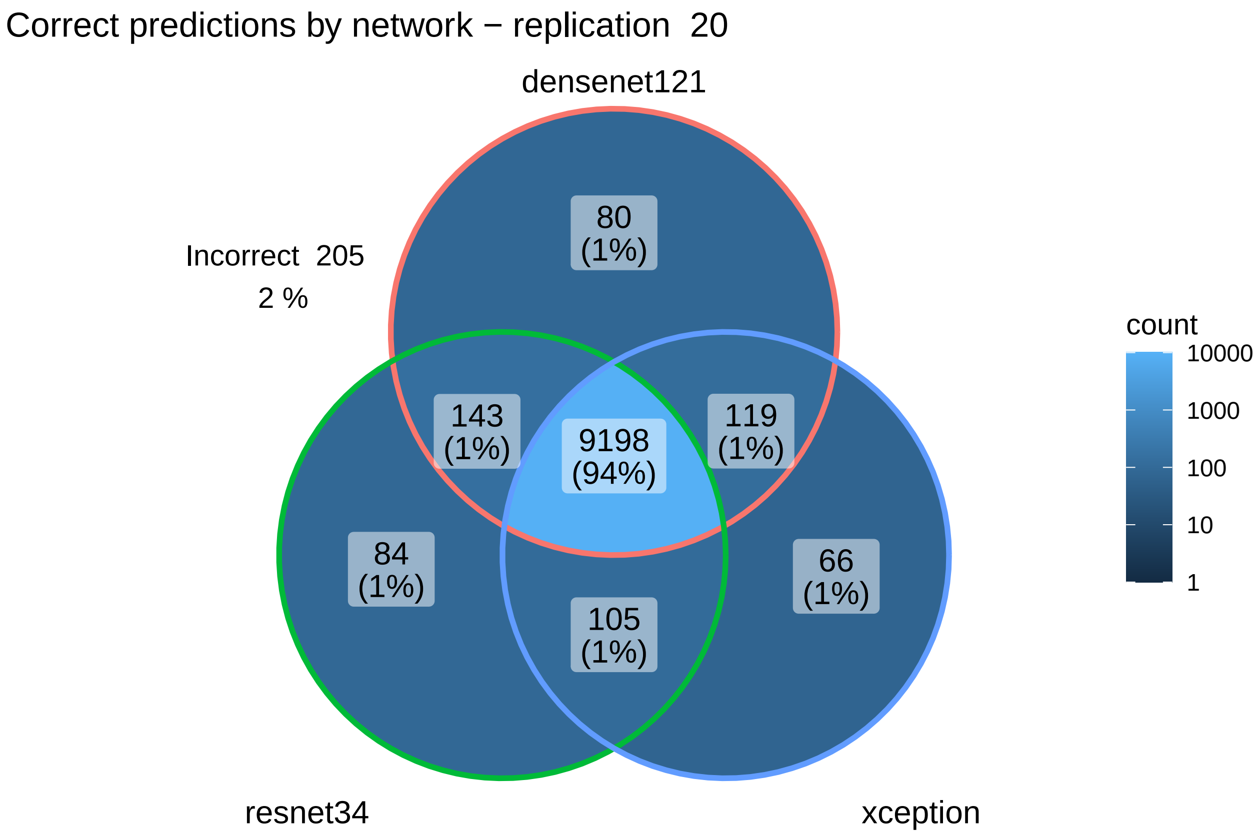

Correct predictions by network − replication 15
densenet121
Incorrect 207
2 %
82
(1%)
190
(2%)
9163
(94%)
121
(1%)
68
(1%)
105
(1%)
64
(1%)
resnet34
xception
count
10000
1000
100
10
1

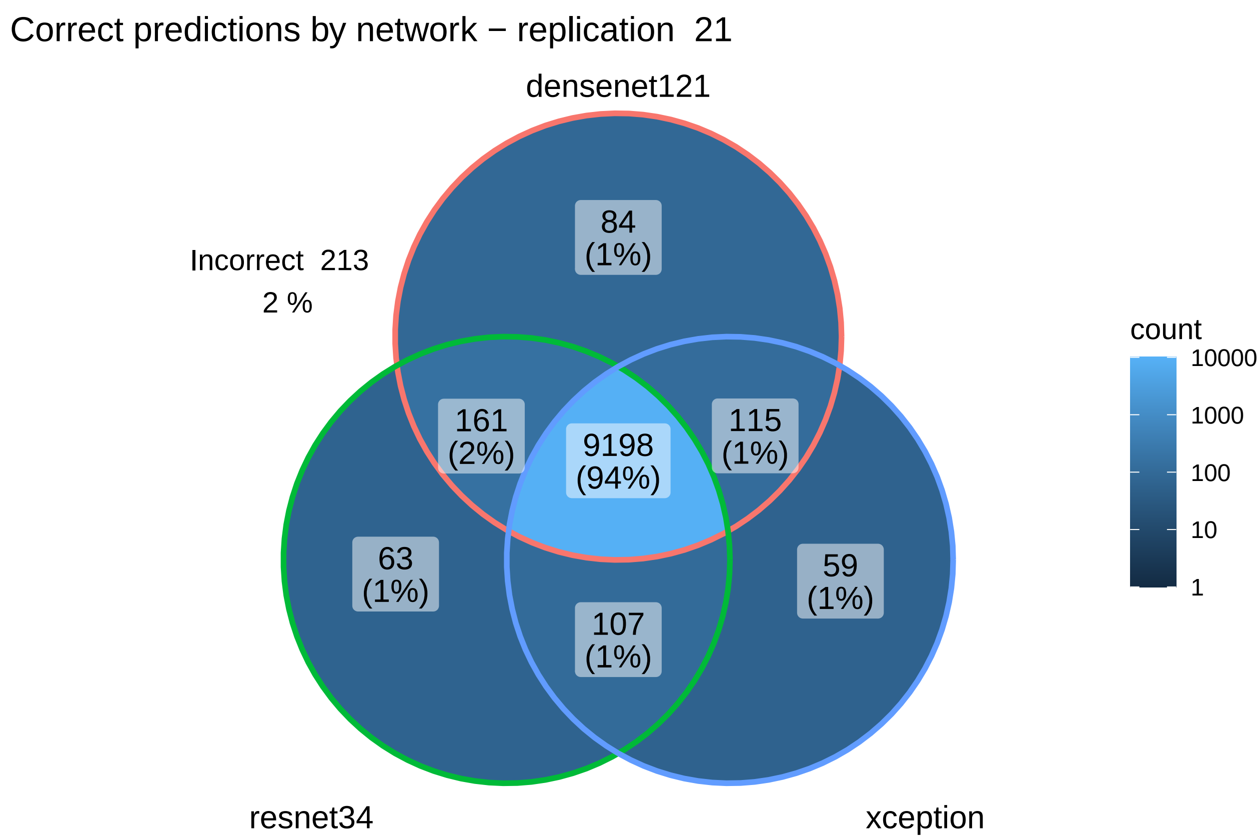

Correct predictions by network − replication 16
densenet121
Incorrect 222
2 %
70
(1%)
192
(2%)
9145
(94%)
142
(1%)
65
(1%)
98
(1%)
66
(1%)
resnet34
xception
count
10000
1000
100
10
1

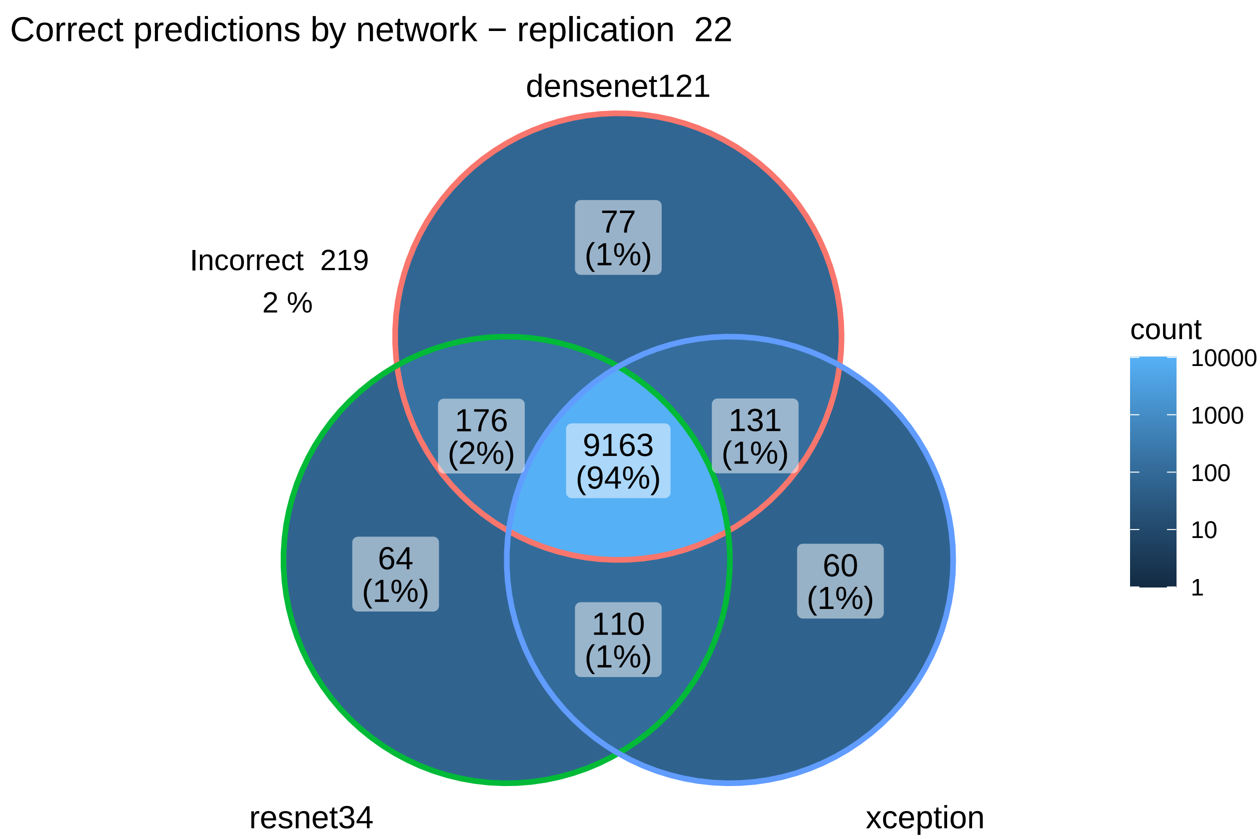

Correct predictions by network − replication 17
densenet121
69
(1%)
Incorrect 216
2 %
169
(2%)
9162
(94%)
140
(1%)
75
(1%)
108
(1%)
61
(1%)
resnet34
xception
count
10000
1000
100
10
1

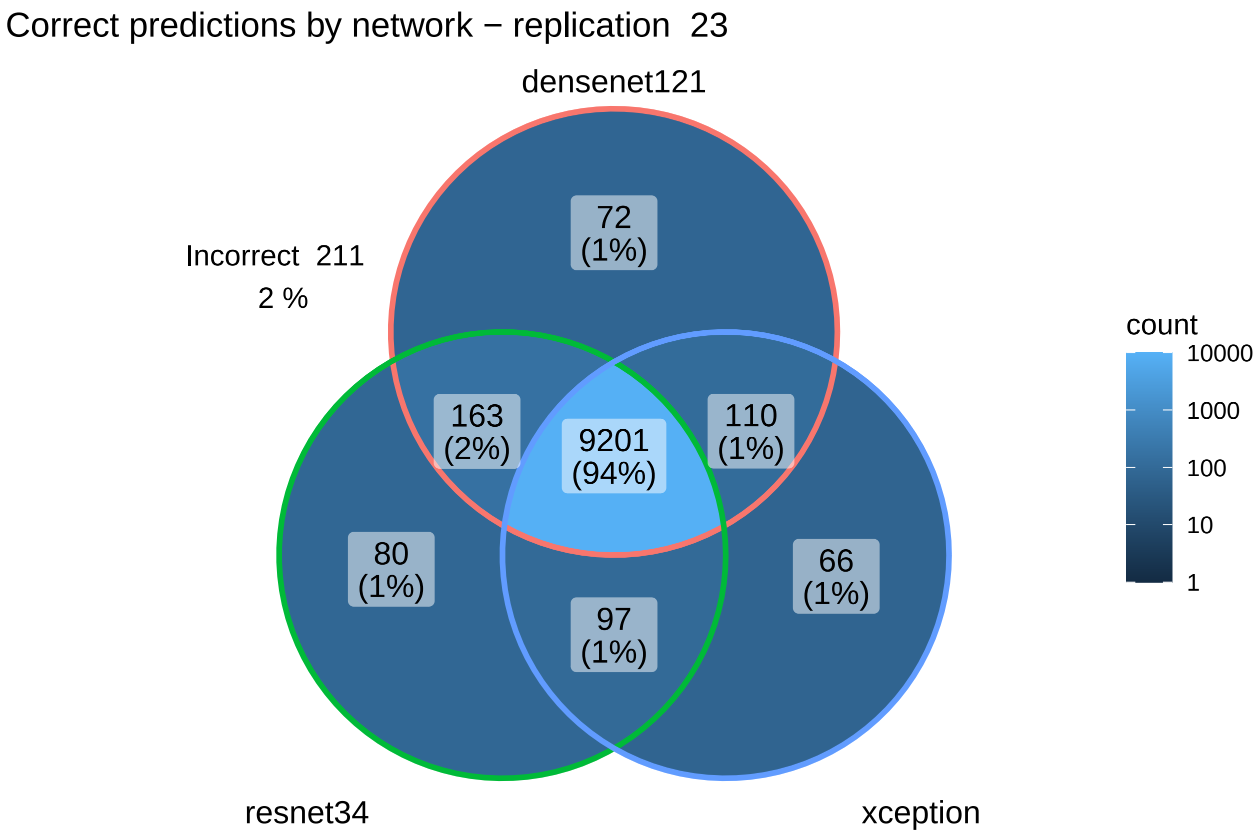

Correct predictions by network − replication 18
densenet121
81
(1%)
Incorrect 210
2 %
171
(2%)
9176
(94%)
131
(1%)
82
(1%)
92
(1%)
57
(1%)
resnet34
xception
count
10000
1000
100
10
1

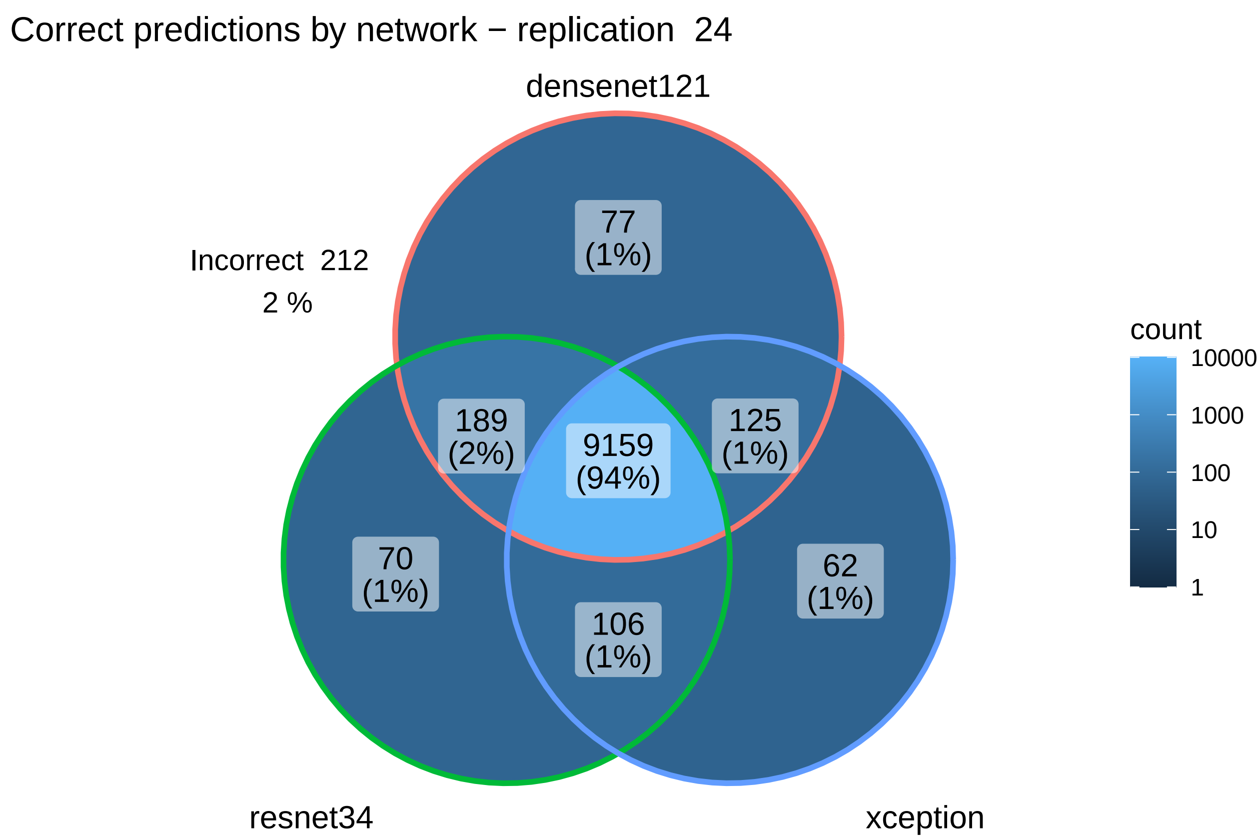

Correct predictions by network − replication 19
densenet121
Incorrect 224
2 %
60
(1%)
179
(2%)
9159
(94%)
138
(1%)
80
(1%)
103
(1%)
57
(1%)
resnet34
xception
count
10000
1000
100
10
1


Correct predictions by network − replication 20
densenet121
80
(1%)
Incorrect 205
2 %
143
(1%)
9198
(94%)
119
(1%)
84
(1%)
66
(1%)
105
(1%)
resnet34
xception
count
10000
1000
100
10
1

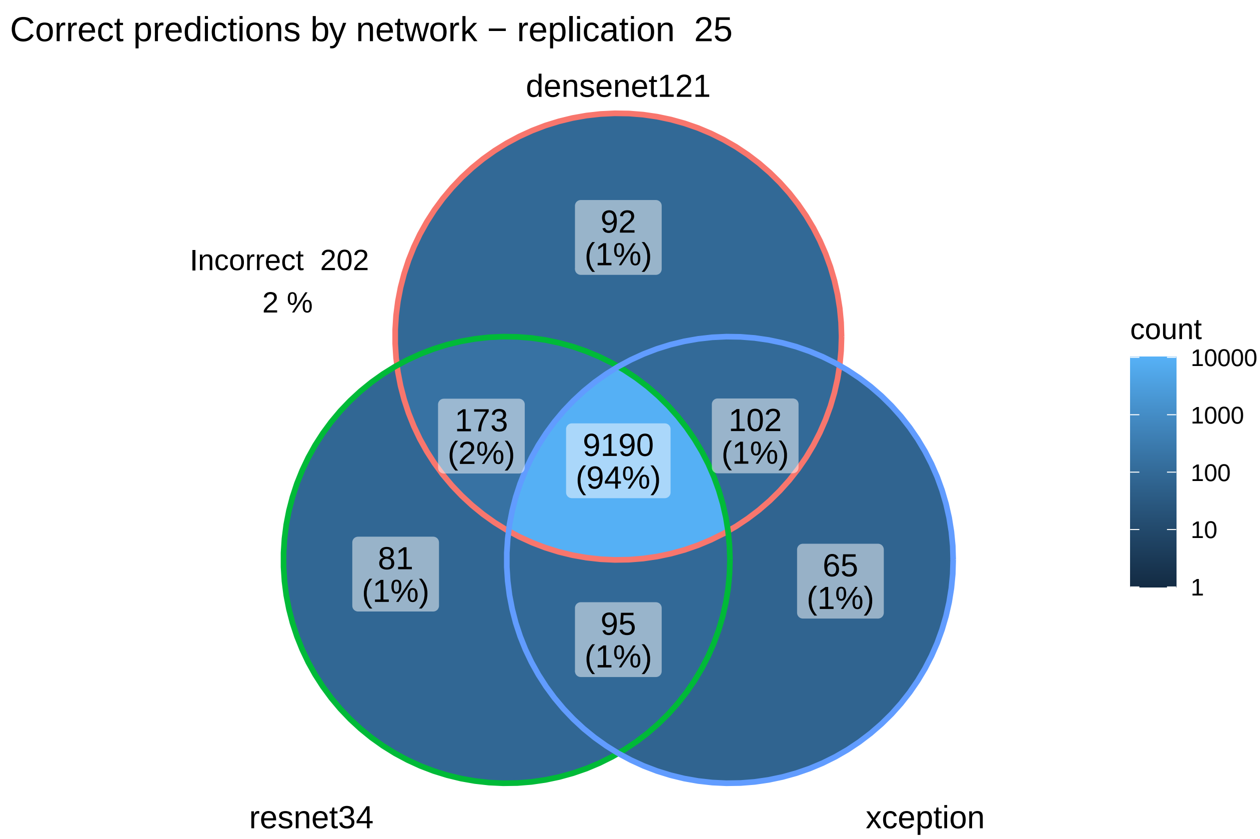

Correct predictions by network − replication 21
densenet121
84
(1%)
Incorrect 213
2 %
161
(2%)
9198
(94%)
115
(1%)
63
(1%)
107
(1%)
59
(1%)
resnet34
xception
count
10000
1000
100
10
1

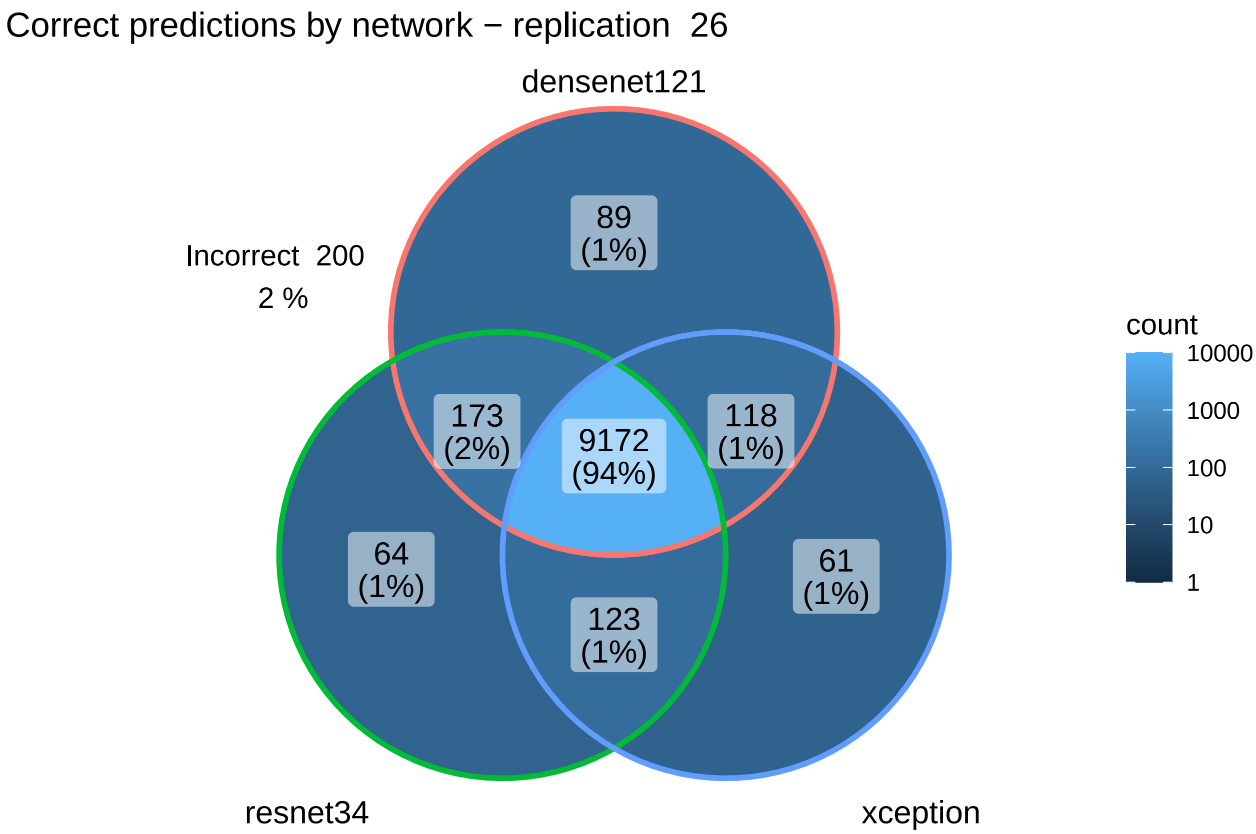

Correct predictions by network − replication 22
densenet121
Incorrect 219
2 %
77
(1%)
176
(2%)
9163
(94%)
131
(1%)
64
(1%)
110
(1%)
60
(1%)
resnet34
xception
count
10000
1000
100
10
1

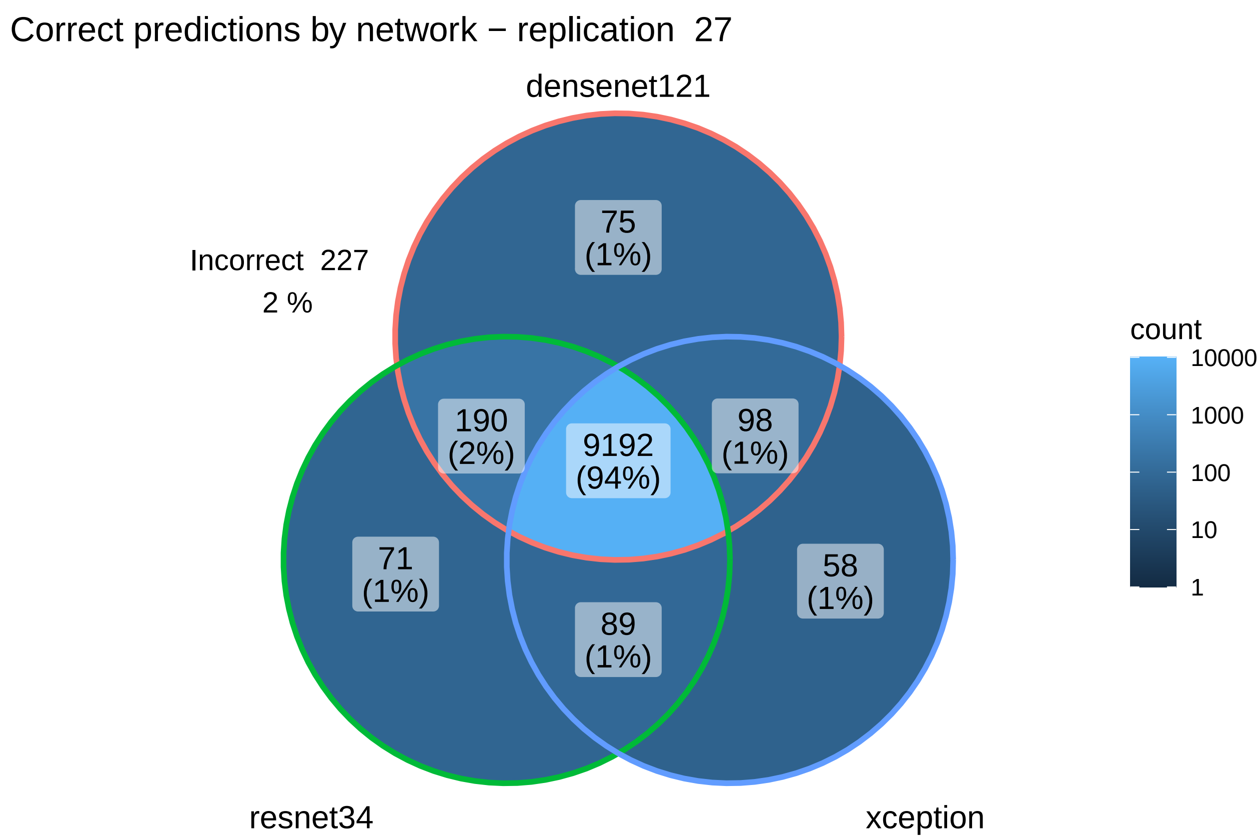

Correct predictions by network − replication 23
densenet121
72
(1%)
Incorrect 211
2 %
163
(2%)
9201
(94%)
110
(1%)
80
(1%)
97
(1%)
66
(1%)
resnet34
xception
count
10000
1000
100
10
1

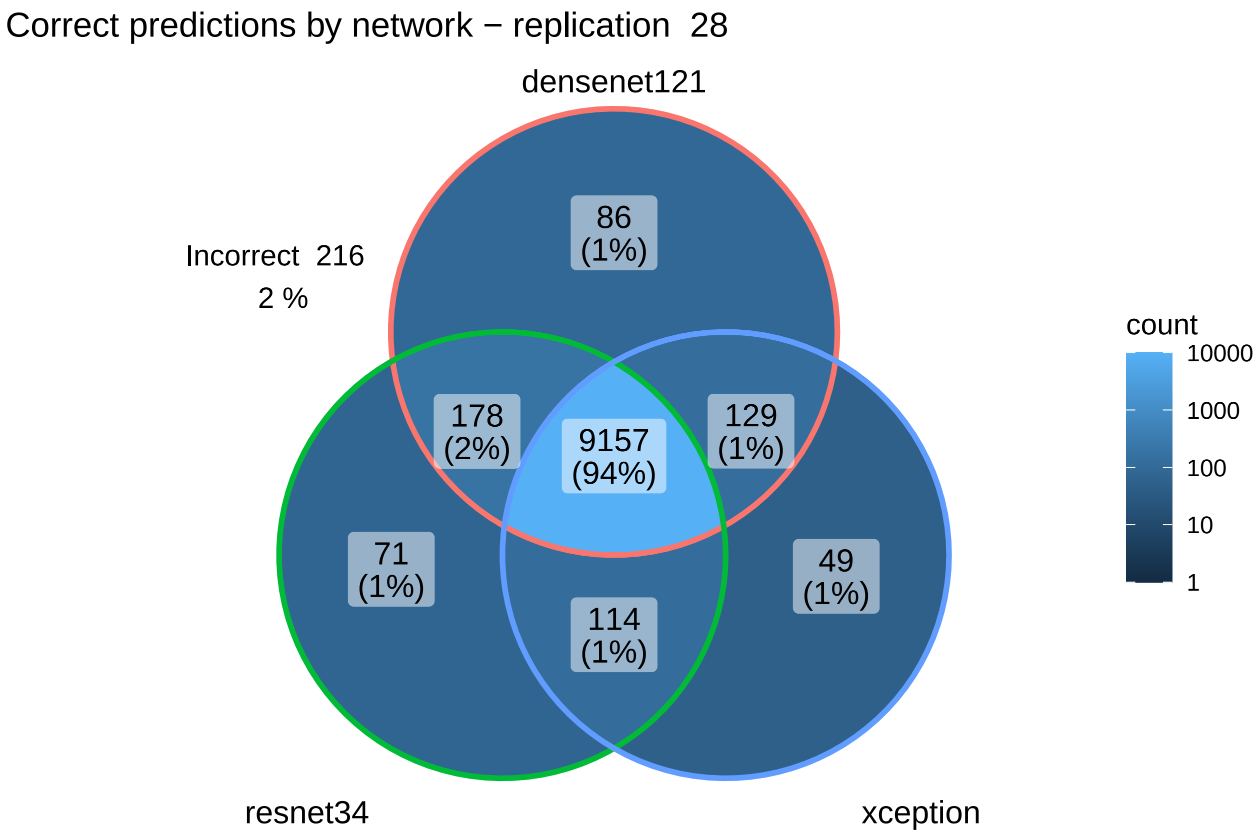


Correct predictions by network − replication 24
densenet121
Incorrect 212
2 %
77
(1%)
189
(2%)
9159
(94%)
125
(1%)
70
(1%)
106
(1%)
62
(1%)
resnet34
xception
count
10000
1000
100
10
1

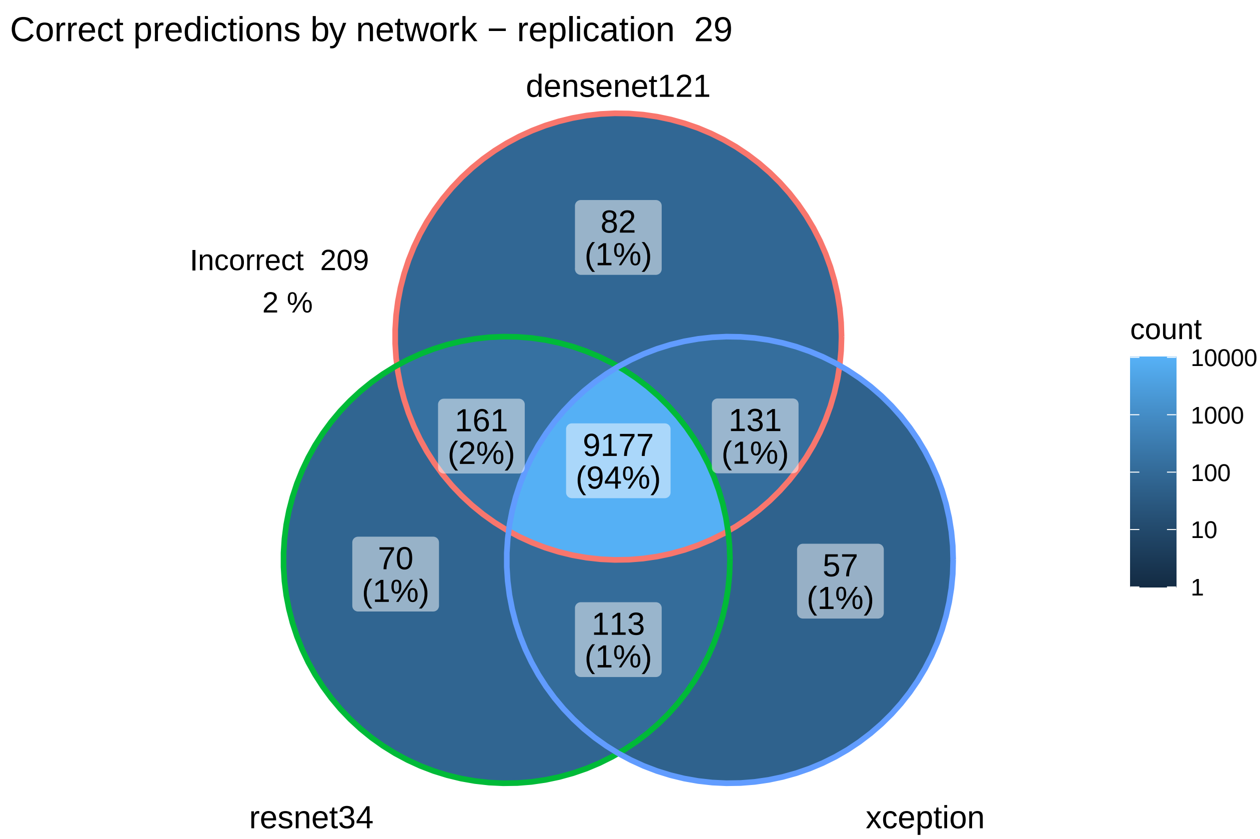

Correct predictions by network − replication 25
densenet121
Incorrect 202
2 %
92
(1%)
173
(2%)
9190
(94%)
102
(1%)
81
(1%)
95
(1%)
65
(1%)
resnet34
xception
count
10000
1000
100
10
1

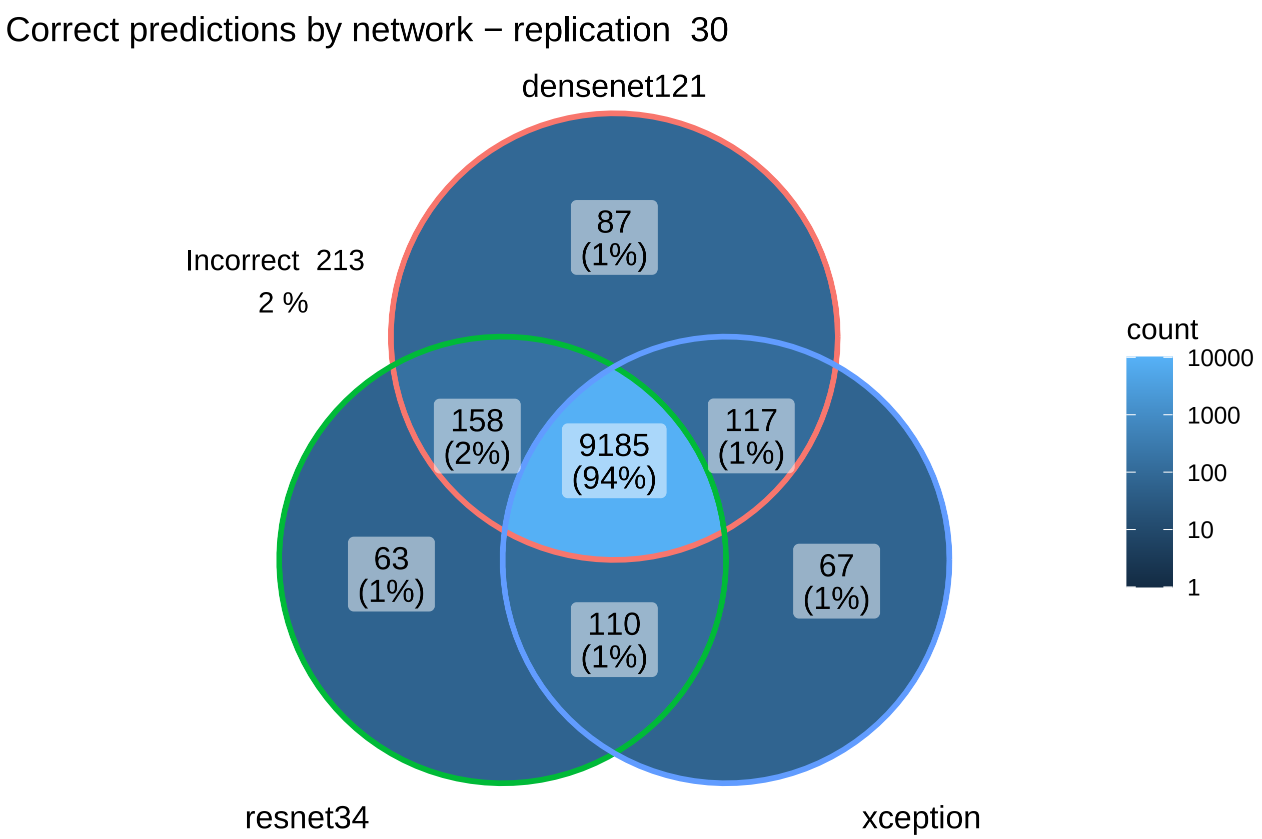

Correct predictions by network − replication 26
densenet121
Incorrect 200
2 %
89
(1%)
173
(2%)
9172
(94%)
118
(1%)
64
(1%)
123
(1%)
61
(1%)
resnet34
xception
count
10000
1000
100
10
1


Correct predictions by network − replication 27
densenet121
Incorrect 227
2 %
75
(1%)
190
(2%)
9192
(94%)
98
(1%)
71
(1%)
89
(1%)
58
(1%)
resnet34
xception
count
10000
1000
100
10
1



Correct predictions by network − replication 28
densenet121
Incorrect 216
2 %
86
(1%)
178
(2%)
9157
(94%)
129
(1%)
71
(1%)
114
(1%)
49
(1%)
resnet34
xception
count
10000
1000
100
10
1


Correct predictions by network − replication 29
densenet121
Incorrect 209
2 %
82
(1%)
161
(2%)
9177
(94%)
131
(1%)
70
(1%)
113
(1%)
57
(1%)
resnet34
xception
count
10000
1000
100
10
1

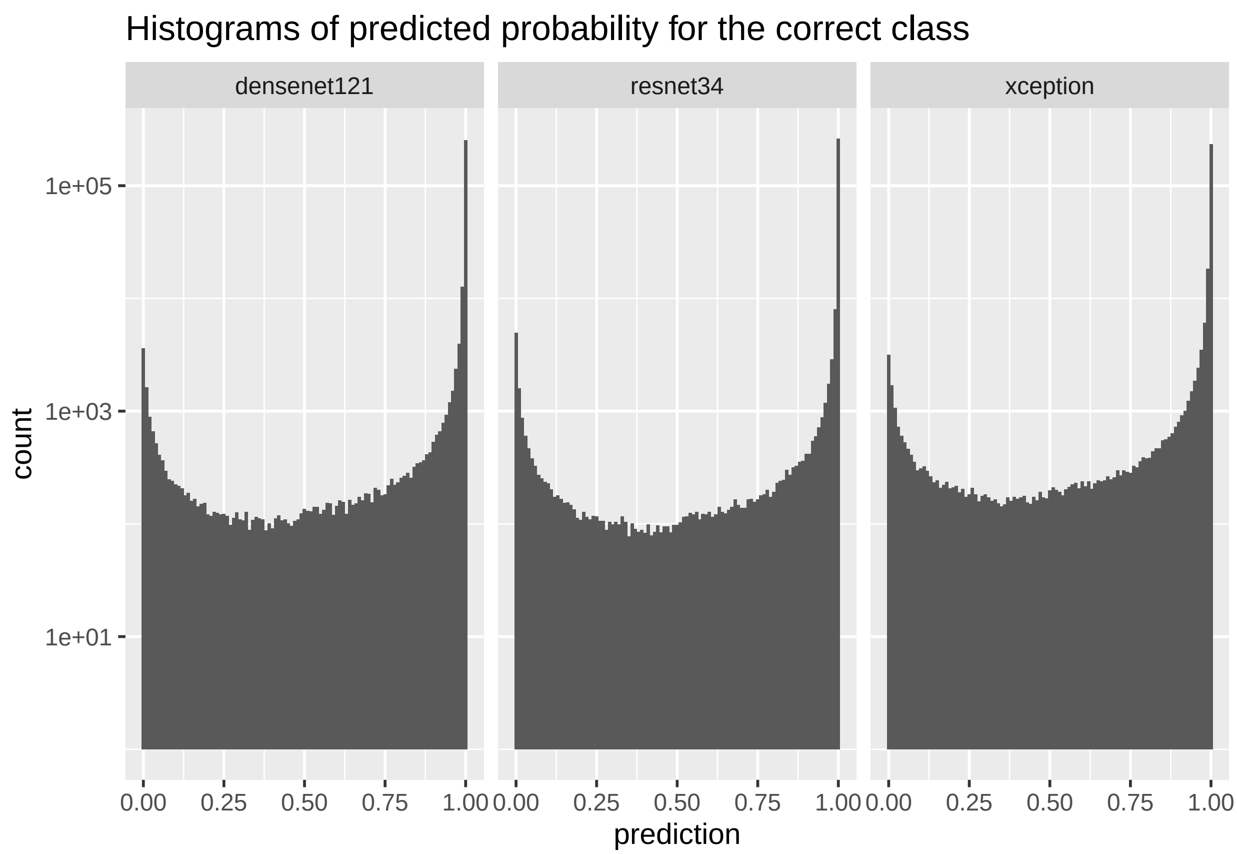


In all replications, around 94% of data that was correctly classified was correctly classified by all networks. Network with most exclusively correct predictions is in majority of replications densenet. Densenet annd resnet have in majority of replications most common correct predictions amongs the pairs of networks.

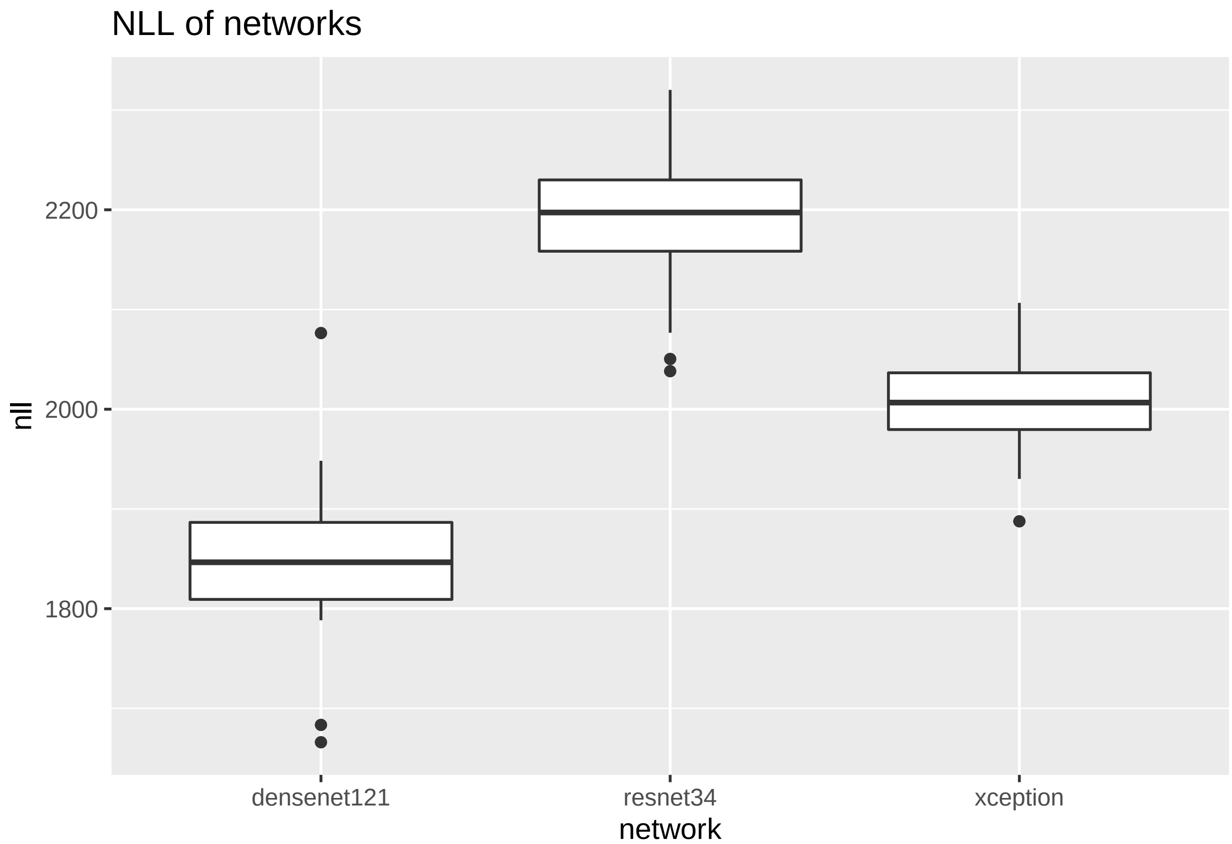
For clearer visualization, we will plot just the predicted probability of the correct class for all the methods.

```
preds <- nets_outputs$test_outputs
for (ri in repls + 1)
{
  for (net_i in seq_along(nets_outputs[["networks"]]))
  {
    preds[ri, net_i, ,] <- softmax(preds[ri, net_i, , ])
  }
}
nets_test_cor_probs <- gather(preds, 1 + nets_outputs$test_labels[1, ], 3, 4)
nets_test_cor_probs <- melt(nets_test_cor_probs)
nets_test_cor_probs <- nets_test_cor_probs[, c(-3, -4)]
names(nets_test_cor_probs) <- c("replication", "network", "prediction")
nets_test_cor_probs$network <- as.factor(nets_test_cor_probs$network)
levels(nets_test_cor_probs$network) <- nets_outputs$networks
```

```
nets_cor_preds_histo <- ggplot(data=nets_test_cor_probs) + geom_histogram(mapping=aes(x=prediction), bi
  ggtitle("Histograms of predicted probability for the correct class") + facet_wrap(~network) + scale_y
nets_cor_preds_histo
```



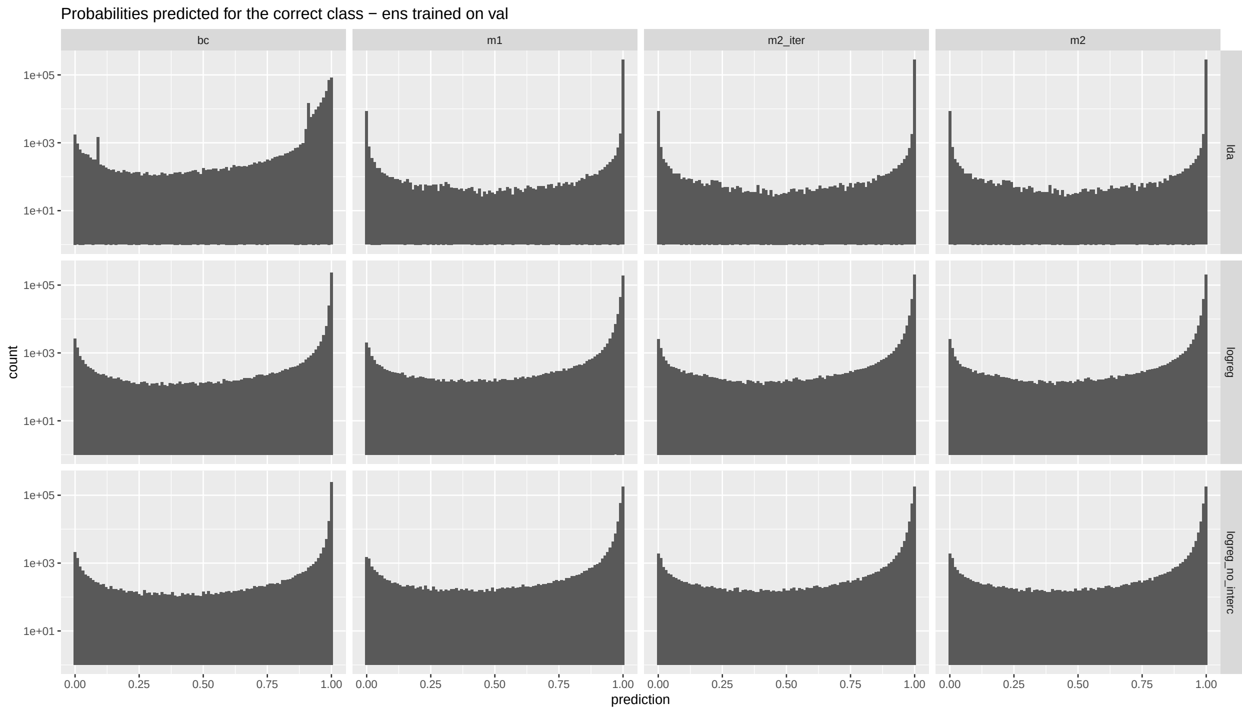

```
networks_nll <- ggplot(data=net_results) + geom_boxplot(mapping=aes(x=network, y=nll)) + ggtitle("NLL o
networks_nll
```



Networks nll seems to have good correspondence with lowest prediction probability column in the previous histograms.

```
val_ens_cor_probs <- gather(ens_outputs$val_training, 1 + nets_outputs$test_labels[1, ], 4, 5)
val_ens_cor_probs <- melt(val_ens_cor_probs)
val_ens_cor_probs <- val_ens_cor_probs[, c(-4, -5)]
names(val_ens_cor_probs) <- c("replication", "combining_method", "coupling_method", "prediction")
val_ens_cor_probs[, c("combining_method", "coupling_method")] <- lapply(val_ens_cor_probs[, c("combining
levels(val_ens_cor_probs$combining_method) <- ens_outputs$combining_methods
levels(val_ens_cor_probs$coupling_method) <- ens_outputs$coupling_methods
```

```
val_ens_cor_preds_histo <- ggplot(data=val_ens_cor_probs) + geom_histogram(mapping=aes(x=prediction), bi
val_ens_cor_preds_histo
```

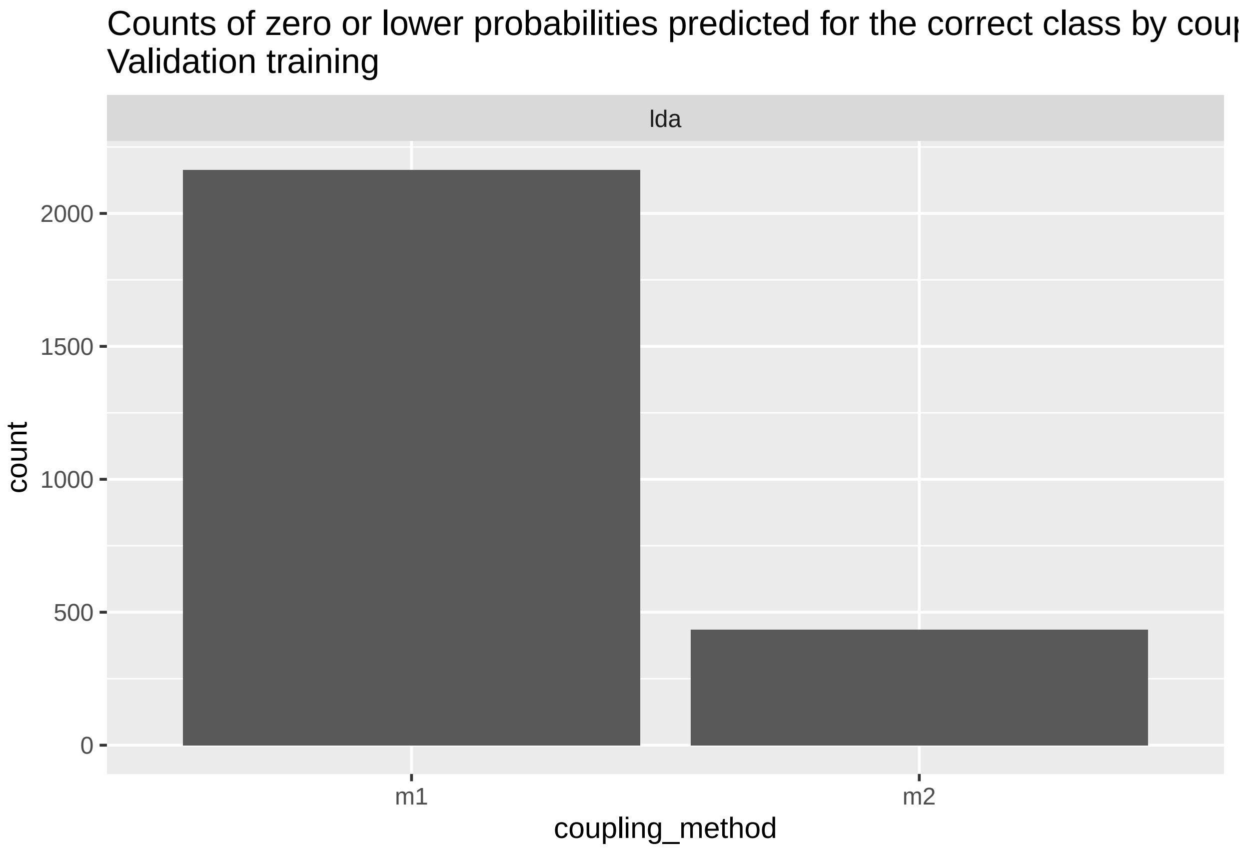


In the case of combining method lda, coupling method bc produces far fewer probabilities falling into the lowest bin for the correct class than m1 and m2. For the logistic regression combining method, predictions are distributed more smoothly and there are fewer predictions in the extreme values.

```
val_ens_zero_counts <- ggplot(data=val_ens_cor_probs[val_ens_cor_probs$prediction <= 0, ]) + geom_histog
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

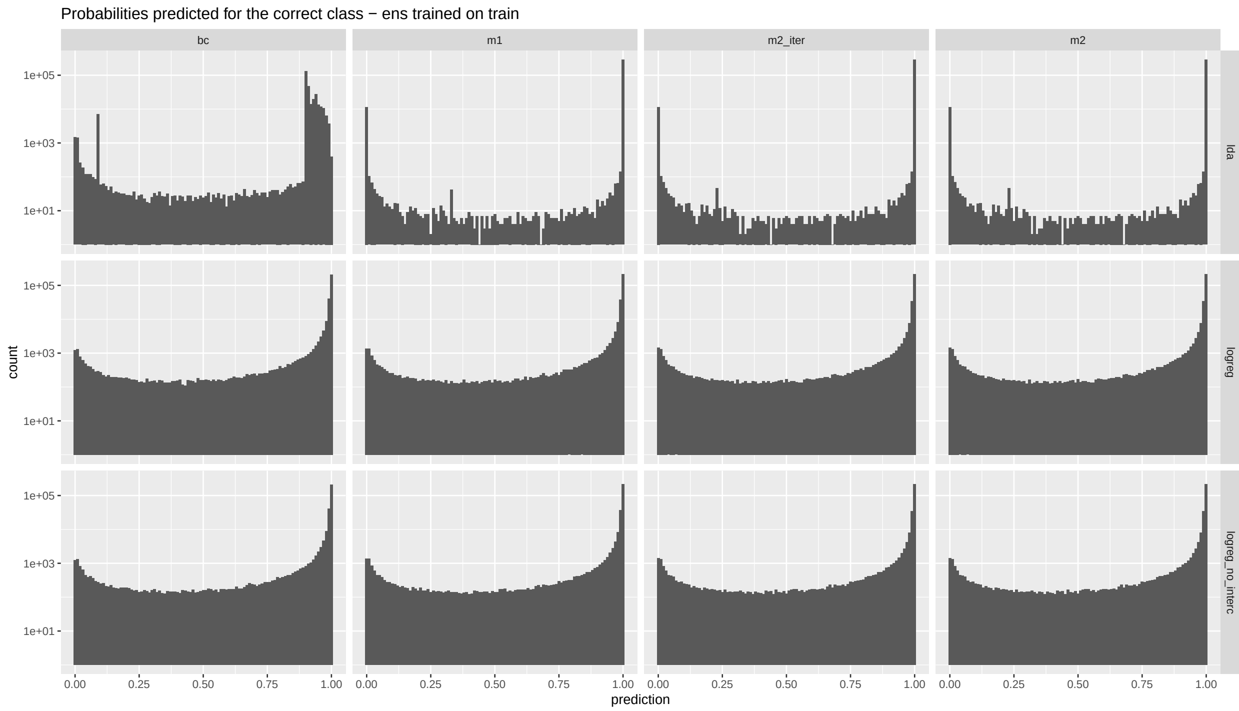
```
val_ens_zero_counts
```



m2_iter and bc didn't produce any zero probability outputs. Neither were there any zero probability outputs for combining method logistic regression.

```
train_ens_cor_probs <- gather(ens_outputs$train_training, 1 + nets_outputs$test_labels[1, ], 4, 5)
train_ens_cor_probs <- melt(train_ens_cor_probs)
train_ens_cor_probs <- train_ens_cor_probs[, c(-4, -5)]
names(train_ens_cor_probs) <- c("replication", "combining_method", "coupling_method", "prediction")
train_ens_cor_probs[, c("combining_method", "coupling_method")] <- lapply(train_ens_cor_probs[, c("comb
levels(train_ens_cor_probs$combining_method) <- ens_outputs$combining_methods
levels(train_ens_cor_probs$coupling_method) <- ens_outputs$coupling_methods
```

```
train_ens_cor_preds_histo <- ggplot(data=train_ens_cor_probs) + geom_histogram(mapping=aes(x=prediction
train_ens_cor_preds_histo
```

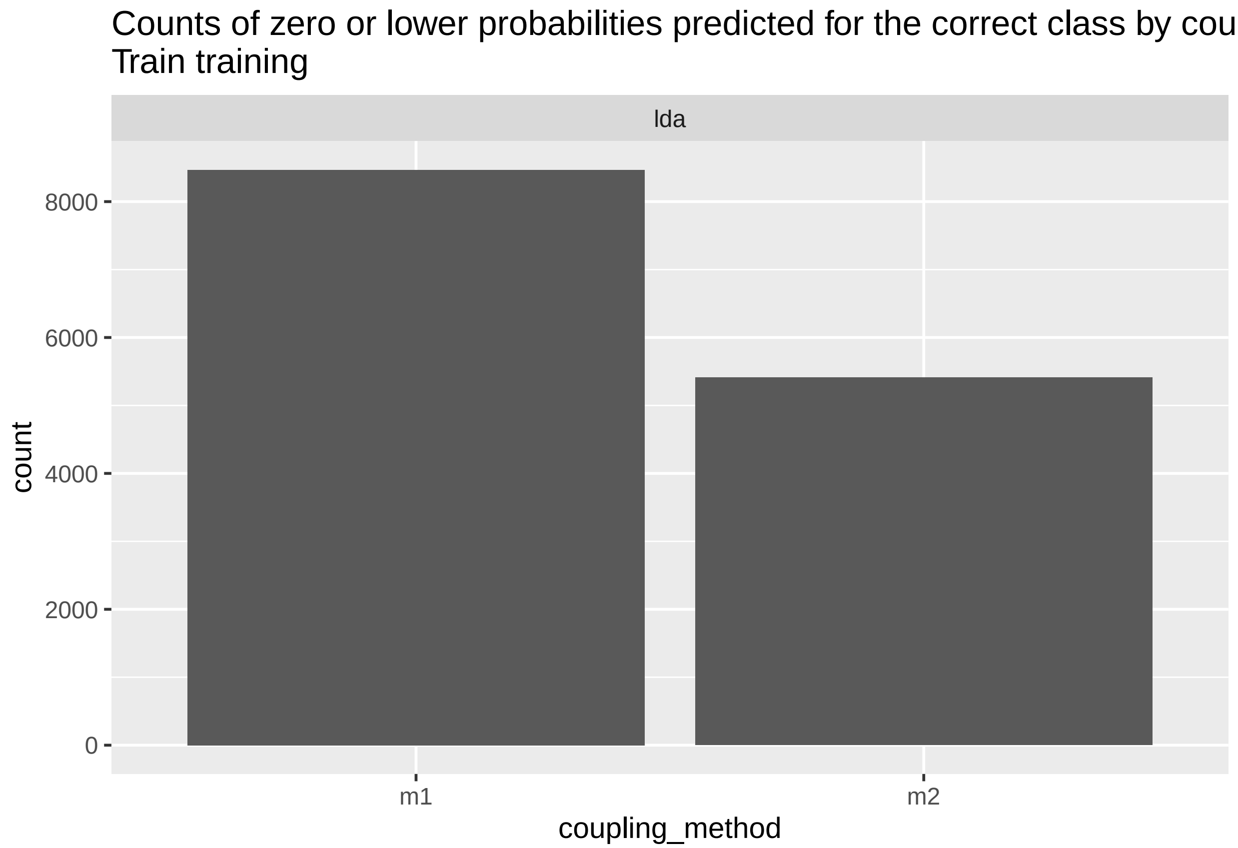


Also in this case, for combining method lda, coupling method bc produces far fewer probabilities falling into the lowest bin for the correct class than m1 and m2. Outputs employing coumbining method logistic regression are much more smoothly distributed.

```
train_ens_zero_counts <- ggplot(data=train_ens_cor_probs[train_ens_cor_probs$prediction <= 0, ]) + geom_
```

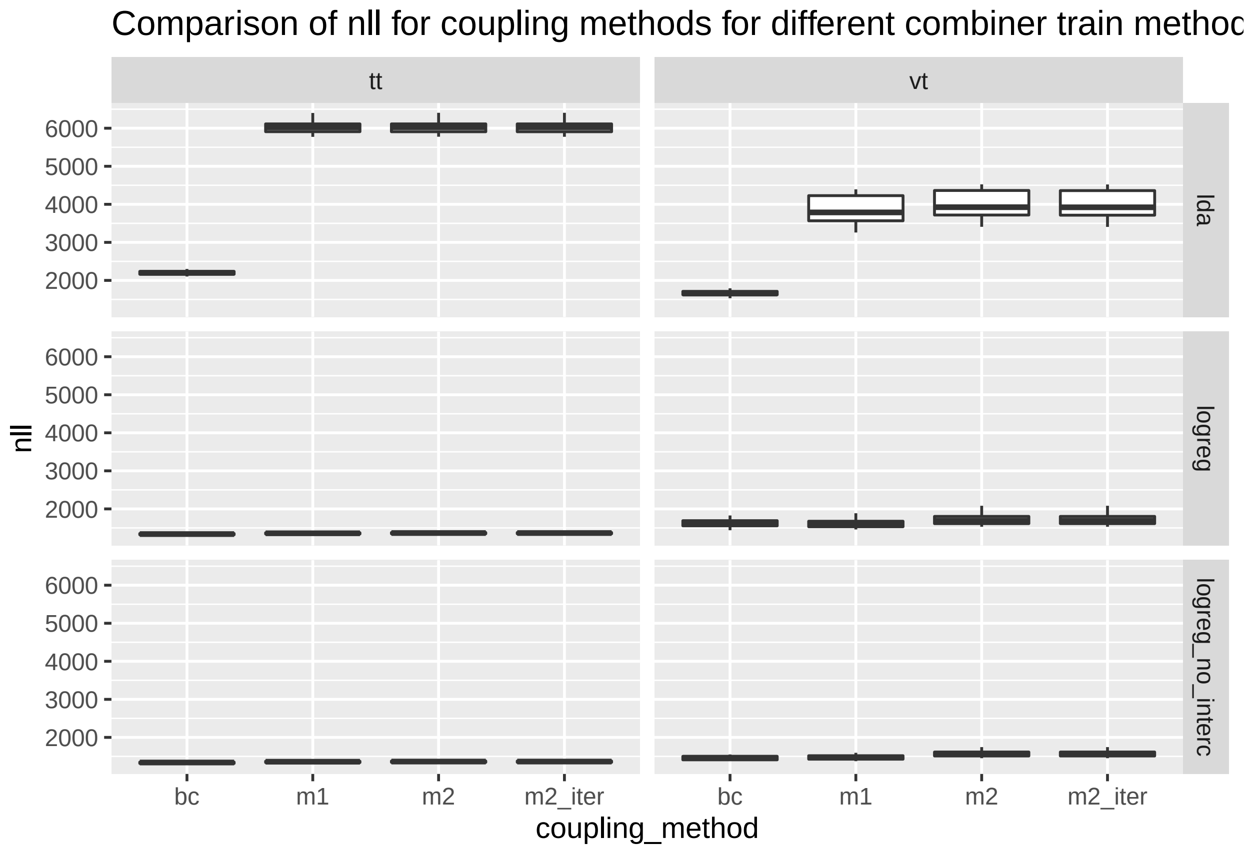
```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

```
train_ens_zero_counts
```



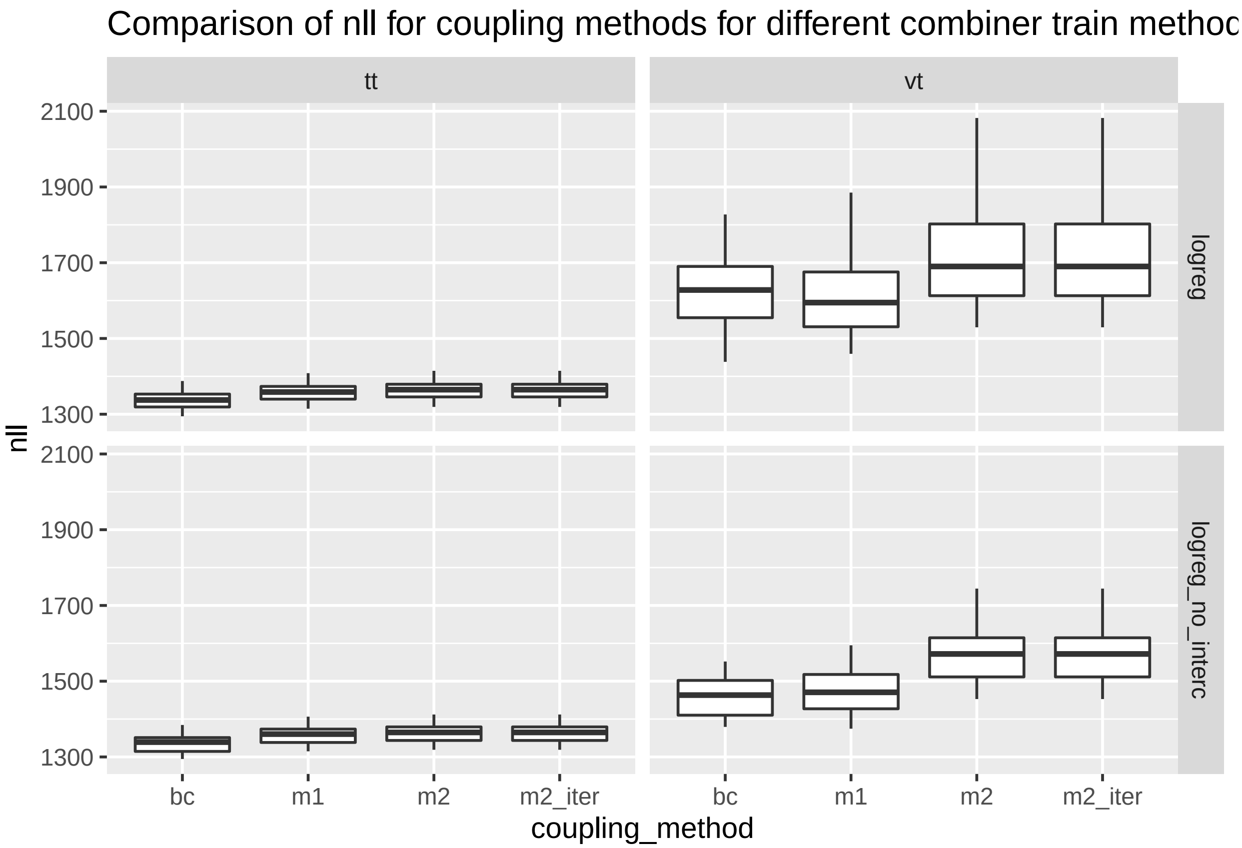
m2_iter and bc didn't produce any zero probability outputs. Neither were there any zero probability outputs for combining method logistic regression.

```
val_ens_nll <- ggplot(data=ens_results) + geom_boxplot(mapping=aes(x=coupling_method, y=nll)) + facet_g
  ggtitle("Comparison of nll for coupling methods for different combiner train methodologies")
val_ens_nll
```



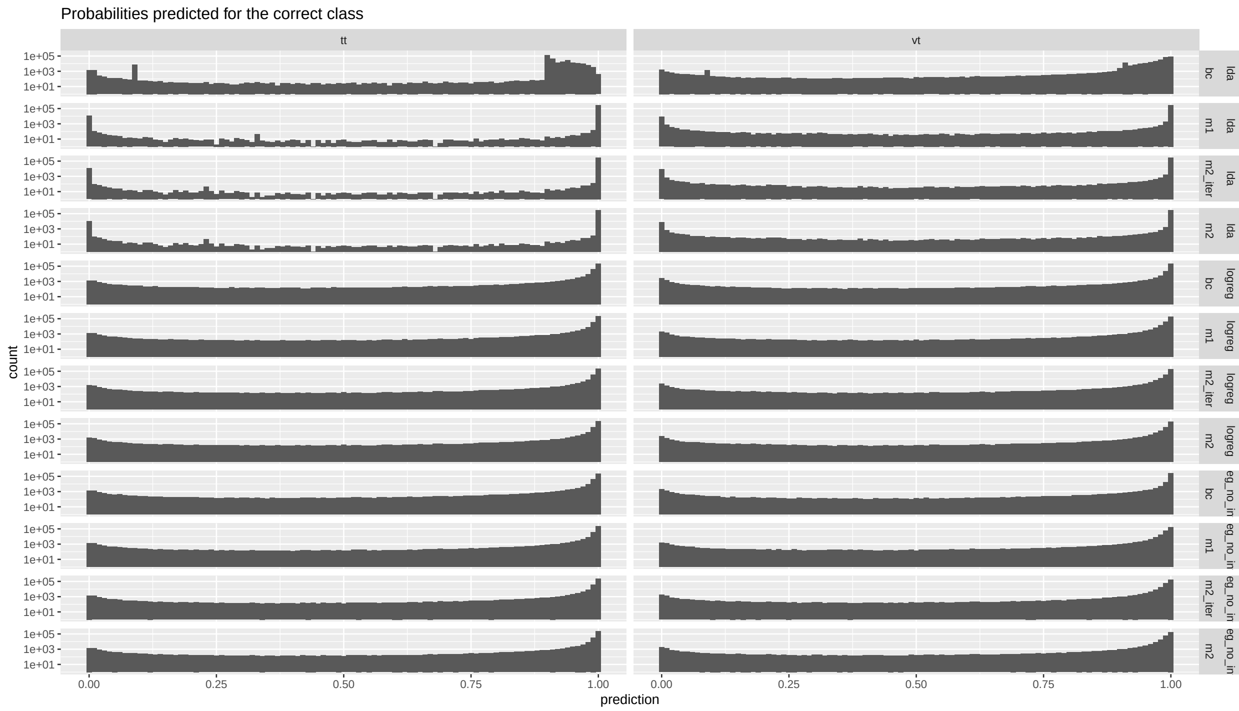
For combining method lda, coupling method bc has by far best results. For other combining methods, bc is still best, but the difference is much smaller.

```
val_ens_nll <- ens_results %>% filter(combining_method!="lda") %>% ggplot() + geom_boxplot(mapping=aes(
  ggtitle("Comparison of nll for coupling methods for different combiner train methodologies")
val_ens_nll
```




```
val_ens_cor_probs$train_type <- "vt"
train_ens_cor_probs$train_type <- "tt"
ens_cor_probs <- rbind(val_ens_cor_probs, train_ens_cor_probs)
```

```
ens_cor_preds_histo <- ggplot(data=ens_cor_probs) + geom_histogram(mapping=aes(x=prediction), binwidth=0
ens_cor_preds_histo
```



For combining method lda Bayes covariant coupling method produces more uniformly distributed predictions than methods m1 and m2. Also, there is a big difference in each method between ensemble trained on validation and ensemble trained on train set. Ensembles trained on validation set produce generally more uniformly distributed predictions. However, ensembles trained on training set attain statistically significantly higher accuracy. For other combining methods, predictions are smoothly distributed for both training methodologies. Also in this case, ensembles trained on training set attain statistically significantly higher accuracy.

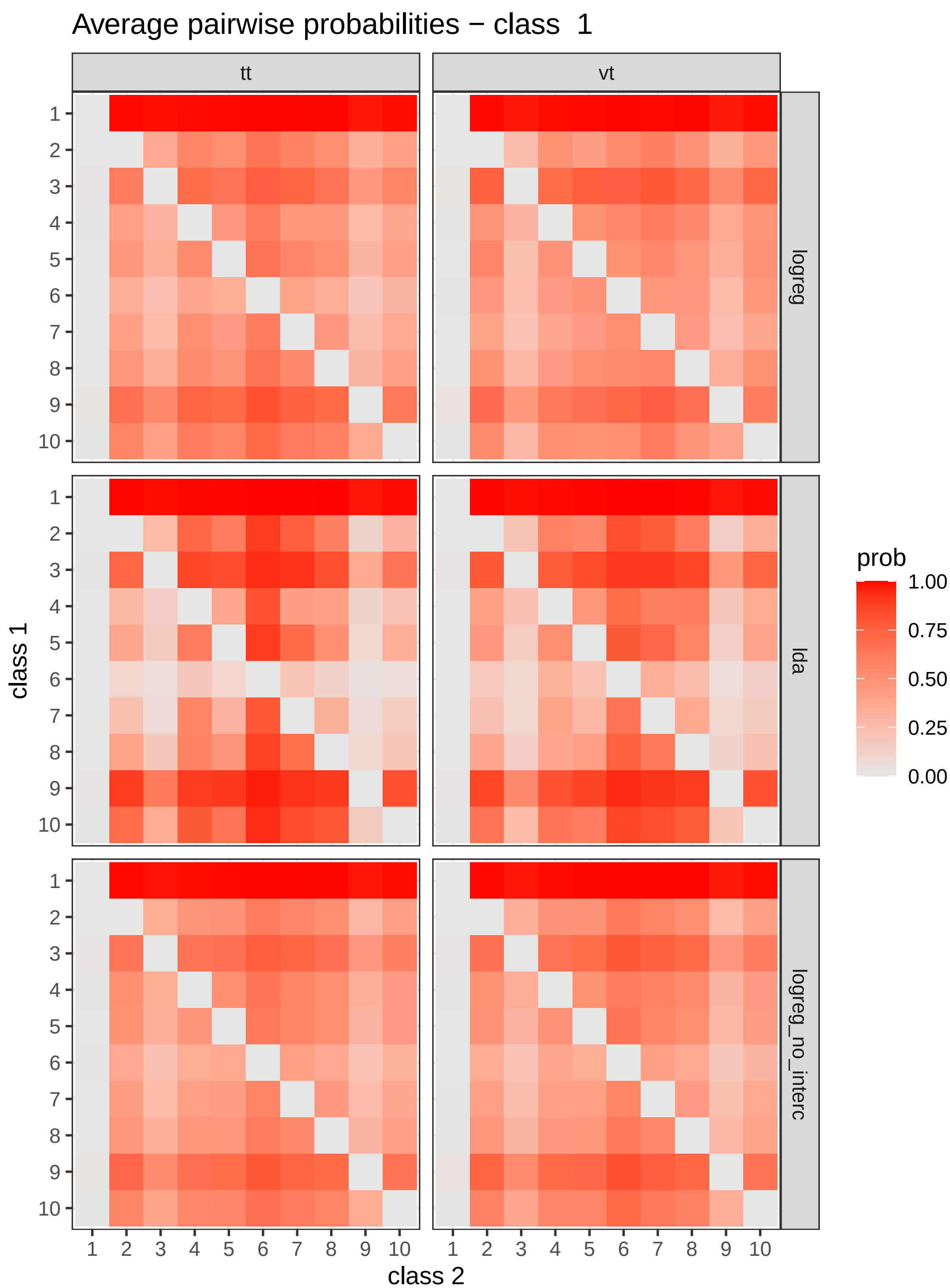
```
df_val_Rs <- melt(np$load(file.path(base_dir, "val_training_class_aggr_R.npy")))
df_train_Rs <- melt(np$load(file.path(base_dir, "train_training_class_aggr_R.npy")))
co_m_R <- read.csv(file.path(base_dir, "R_mat_co_m_names.csv"), header=FALSE)
```

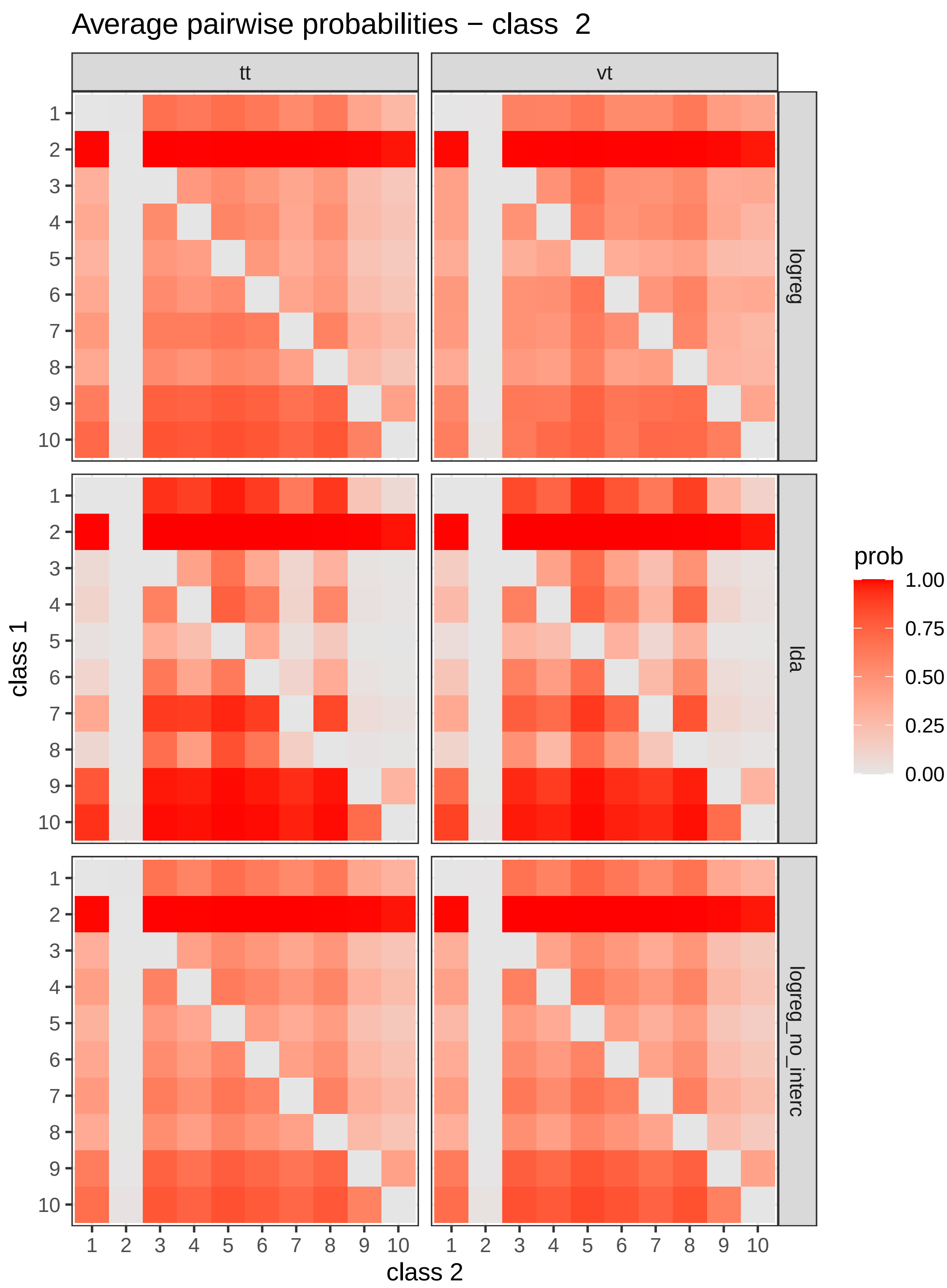
```
names(df_val_Rs) <- c("combining_method", "precision", "class", "class1", "class2", "prob")
names(df_train_Rs) <- c("combining_method", "precision", "class", "class1", "class2", "prob")

df_val_Rs[,c("class", "class1", "class2", "combining_method")] <- lapply(df_val_Rs[,c("class", "class1"
df_train_Rs[,c("class", "class1", "class2", "combining_method")] <- lapply(df_train_Rs[,c("class", "cla

levels(df_val_Rs$combining_method) <- co_m_R$V1
levels(df_train_Rs$combining_method) <- co_m_R$V1
```

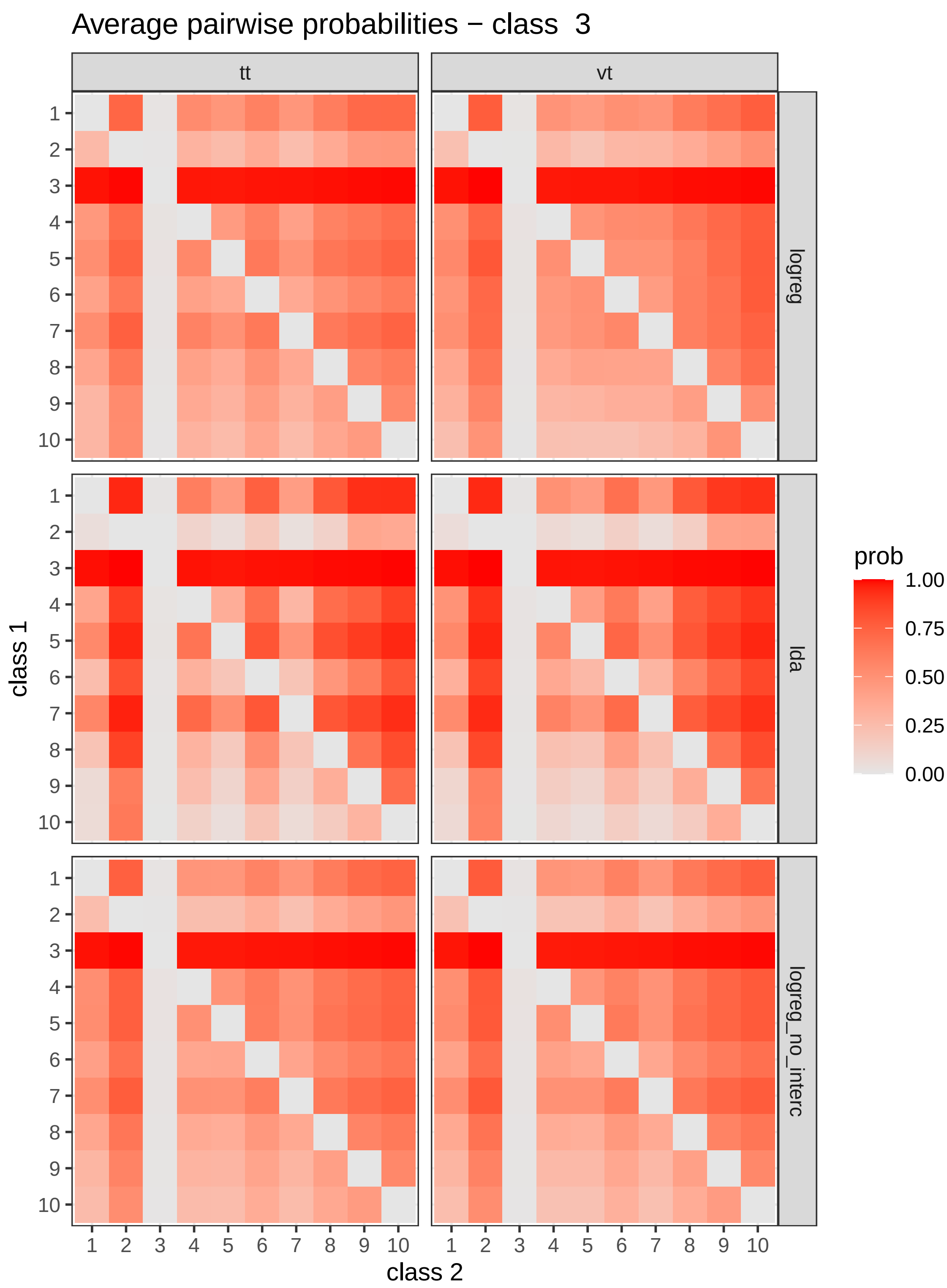
```
df_val_Rs$train_type <- "vt"
df_train_Rs$train_type <- "tt"
class_mean_Rs <- rbind(df_val_Rs, df_train_Rs)

df_aggr_Rs_diff <- class_mean_Rs %>% pivot_wider(names_from = train_type, values_from = prob) %>% mutate
```

```
for (cls in 1:classes)
{
  cur_class_Rs <- class_mean_Rs %>% filter(class == cls)
```

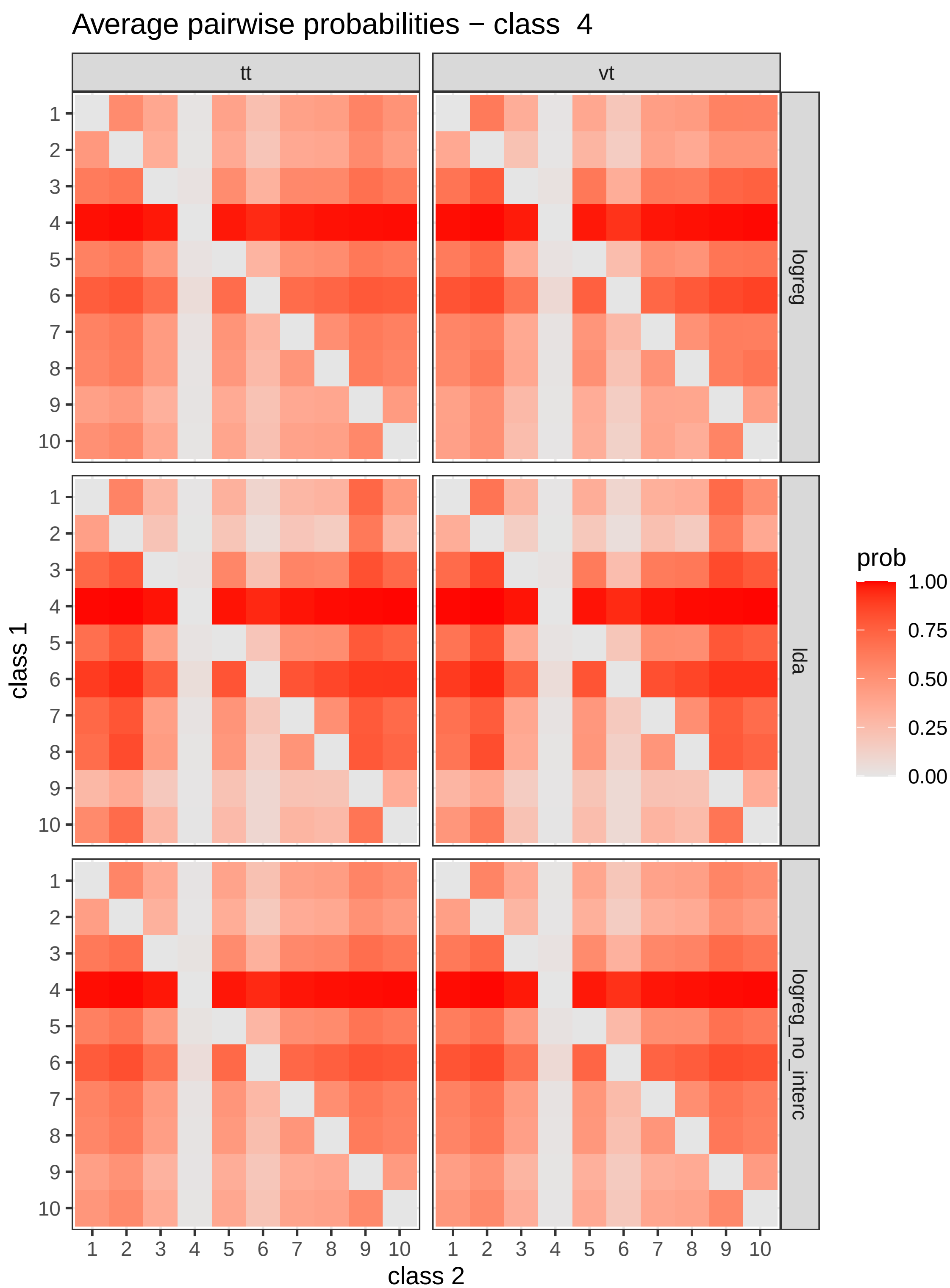
```
  plot_cls <-  ggplot(cur_class_Rs, aes(x = class2, y = class1)) +
    geom_raster(aes(fill=prob)) +
    facet_grid(rows=vars(combining_method), cols=vars(train_type)) +
    scale_fill_gradient(low="grey90", high="red", limits=c(0, 1)) +
    scale_y_discrete(limits=rev) +
    labs(x="class 2", y="class 1", title=paste("Average pairwise probabilities - class ", cls)) +
    theme_bw()

  print(plot_cls)
}
```

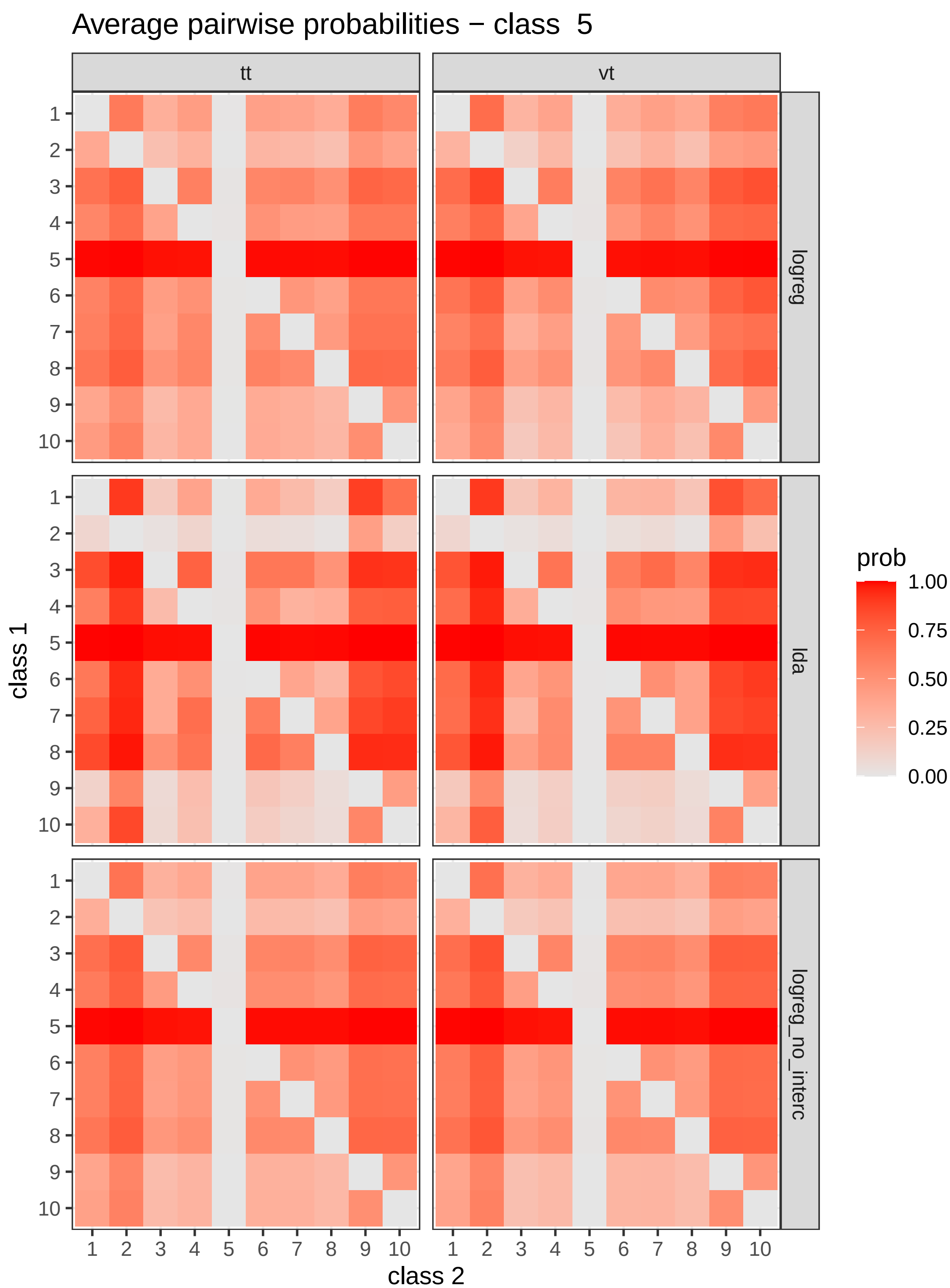

Average pairwise probabilities − class 1
tt
vt
logreg
lda
logreg_no_interc
class 1
class 2
prob
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10

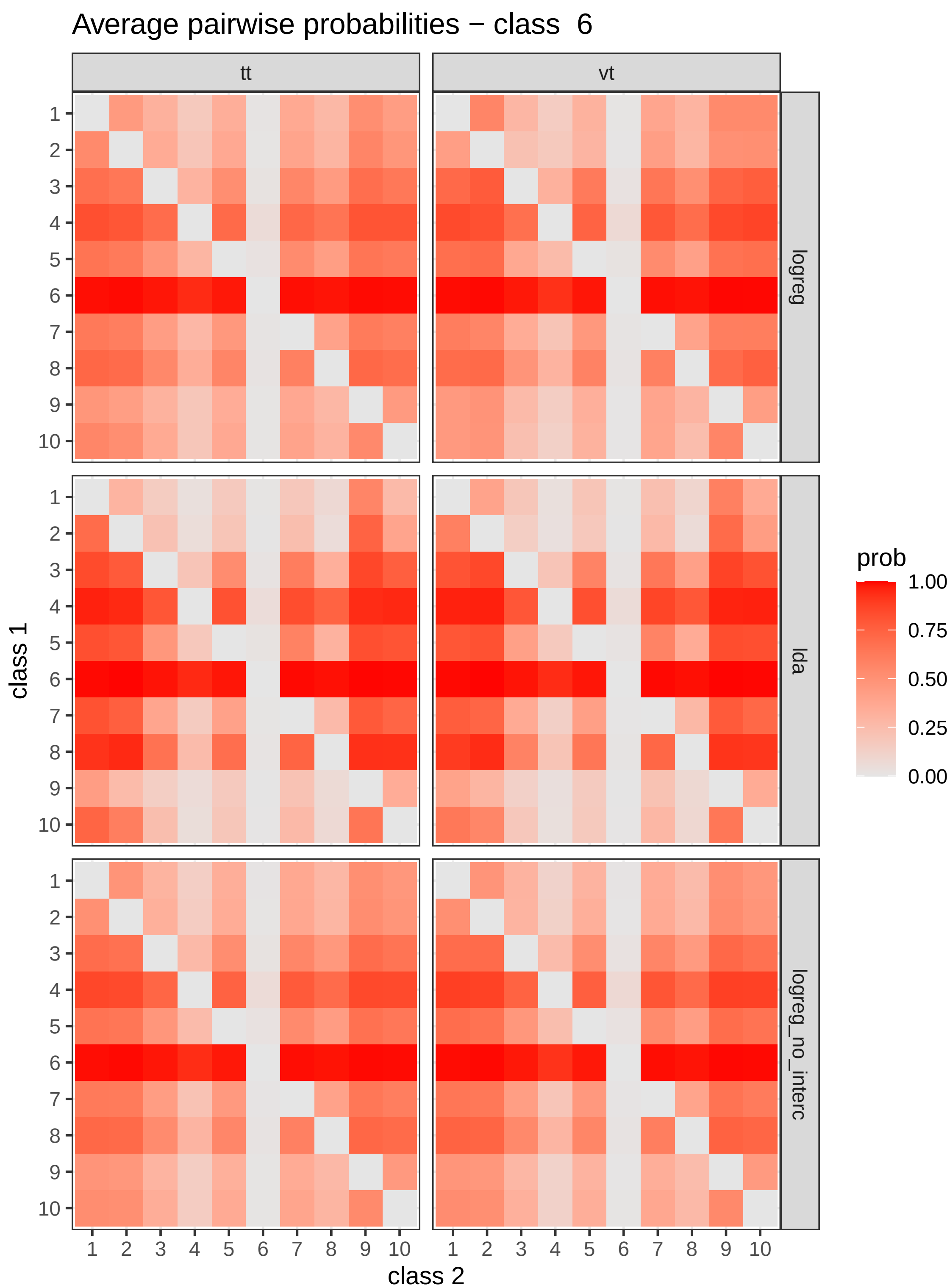

Average pairwise probabilities − class 2
tt
vt
logreg
lda
logreg_no_interc
class 1
class 2
prob
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10



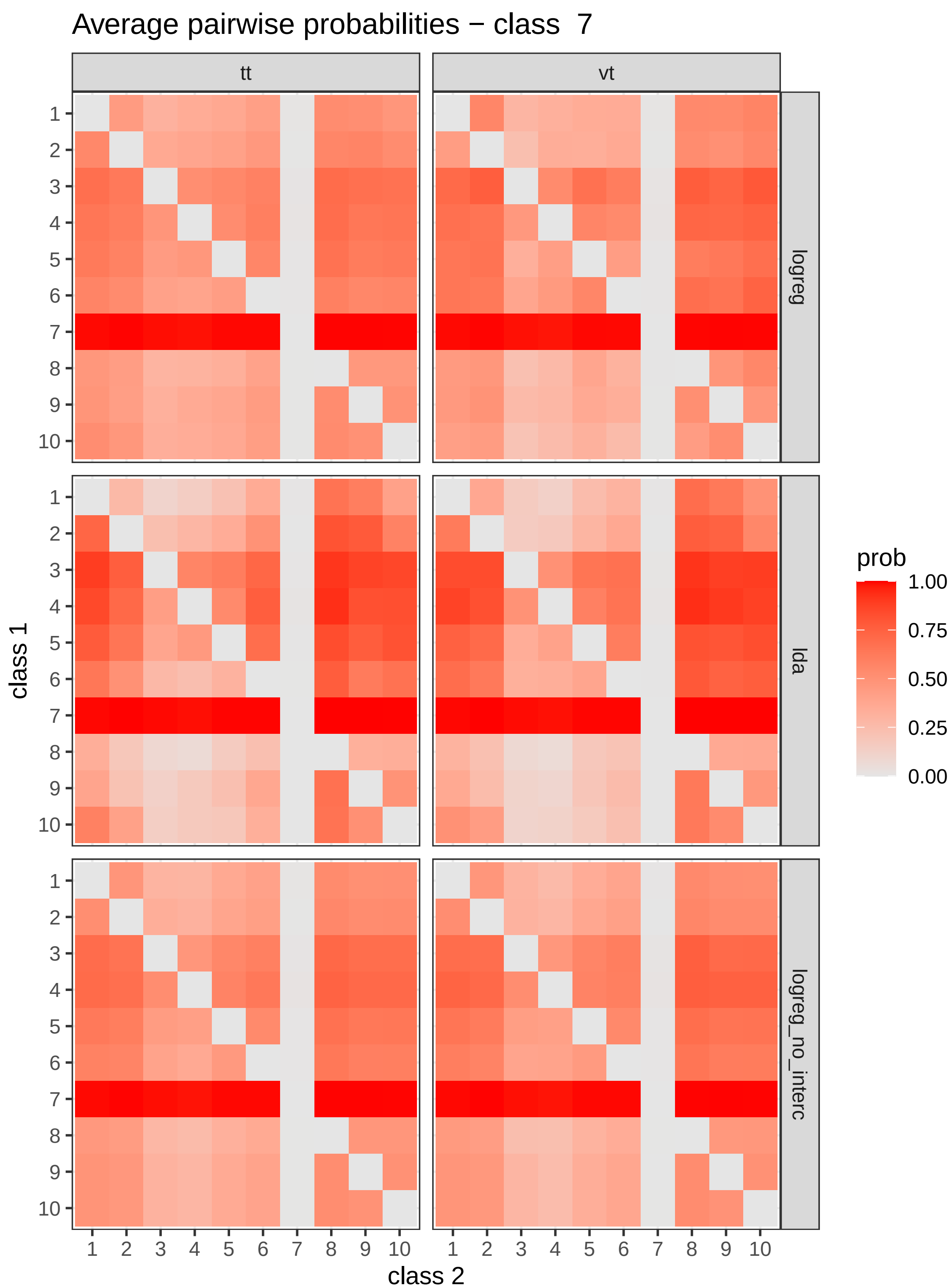
Average pairwise probabilities − class 3
tt
vt
logreg
lda
logreg_no_interc
class 1
class 2
prob
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10

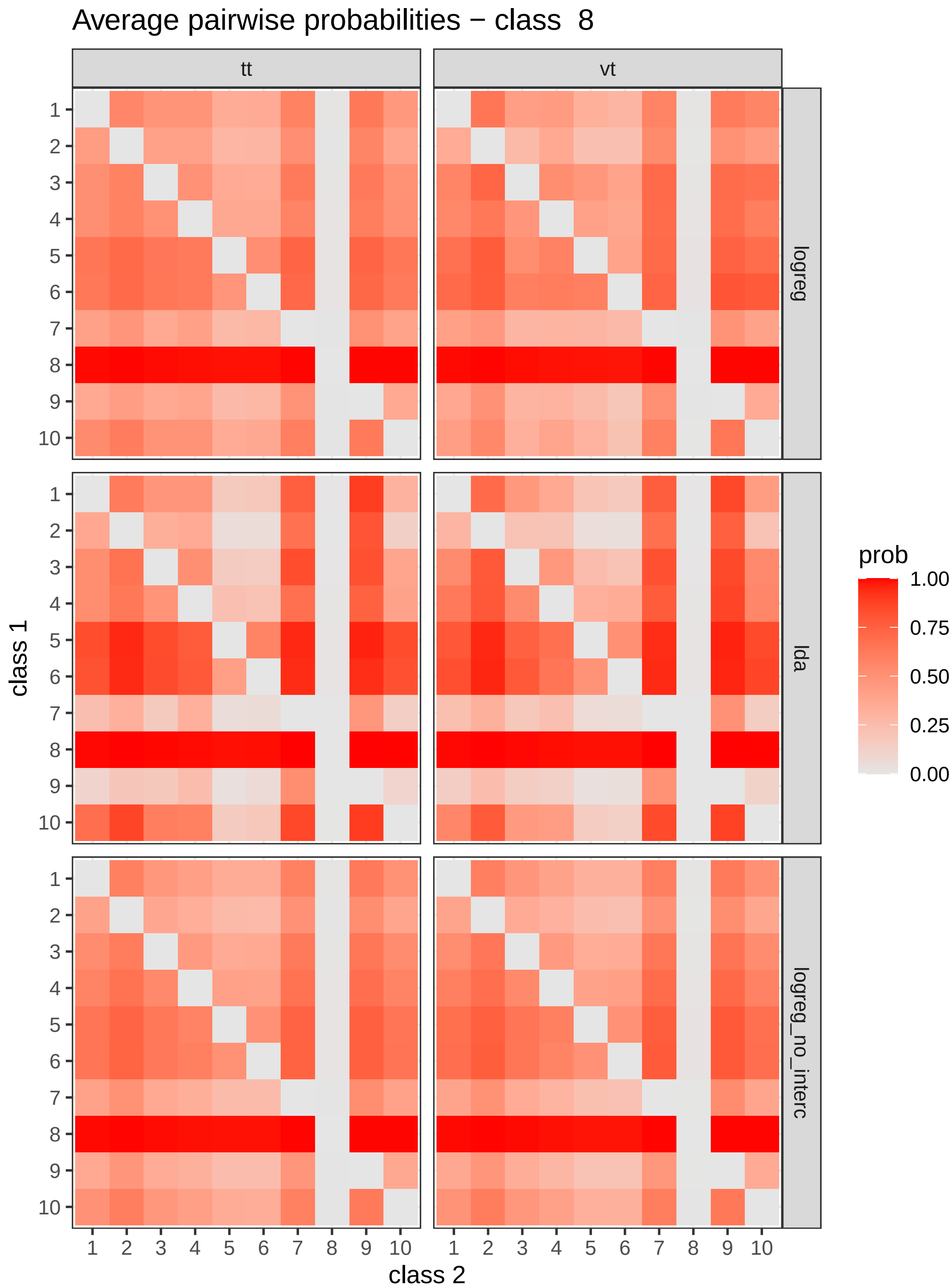

Average pairwise probabilities − class 4
tt
vt
logreg
lda
logreg_no_interc
class 1
class 2
prob
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10


Average pairwise probabilities − class 5
tt
vt
logreg
lda
logreg_no_interc
class 1
class 2
1
2
3
4
5
6
7
8
9
10
prob
1.00
0.75
0.50
0.25
0.00

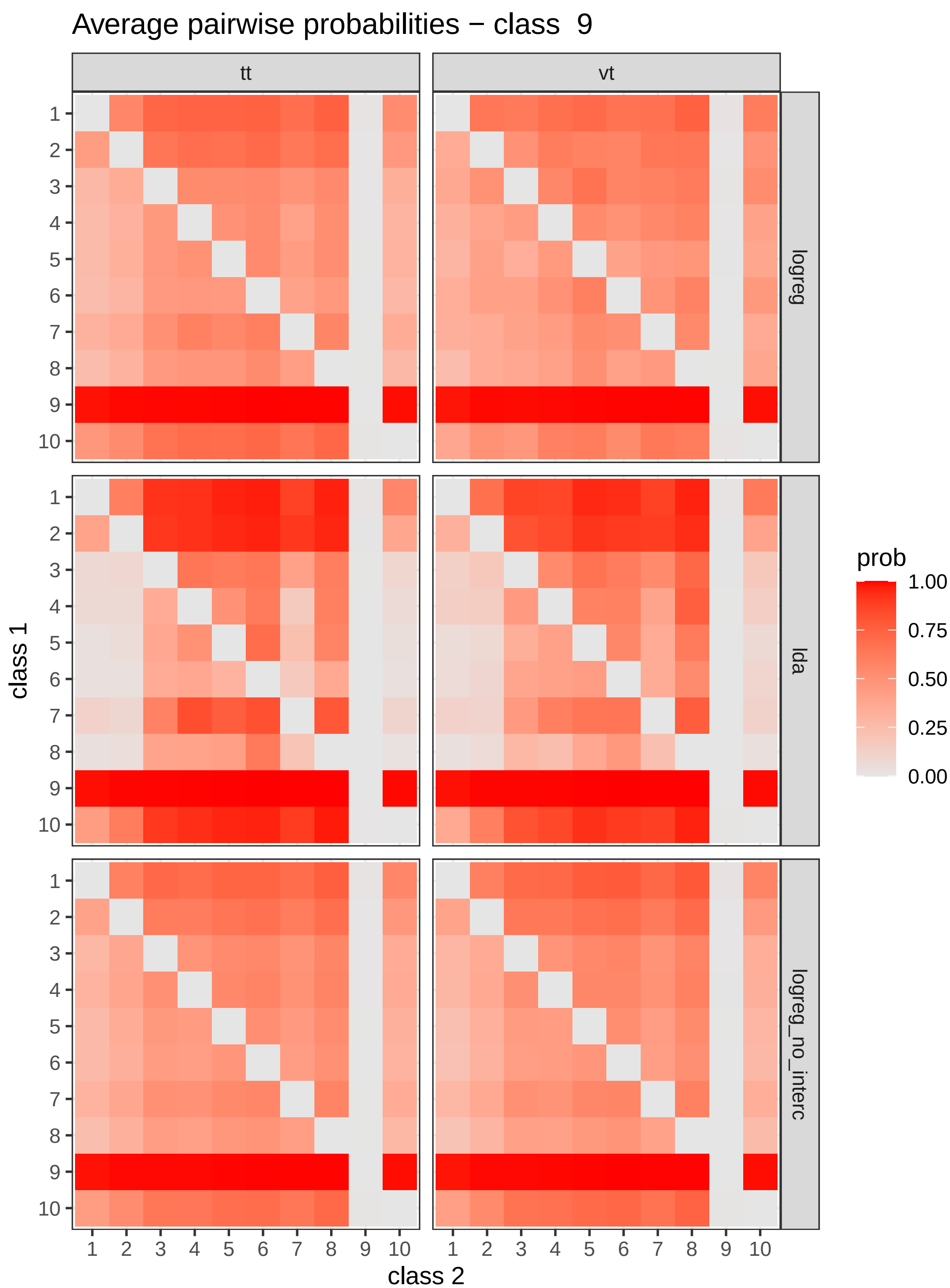


Average pairwise probabilities − class 6
tt
vt
logreg
lda
logreg_no_interc
class 1
class 2
prob
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10

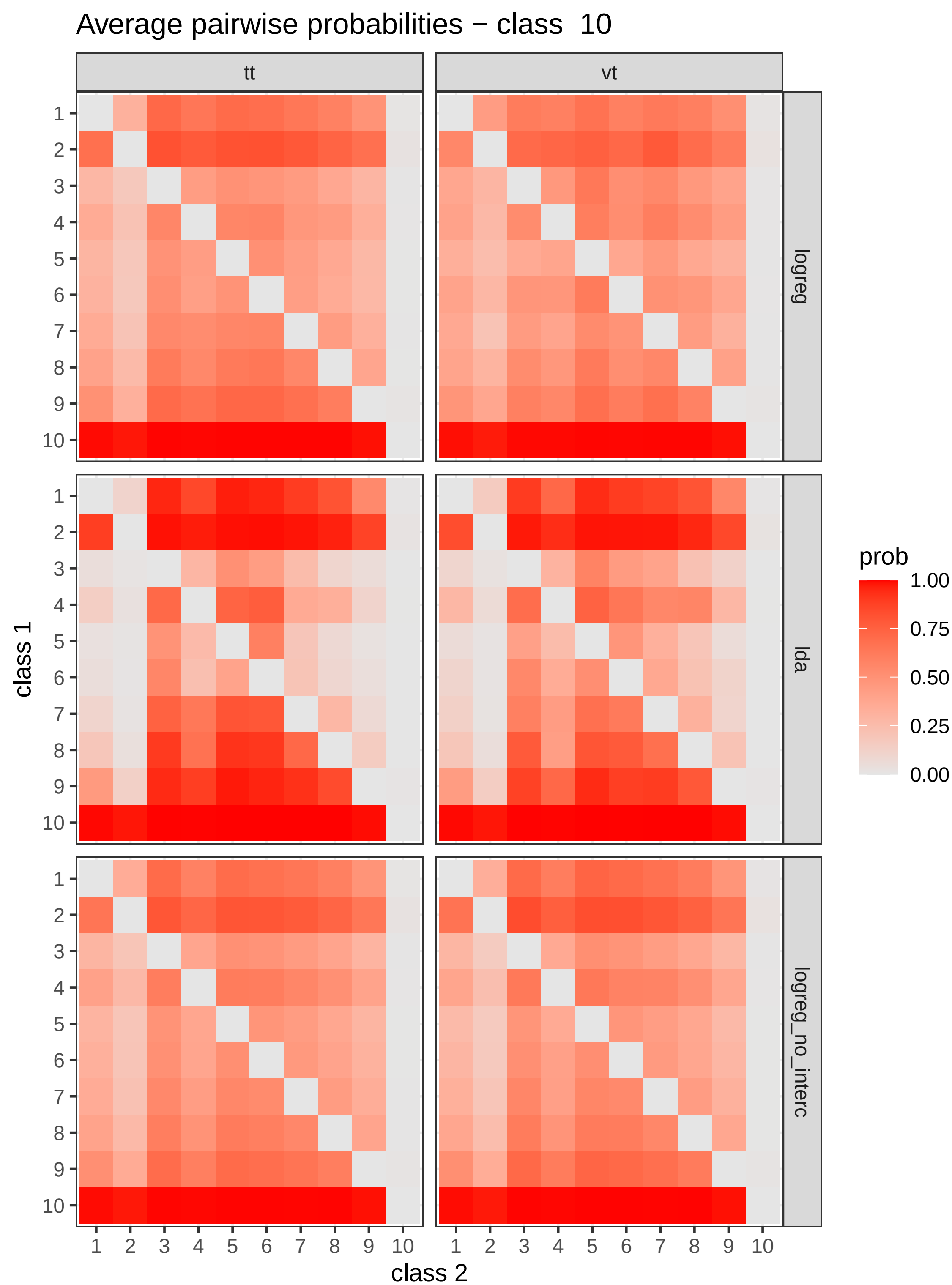

Average pairwise probabilities − class 7
tt
vt
logreg
lda
logreg_no_interc
class 1
class 2
prob
1.00
0.75
0.50
0.25
0.00



Average pairwise probabilities − class 8
tt
vt
logreg
lda
logreg_no_interc
class 1
class 2
1
2
3
4
5
6
7
8
9
10
prob
1.00
0.75
0.50
0.25
0.00



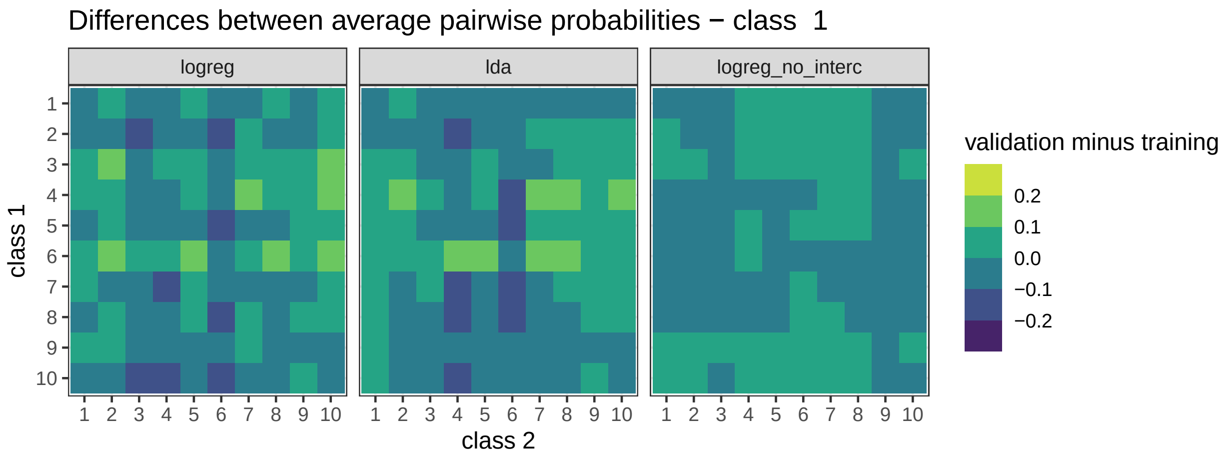
Average pairwise probabilities − class 9
tt
vt
logreg
lda
logreg_no_interc
class 1
class 2
prob
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10

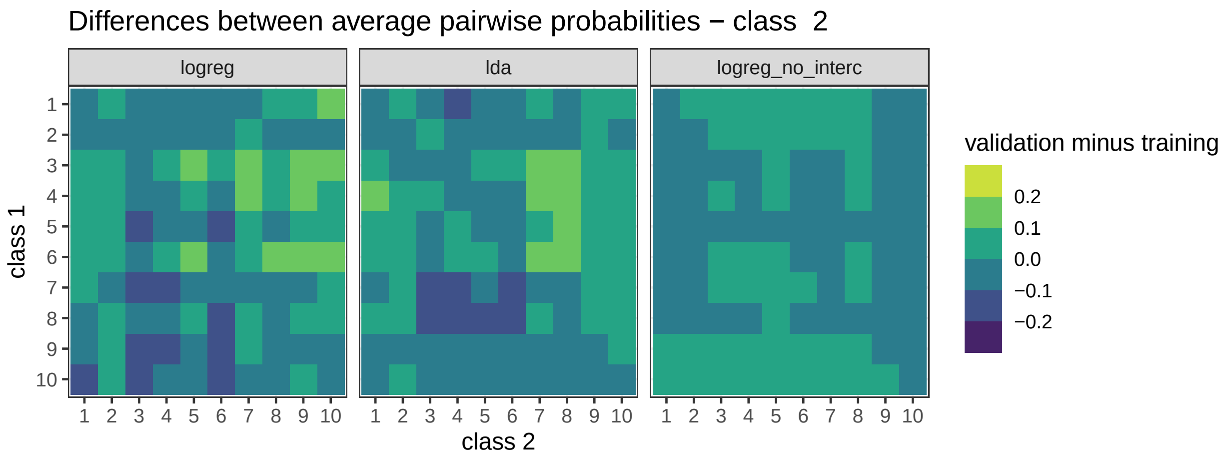


lda has more cells with false high values than logreg.

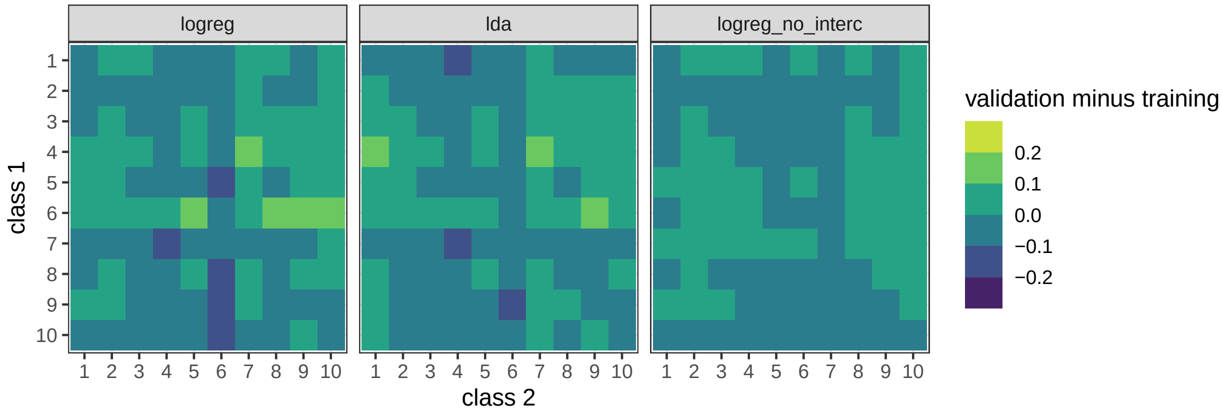
```
for (cls in 1:classes)
{
  cur_class_Rs <- df_aggr_Rs_diff %>% filter(class == cls)
  plot_cls <-  ggplot(cur_class_Rs, aes(x = class2, y = class1)) +
```

```
    geom_raster(aes(fill=val_min_train)) +
    facet_wrap(~combining_method) +
    scale_fill_binned(type="viridis", limits=c(-0.3, 0.3), name="validation minus training") +
    scale_y_discrete(limits=rev) +
    labs(x="class 2", y="class 1", title=paste("Differences between average pairwise probabilities - cla
    theme_bw()

  print(plot_cls)
}
```

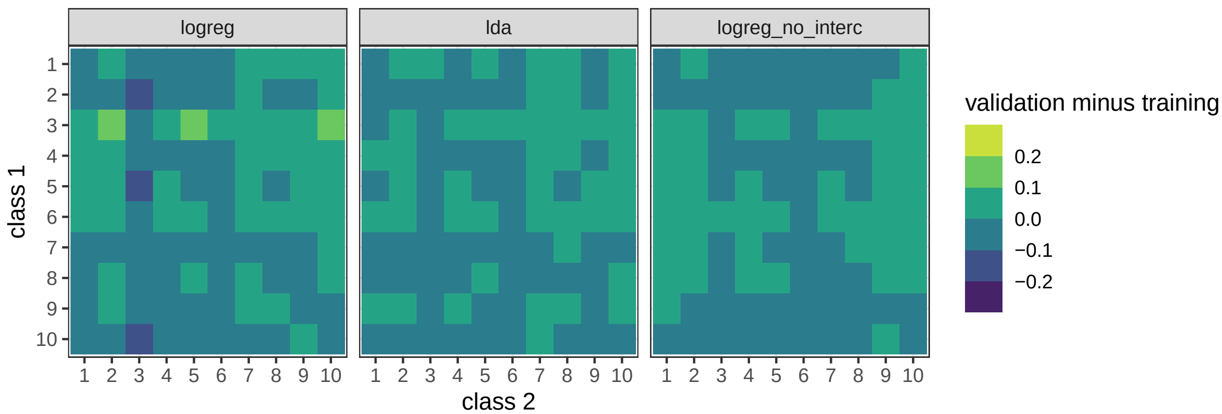




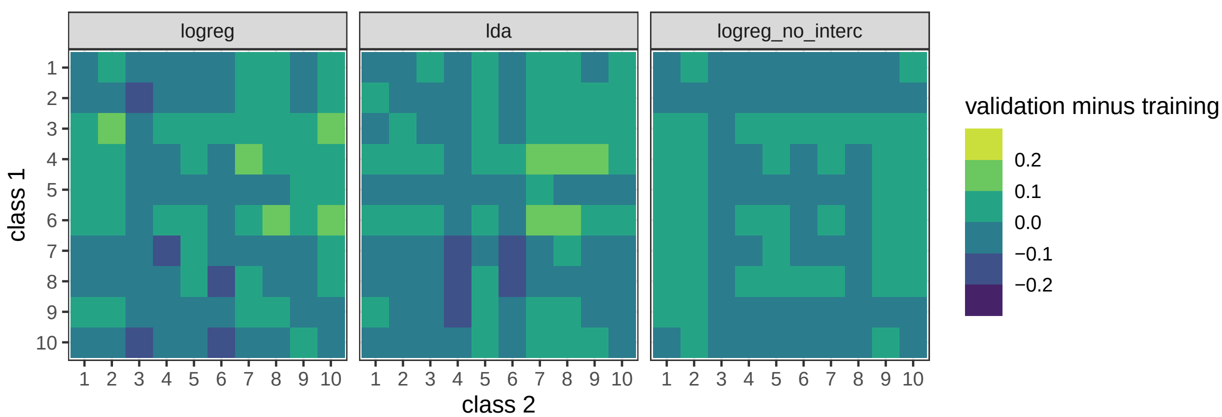
Differences between average pairwise probabilities − class 3



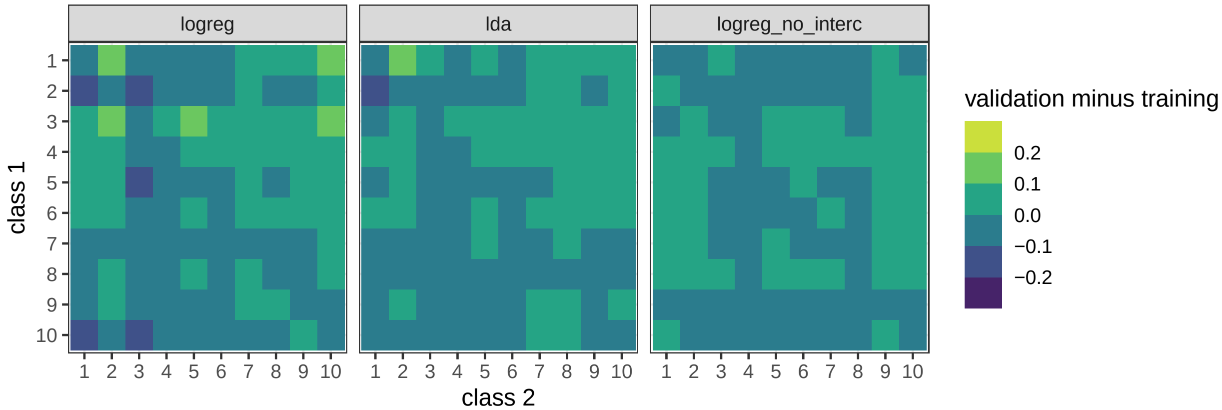
Differences between average pairwise probabilities − class 4



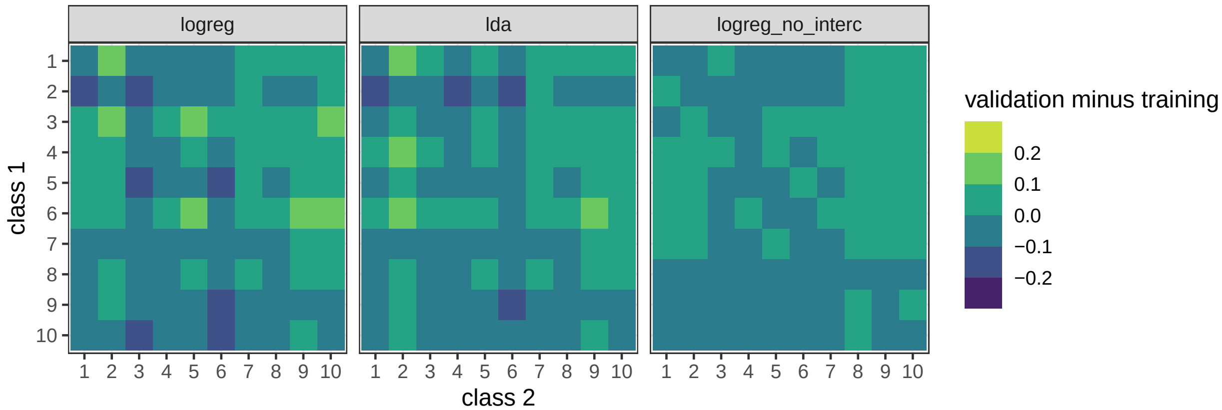
Differences between average pairwise probabilities − class 5



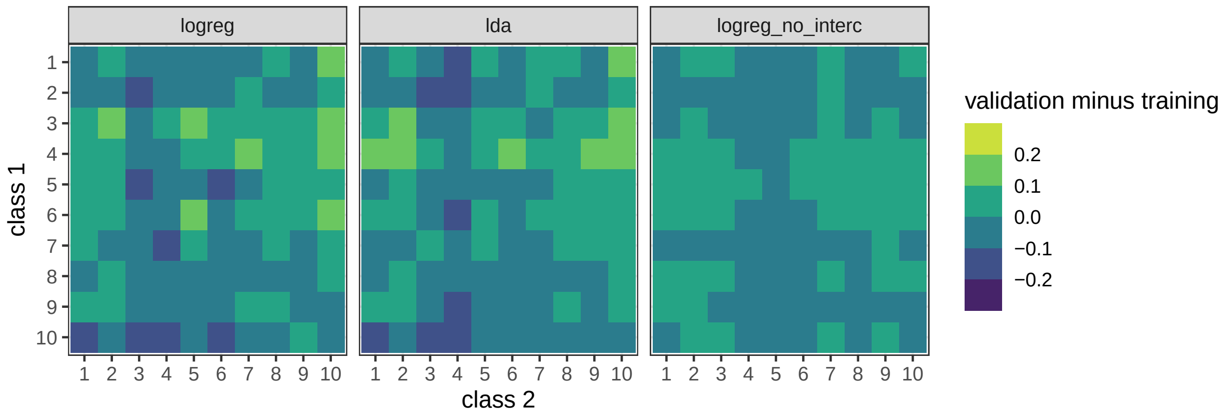
Differences between average pairwise probabilities − class 6

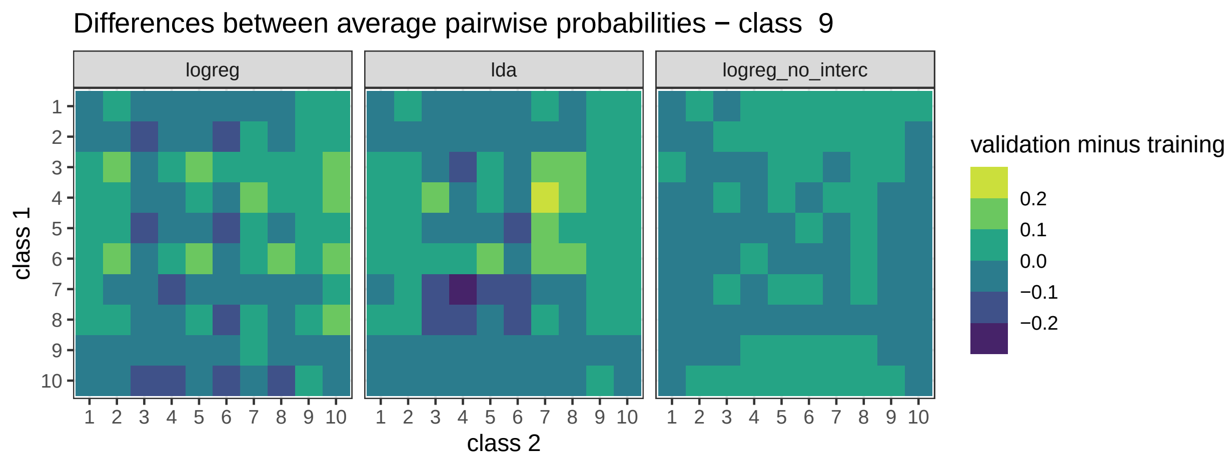


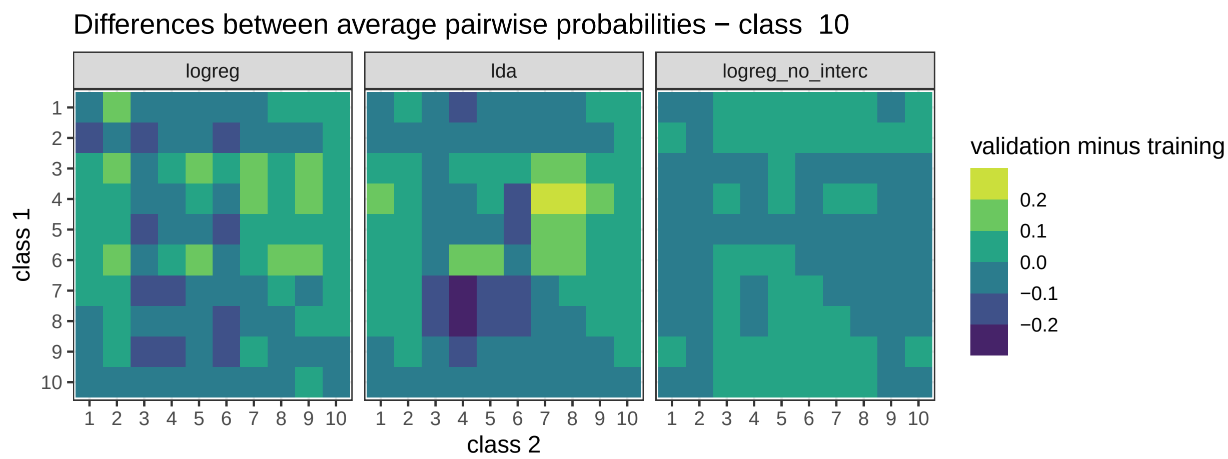
Differences between average pairwise probabilities − class 7



Differences between average pairwise probabilities − class 8



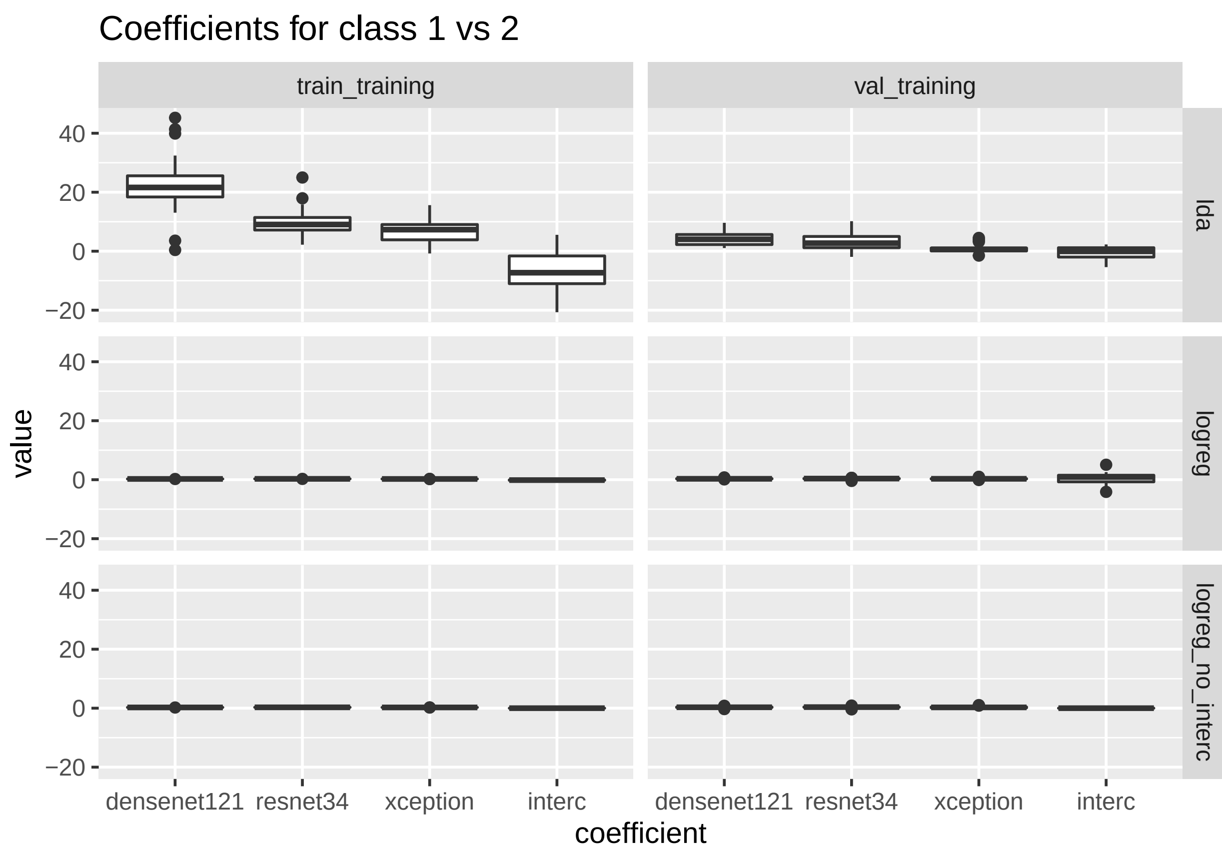


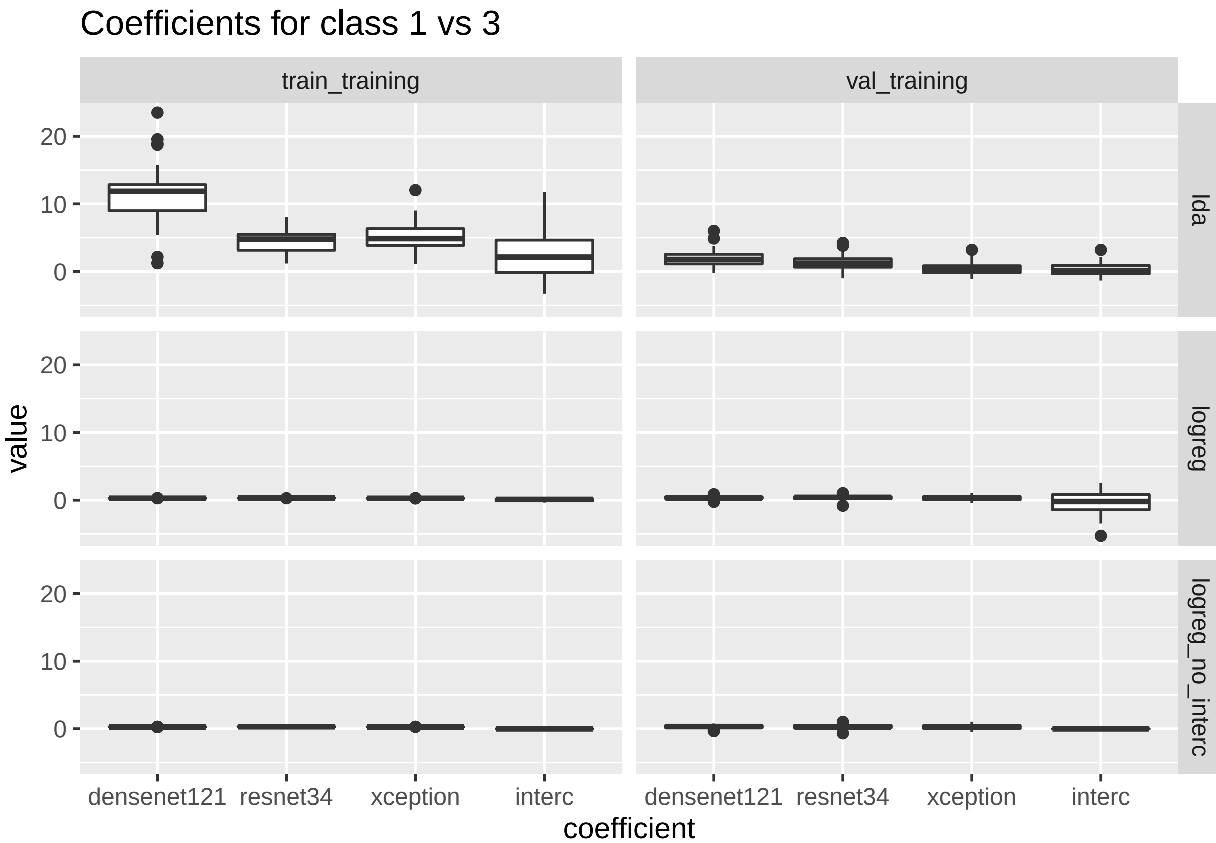


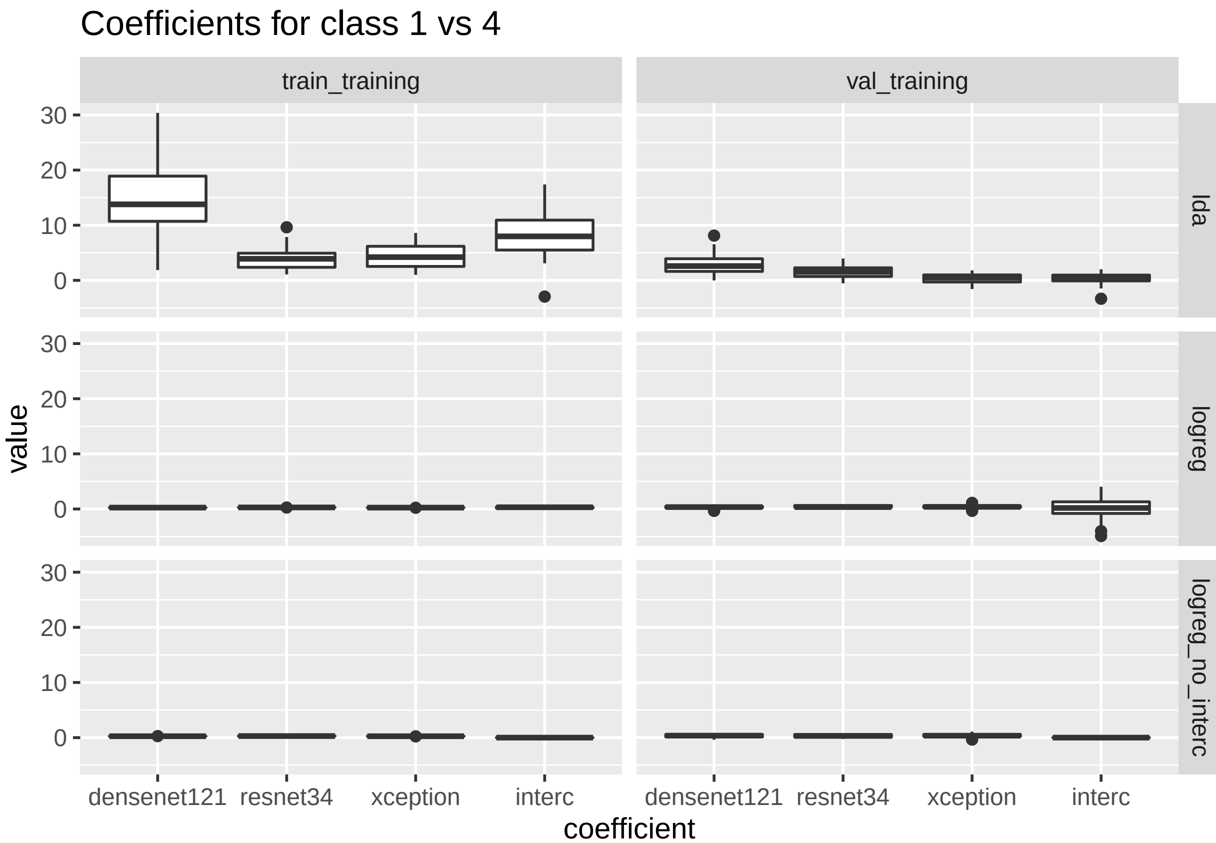
Logistic regression without intercept has lower differences between tt and vt R matrices than other combining methods.

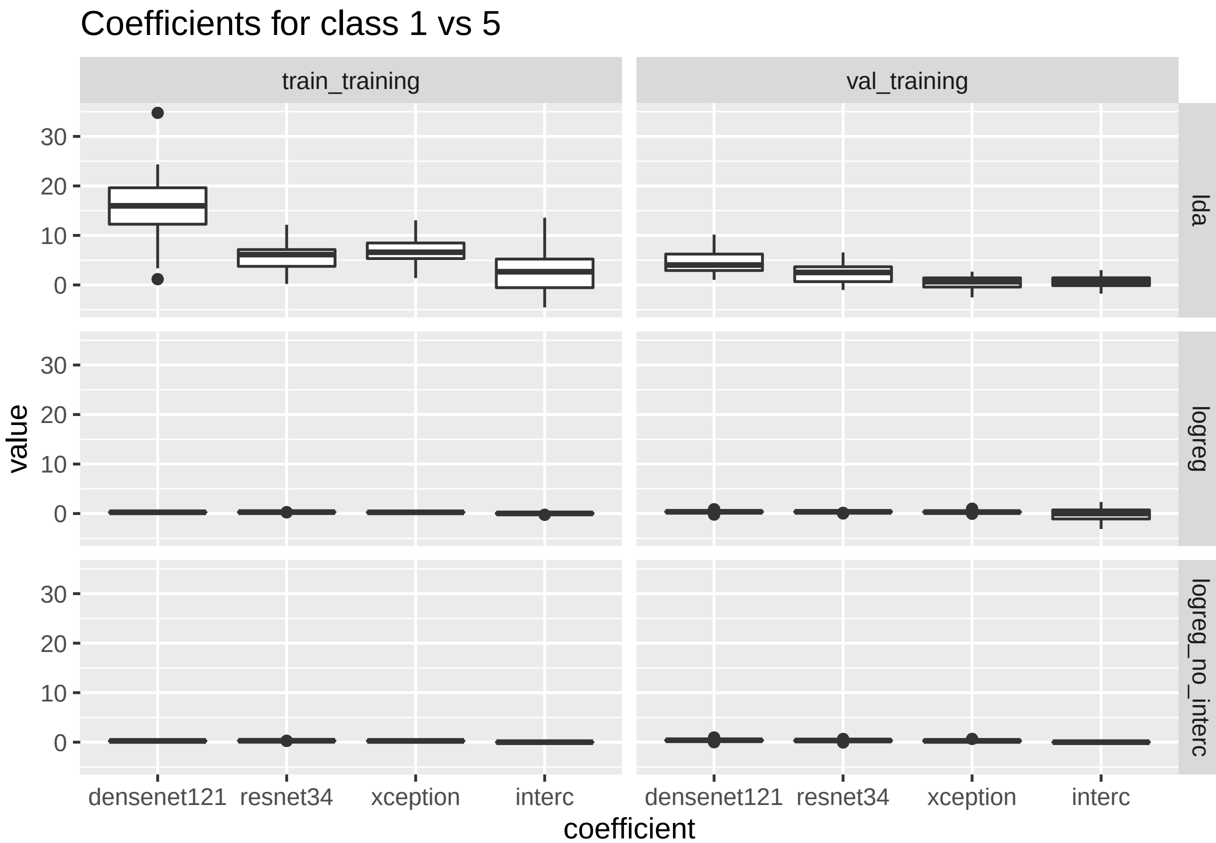
```
combiner_coefs <- load_combiner_coefs(base_dir, repls)
```

```
for (cl1 in 1:(classes - 1))
{
  for (cl2 in (cl1 + 1):classes)
  {
    cur_plt <- combiner_coefs %>% filter(class1 == cl1 & class2 == cl2) %>% ggplot() + geom_boxplot(aes
      facet_grid(cols=vars(train_type), rows=vars(combining_method)) + ggtitle(paste("Coefficients for c
    print(cur_plt)
  }
}
```

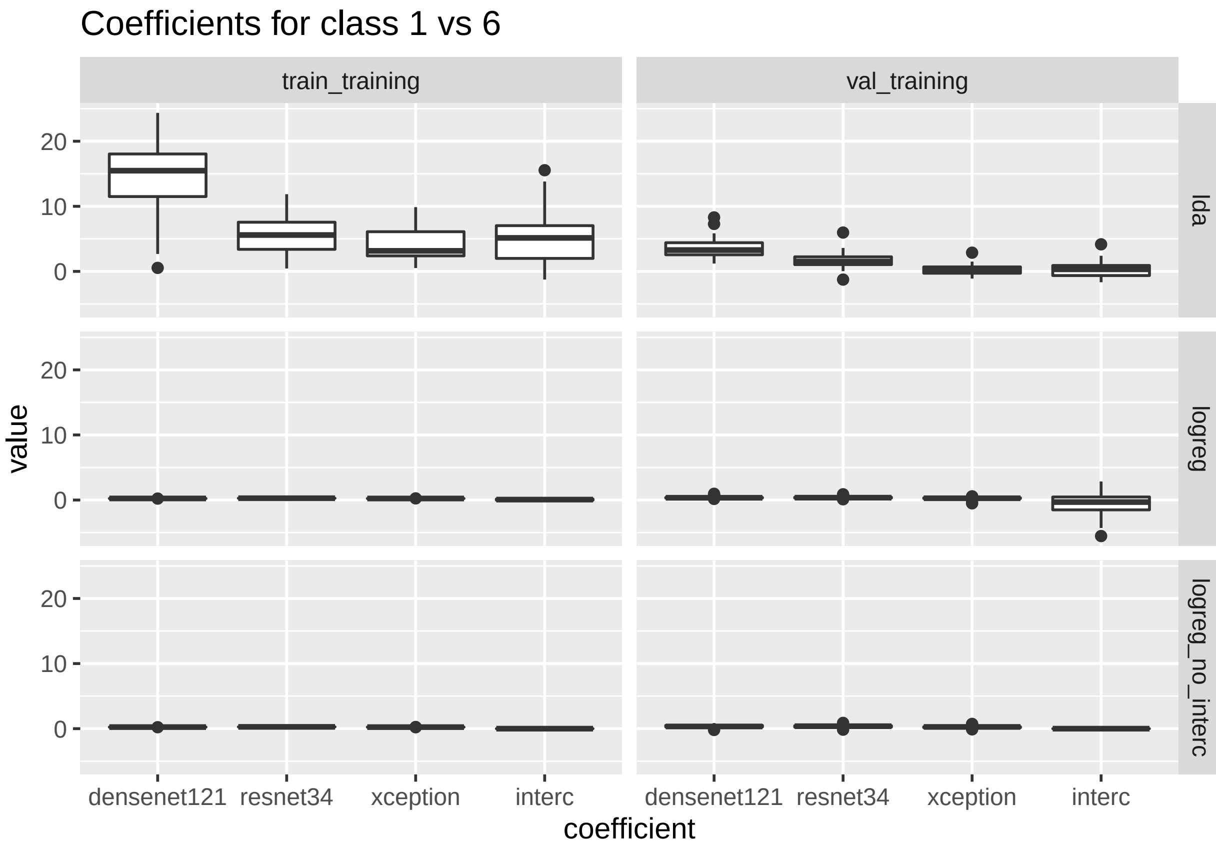

Coefficients for class 1 vs 2
train_training
val_training
lda
logreg
logreg_no_interc
value
40
20
0
−20
densenet121
resnet34
xception
interc
coefficient

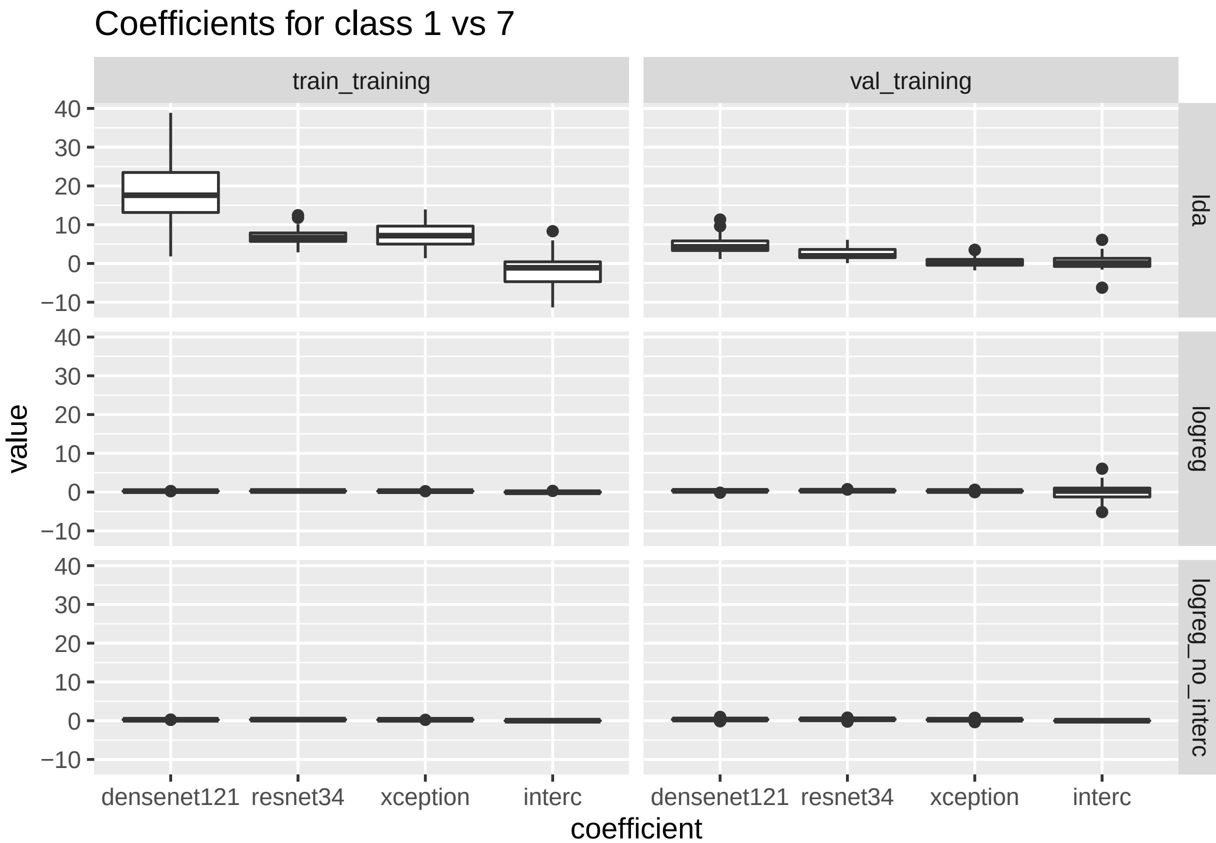

Coefficients for class 1 vs 3
train_training
val_training
lda
logreg
logreg_no_interc
value
20
10
0
densenet121
resnet34
xception
interc
coefficient

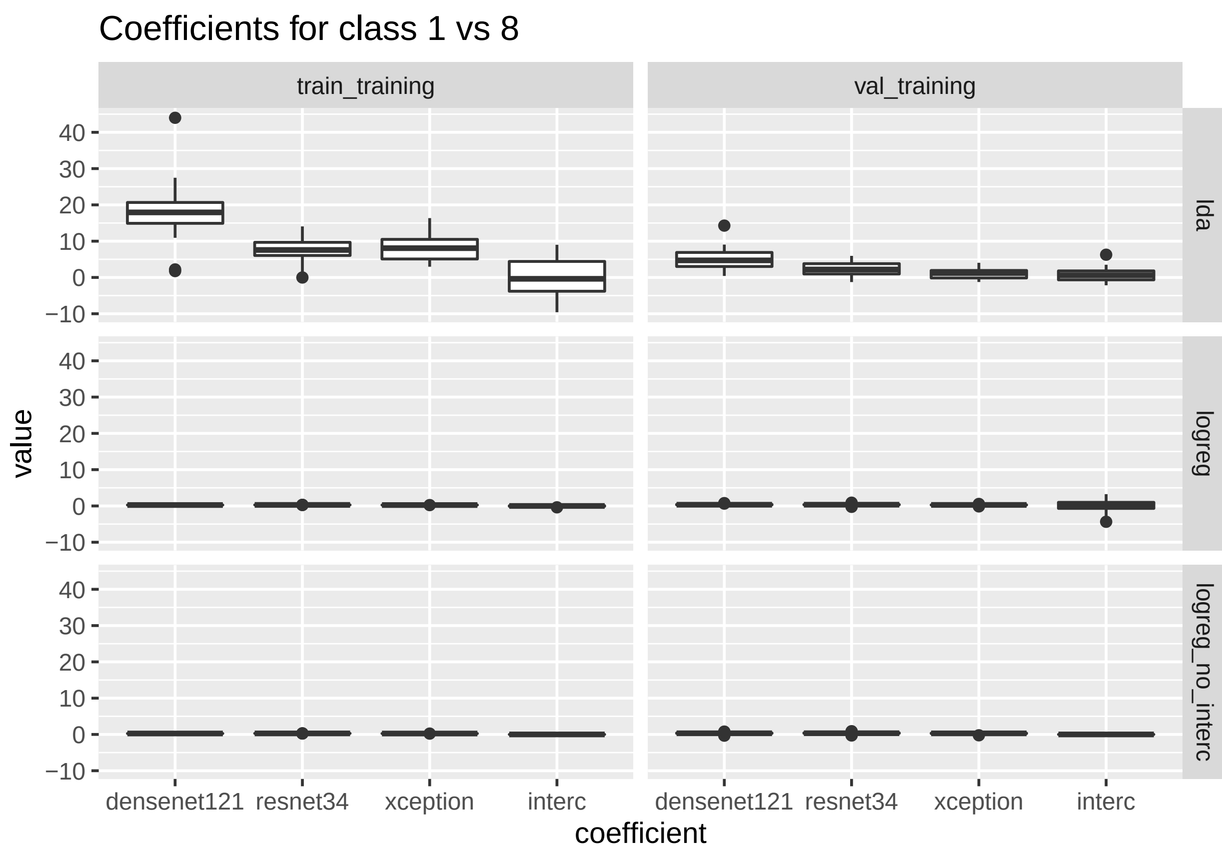

Coefficients for class 1 vs 4
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
densenet121
resnet34
xception
interc
coefficient

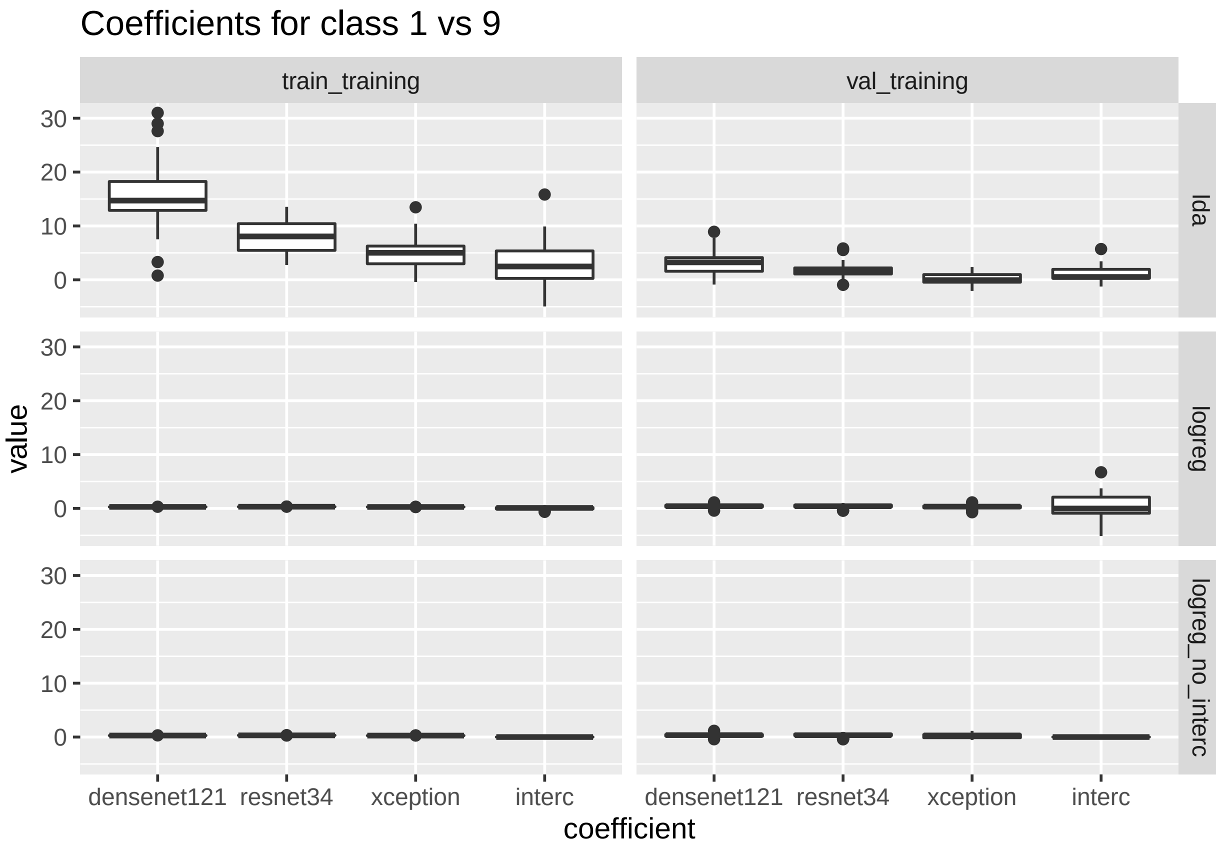


Coefficients for class 1 vs 5
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
densenet121
resnet34
xception
interc
coefficient

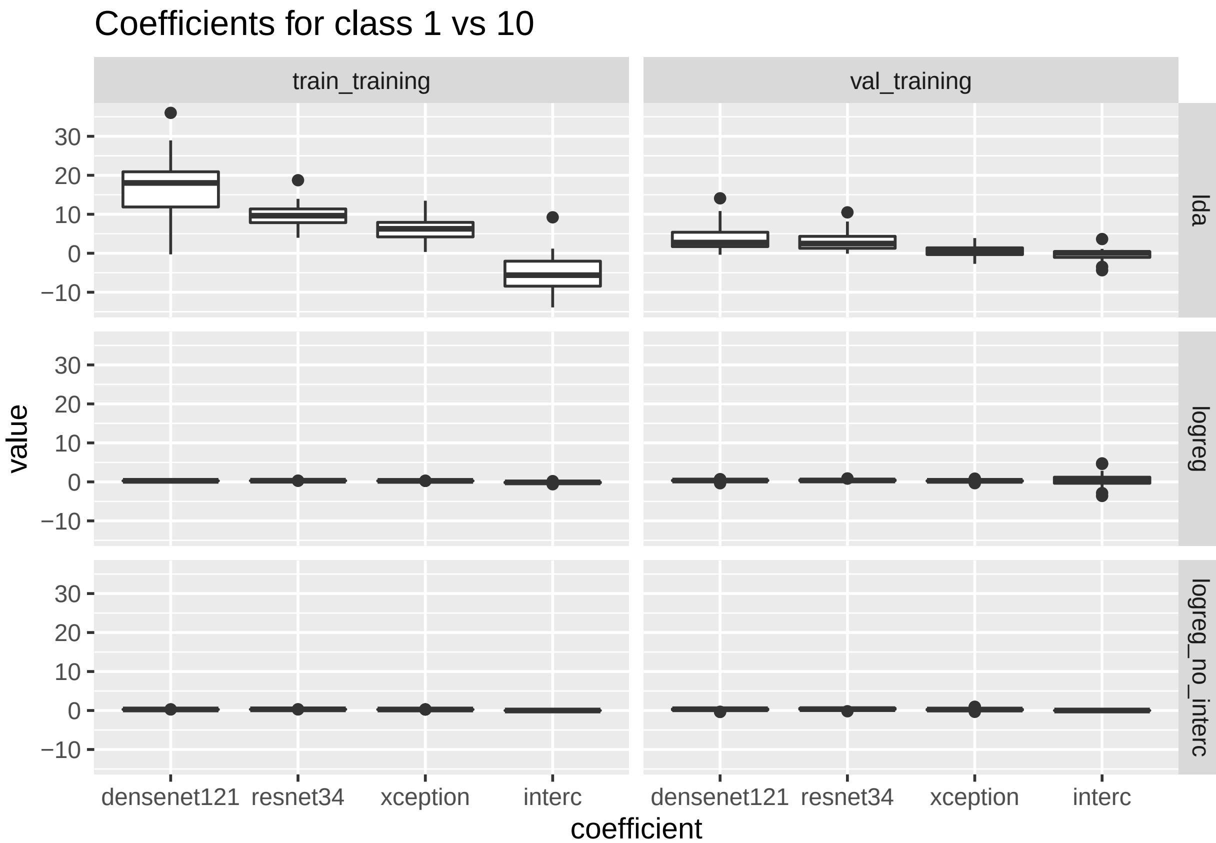

Coefficients for class 1 vs 6
train_training
val_training
lda
logreg
logreg_no_interc
value
20
10
0
densenet121
resnet34
xception
interc
coefficient

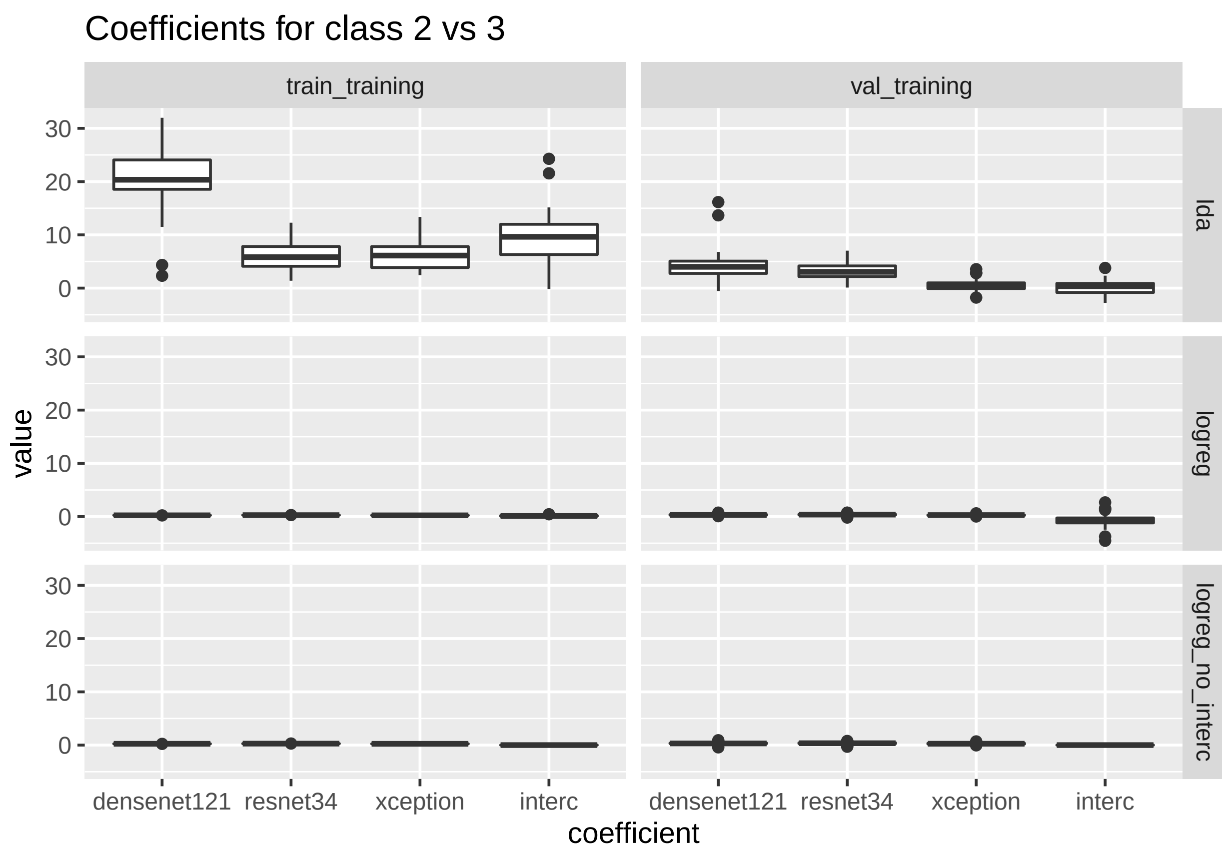

Coefficients for class 1 vs 7
train_training
val_training
lda
logreg
logreg_no_interc
value
40
30
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient


Coefficients for class 1 vs 8
train_training
val_training
lda
logreg
logreg_no_interc
value
40
30
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient

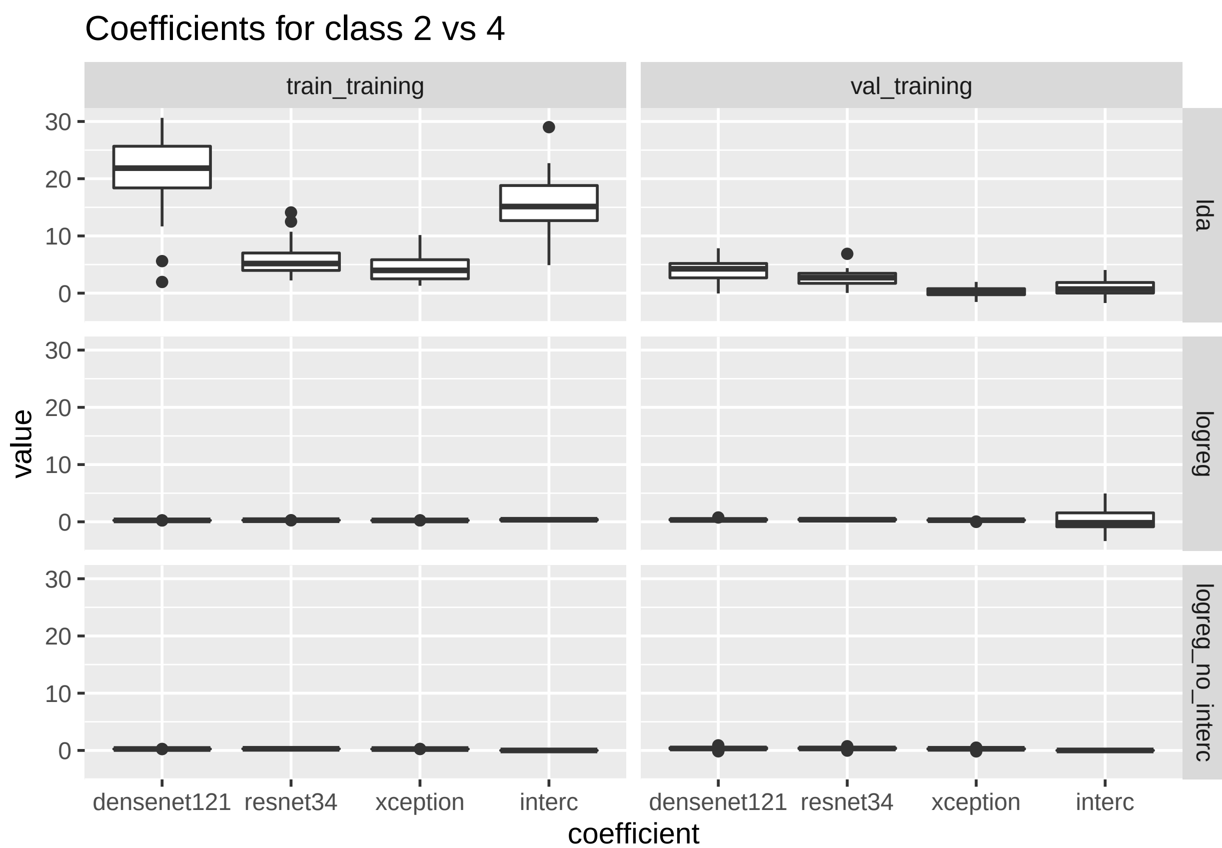


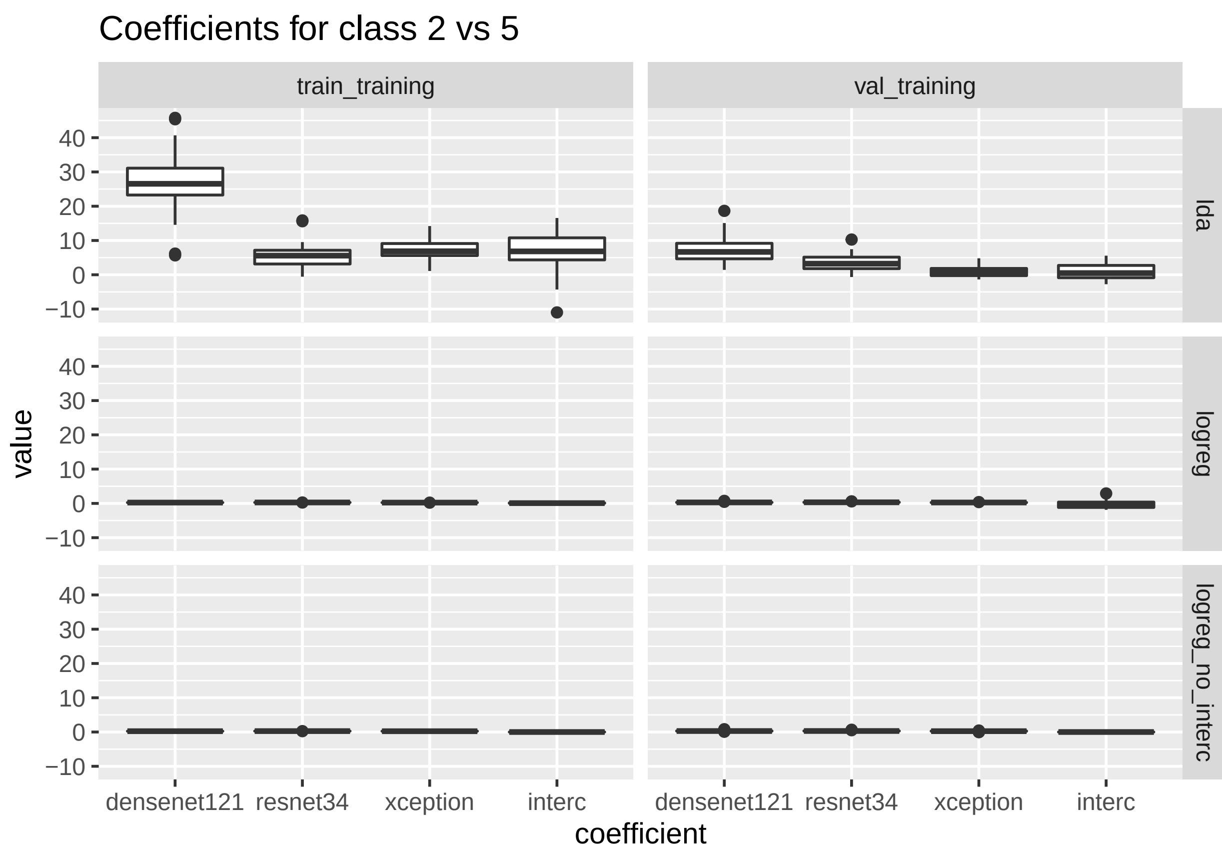
Coefficients for class 1 vs 9
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
densenet121
resnet34
xception
interc
coefficient


Coefficients for class 1 vs 10
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient

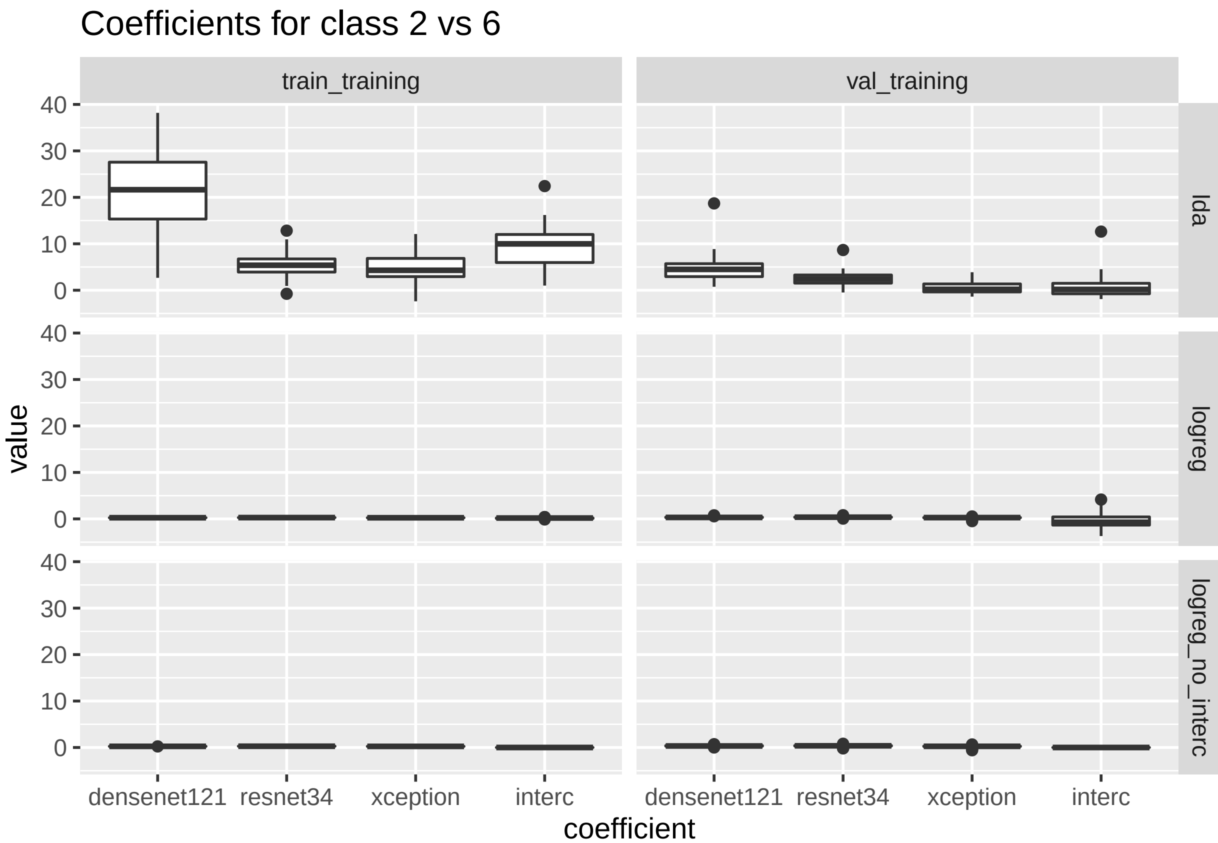


Coefficients for class 2 vs 3
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
densenet121
resnet34
xception
interc
coefficient

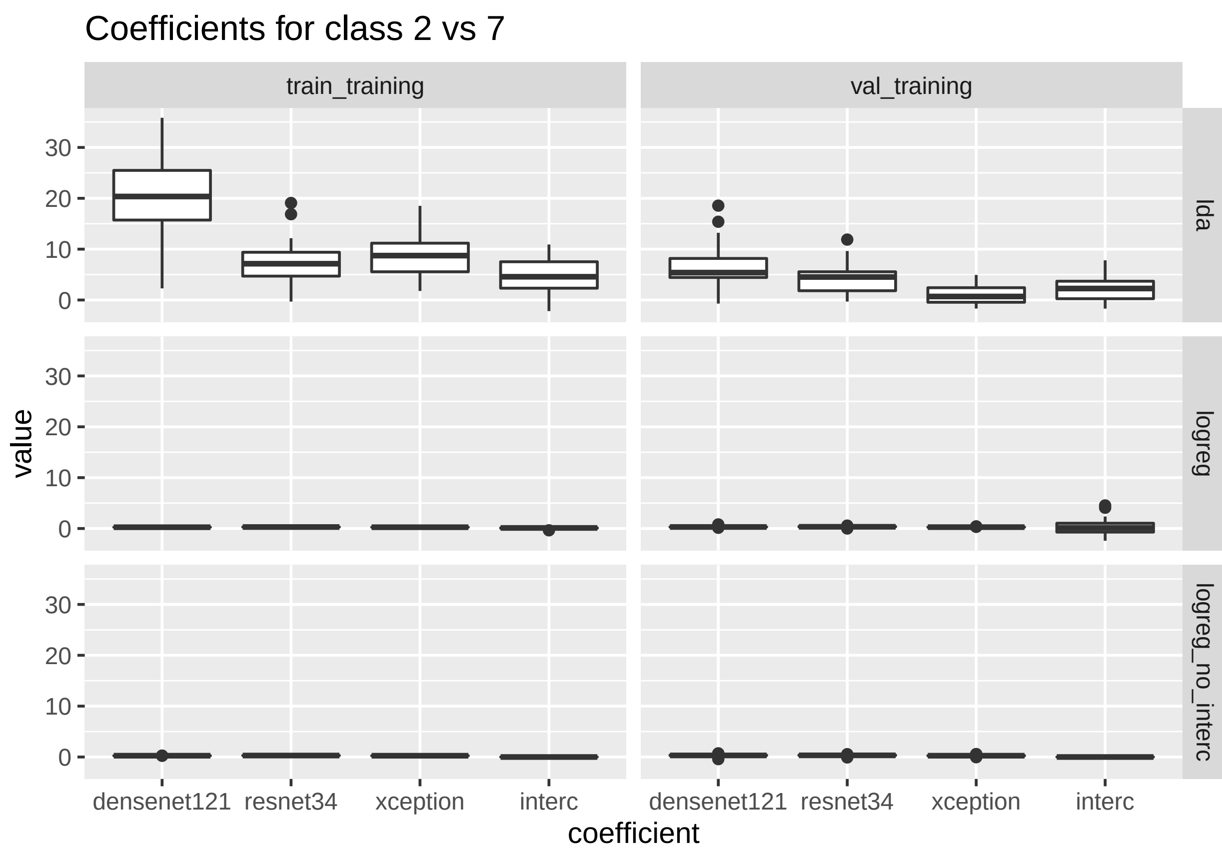

Coefficients for class 2 vs 4
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
densenet121
resnet34
xception
interc
coefficient



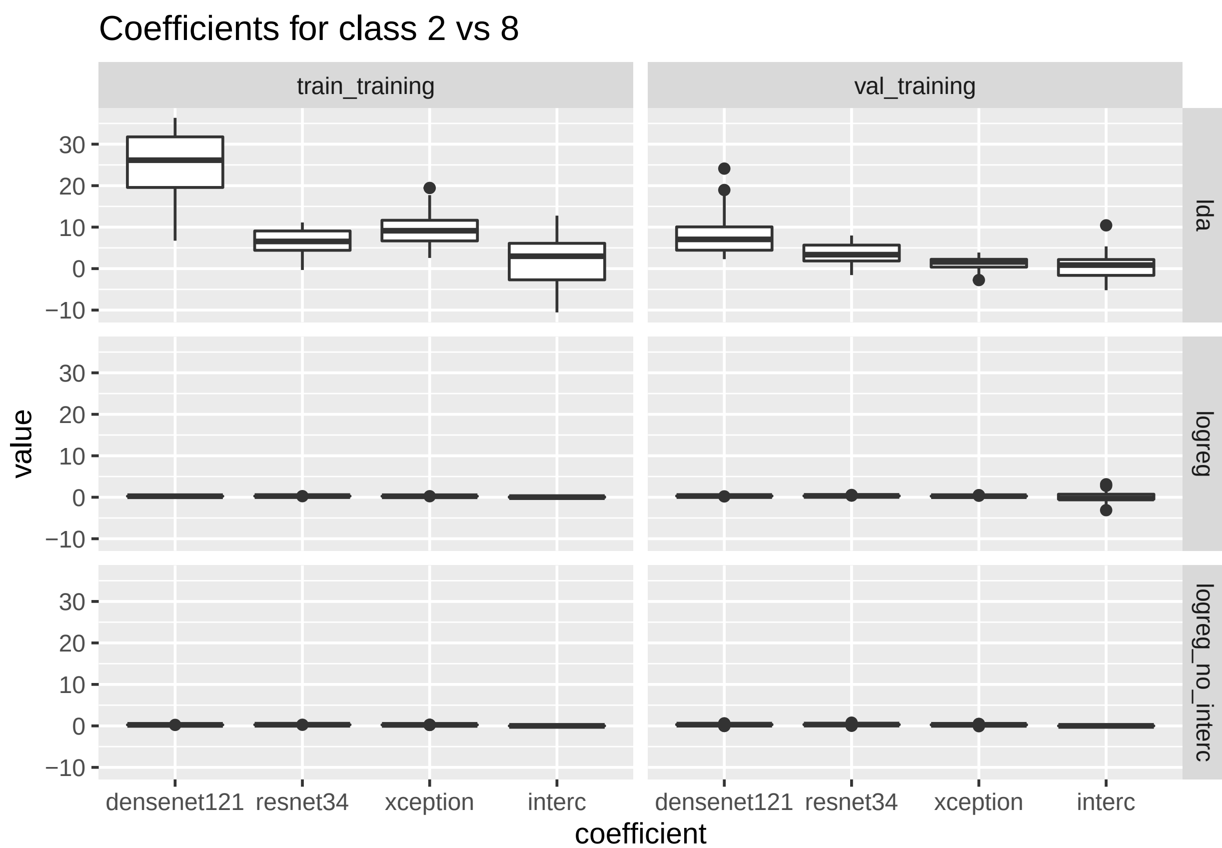
Coefficients for class 2 vs 5
train_training
val_training
lda
logreg
logreg_no_interc
value
40
30
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient



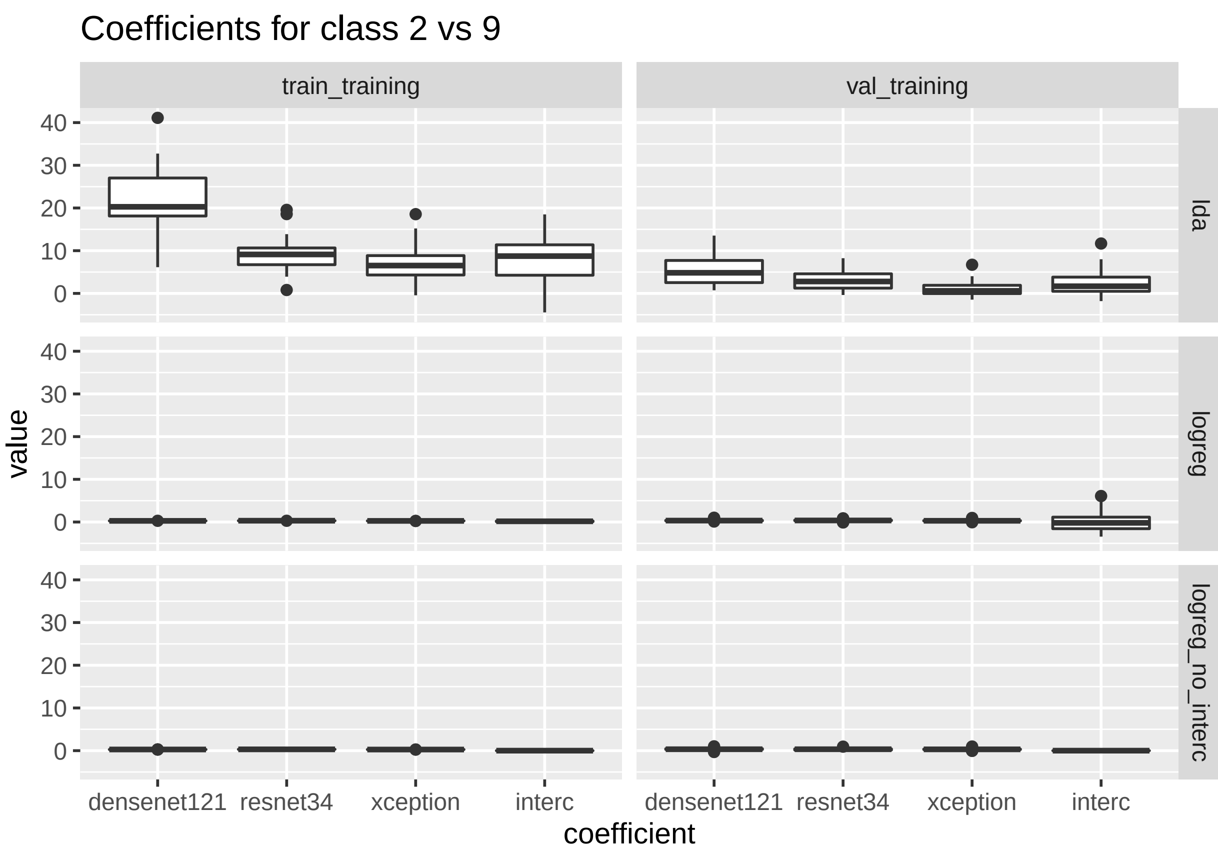
Coefficients for class 2 vs 6
train_training
val_training
lda
logreg
logreg_no_interc
value
40
30
20
10
0
densenet121
resnet34
xception
interc
coefficient

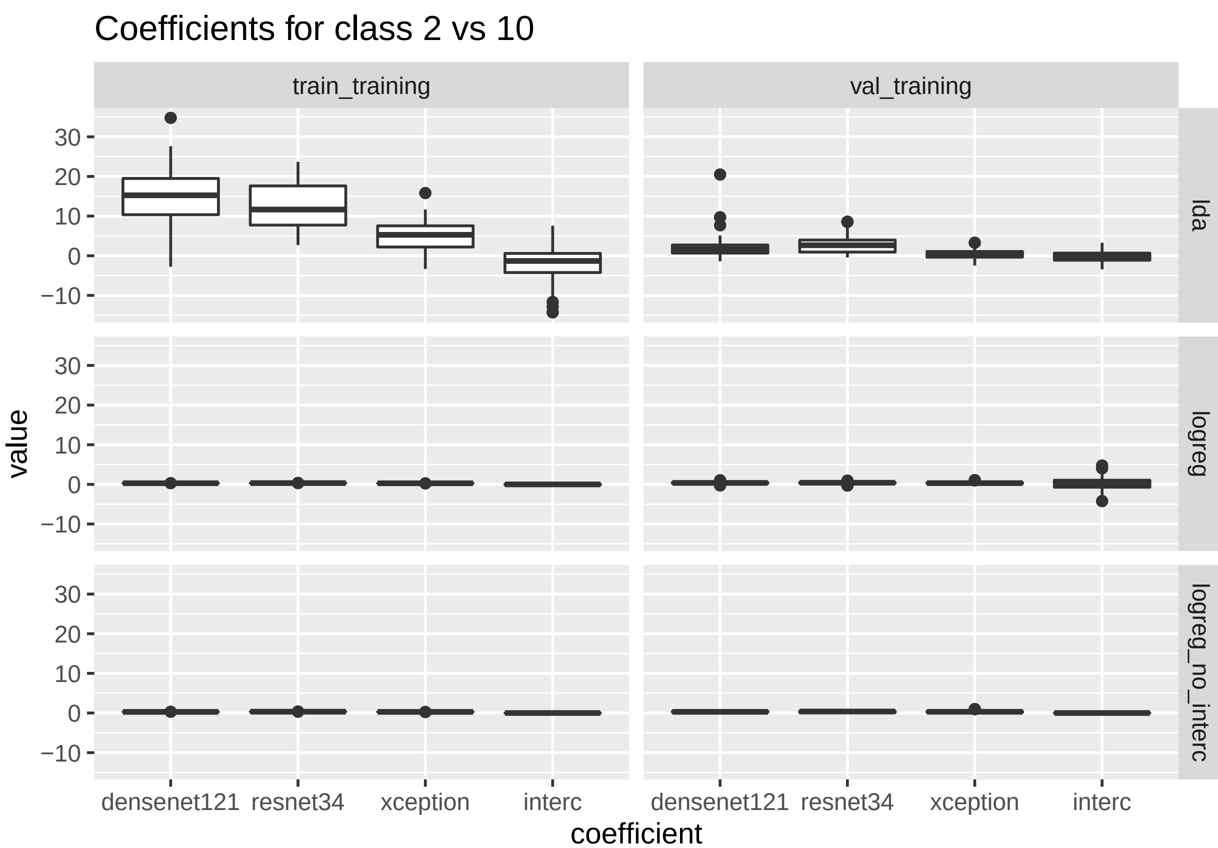


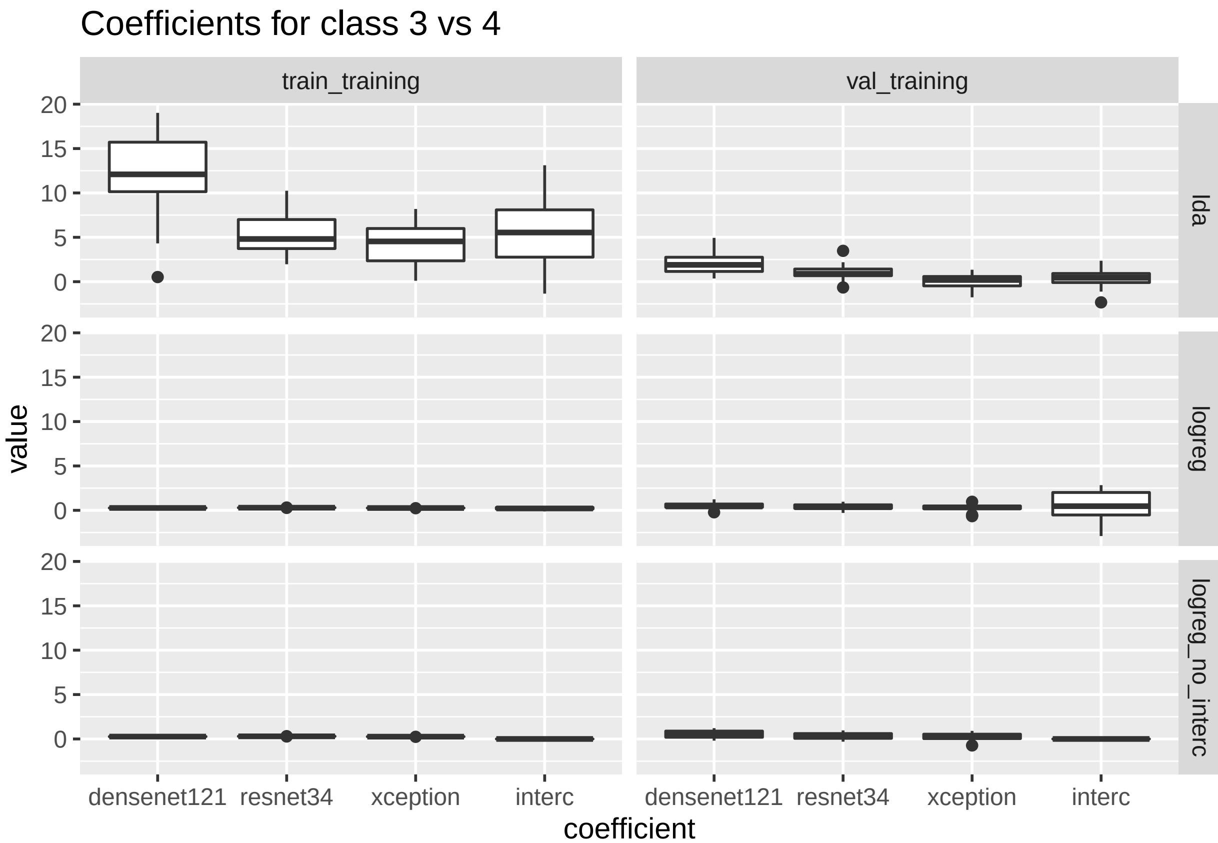
Coefficients for class 2 vs 7
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
densenet121
resnet34
xception
interc
coefficient

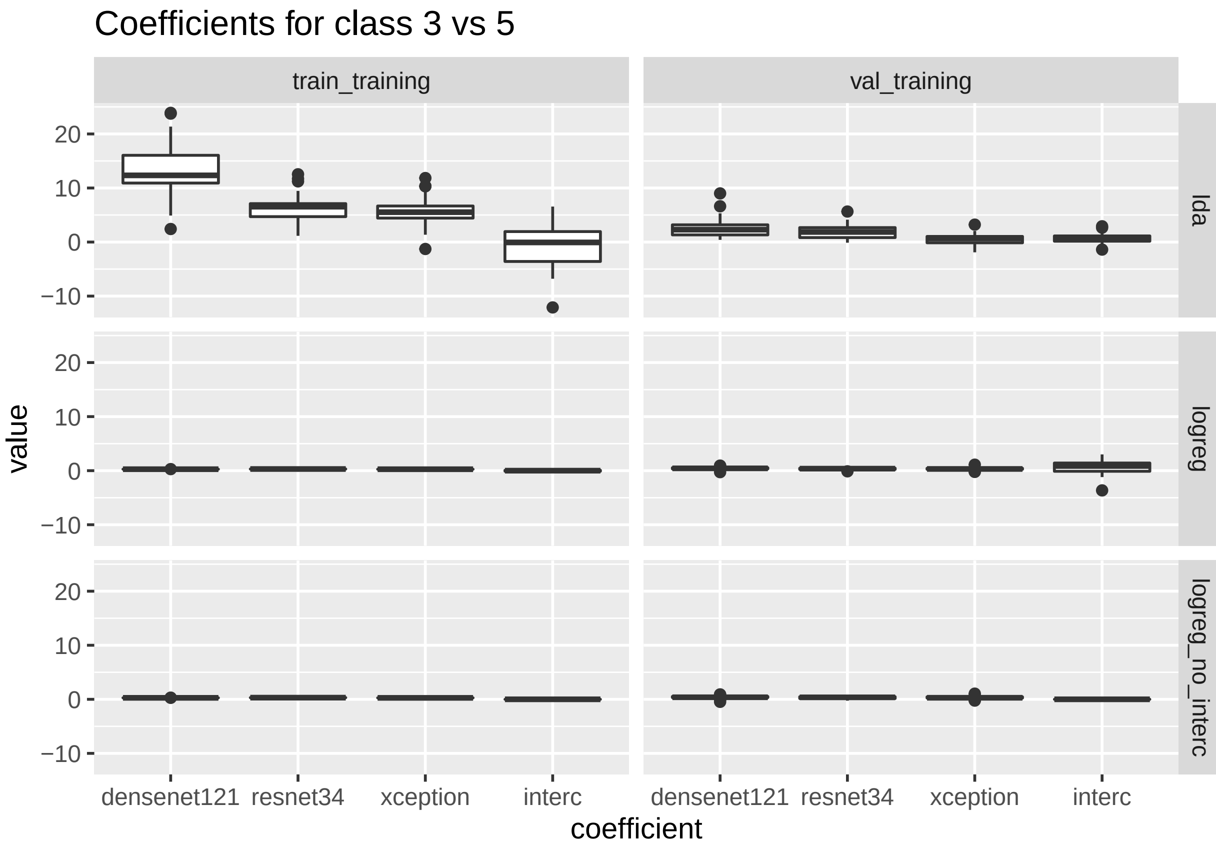


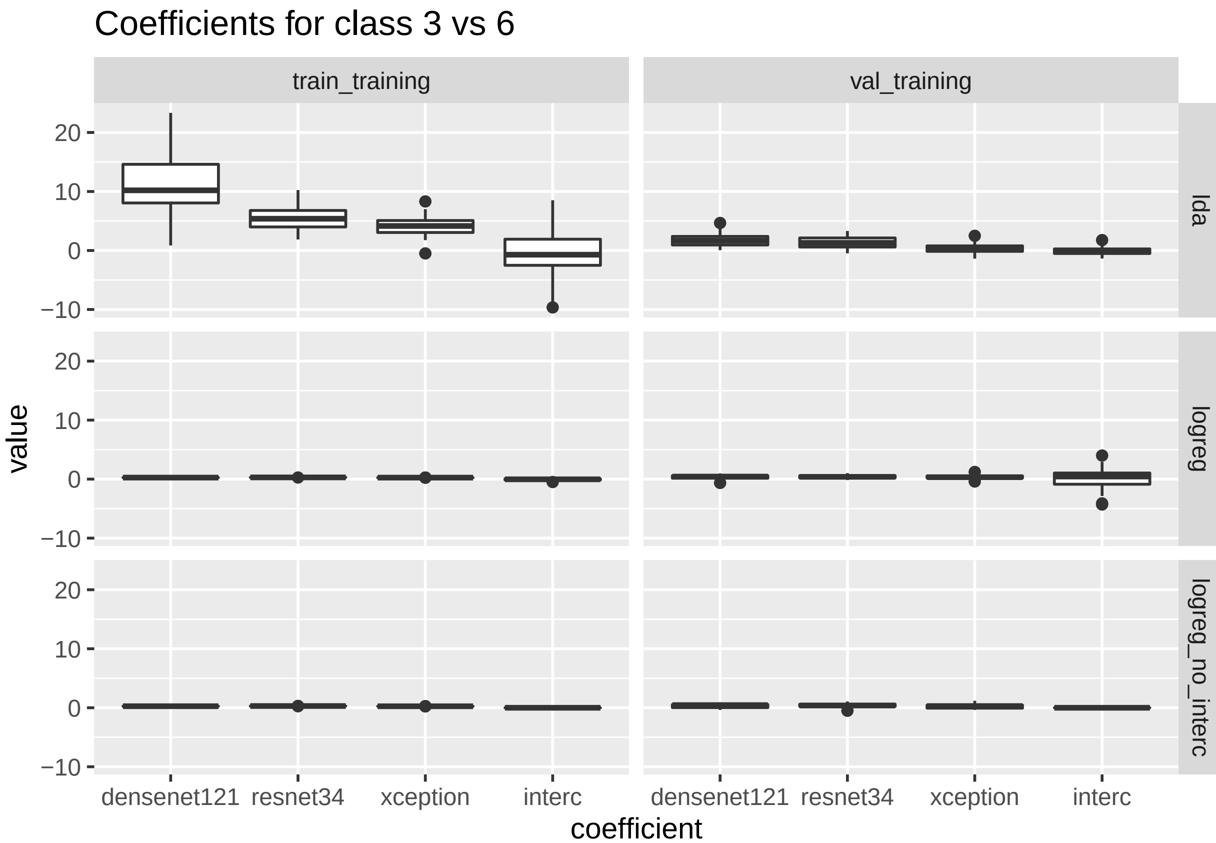
Coefficients for class 2 vs 8
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient

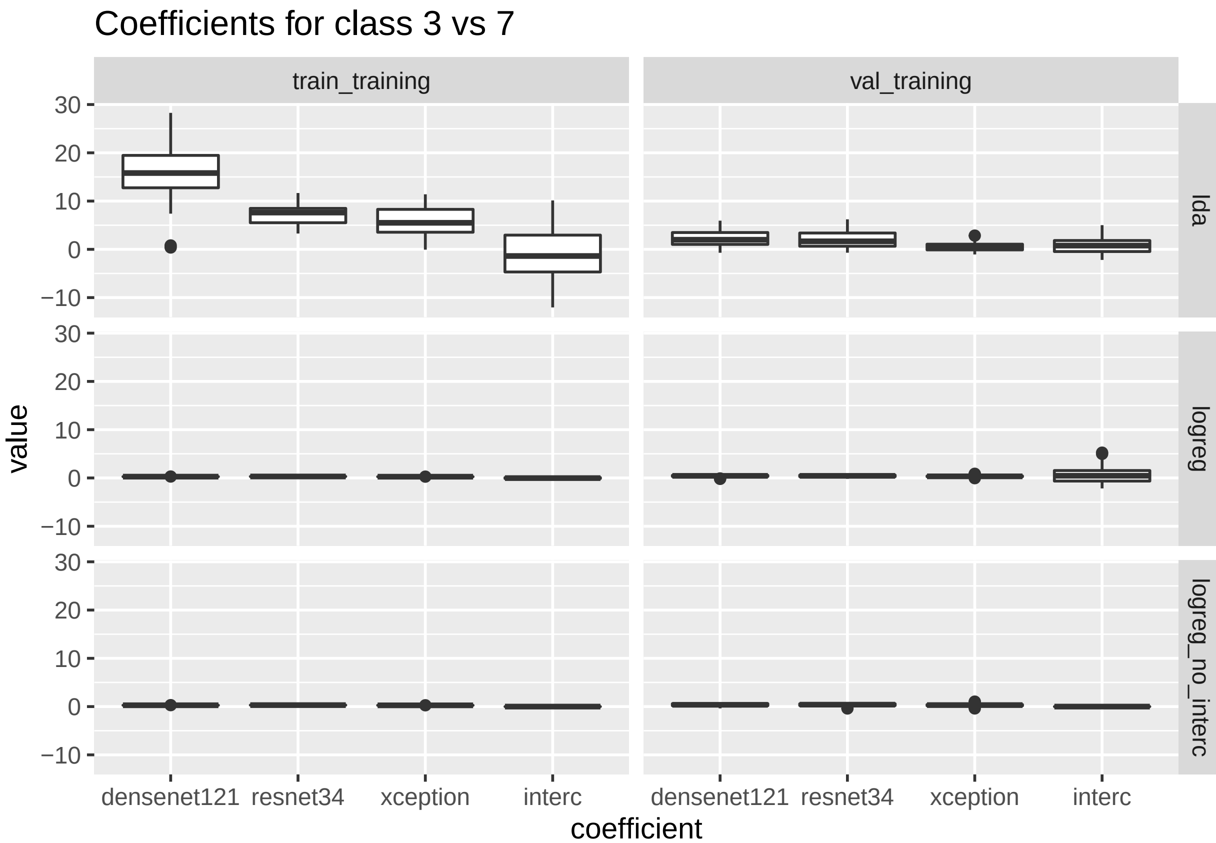

Coefficients for class 2 vs 9
train_training
val_training
lda
logreg
logreg_no_interc
value
40
30
20
10
0
densenet121
resnet34
xception
interc
coefficient

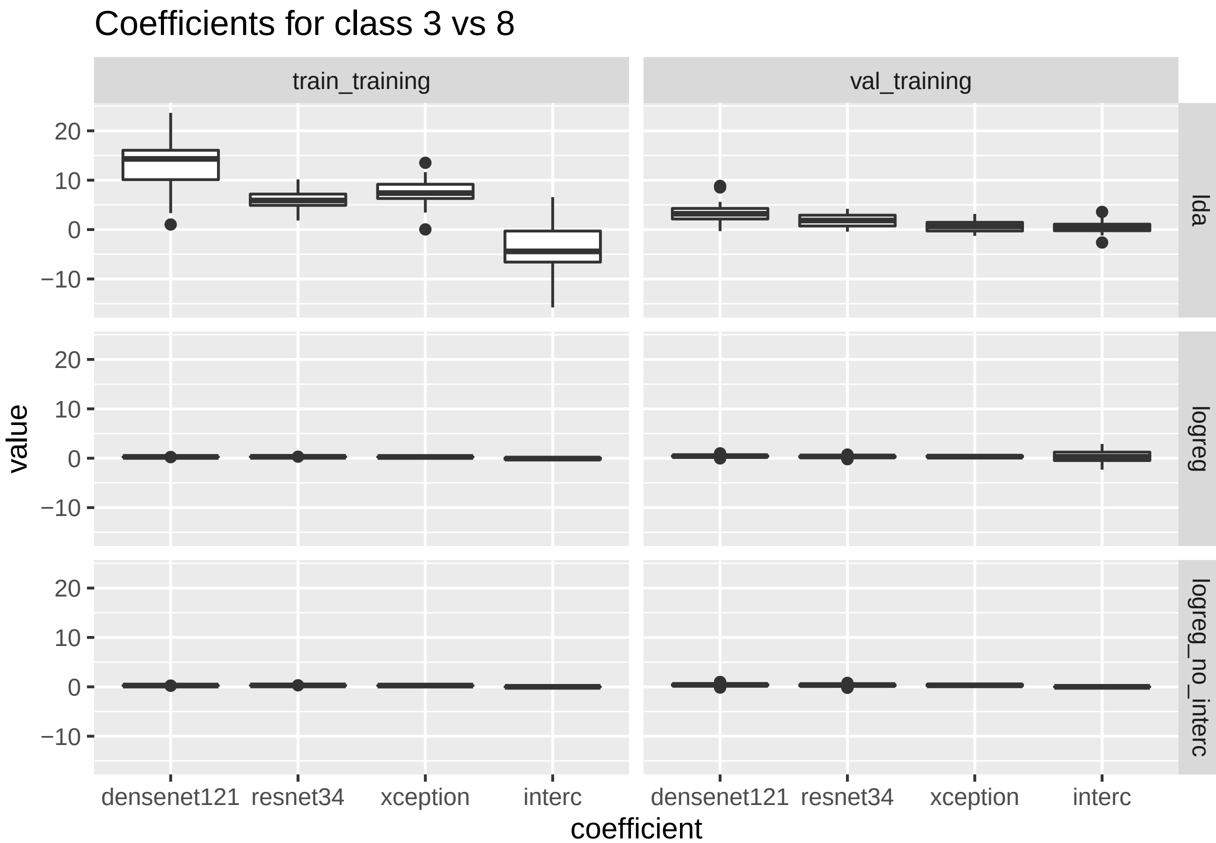

Coefficients for class 2 vs 10
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient

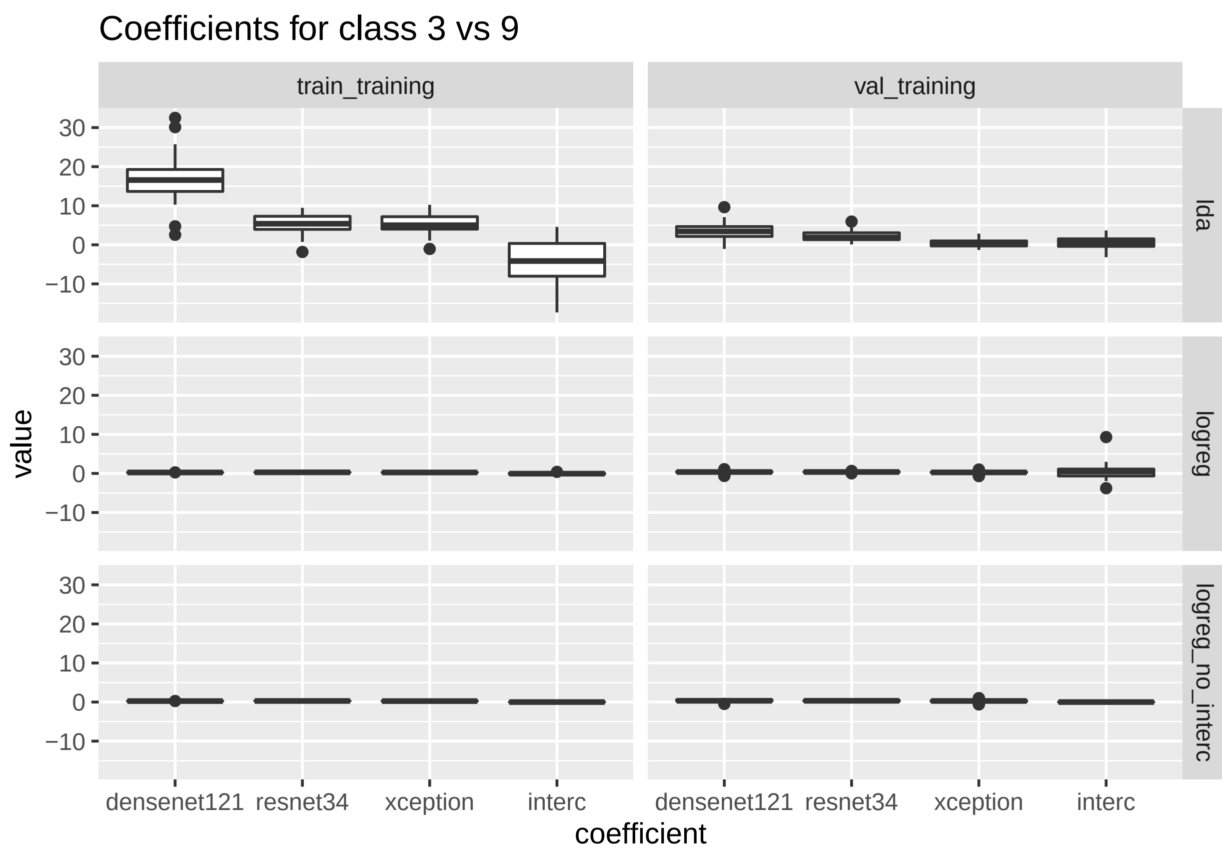

Coefficients for class 3 vs 4
train_training
val_training
lda
logreg
logreg_no_interc
value
20
15
10
5
0
densenet121
resnet34
xception
interc
coefficient

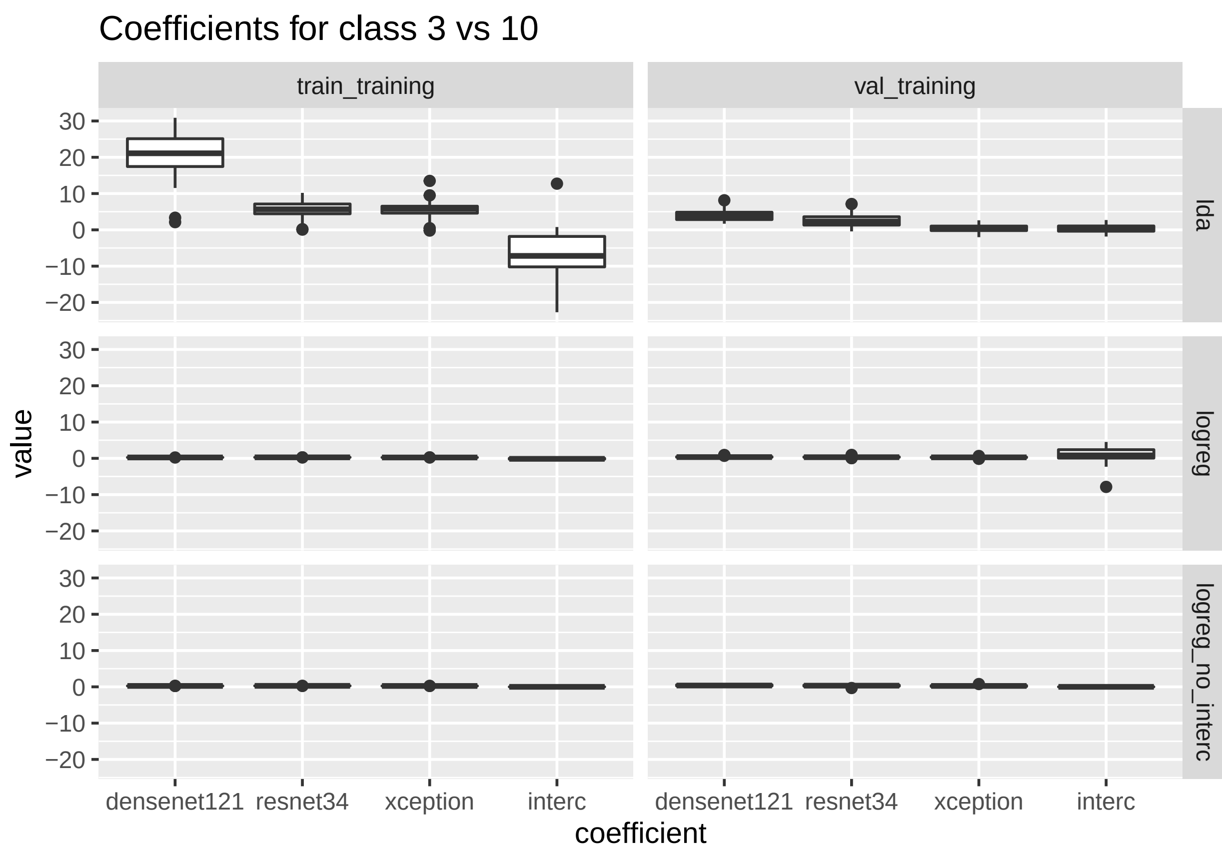

Coefficients for class 3 vs 5
train_training
val_training
lda
logreg
logreg_no_interc
value
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient


Coefficients for class 3 vs 6
train_training
val_training
lda
logreg
logreg_no_interc
value
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient

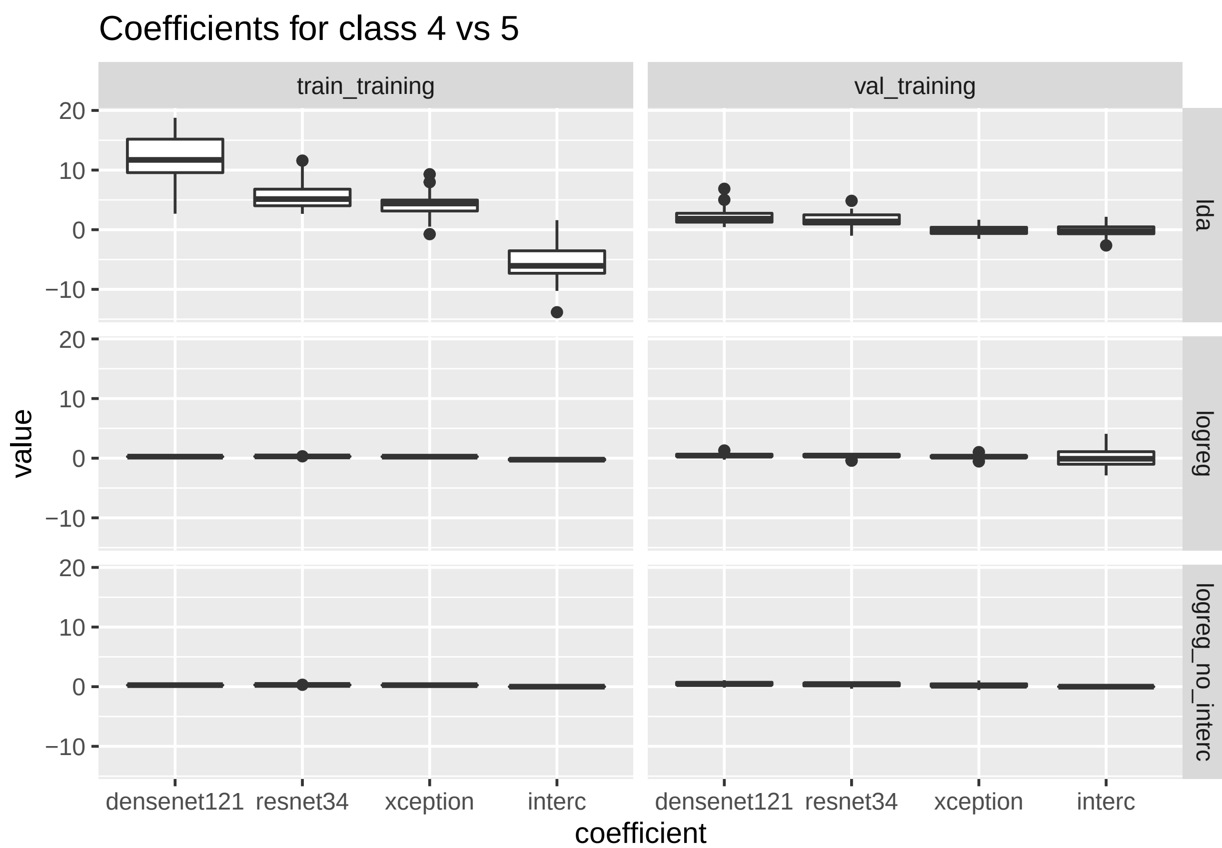

Coefficients for class 3 vs 7
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient

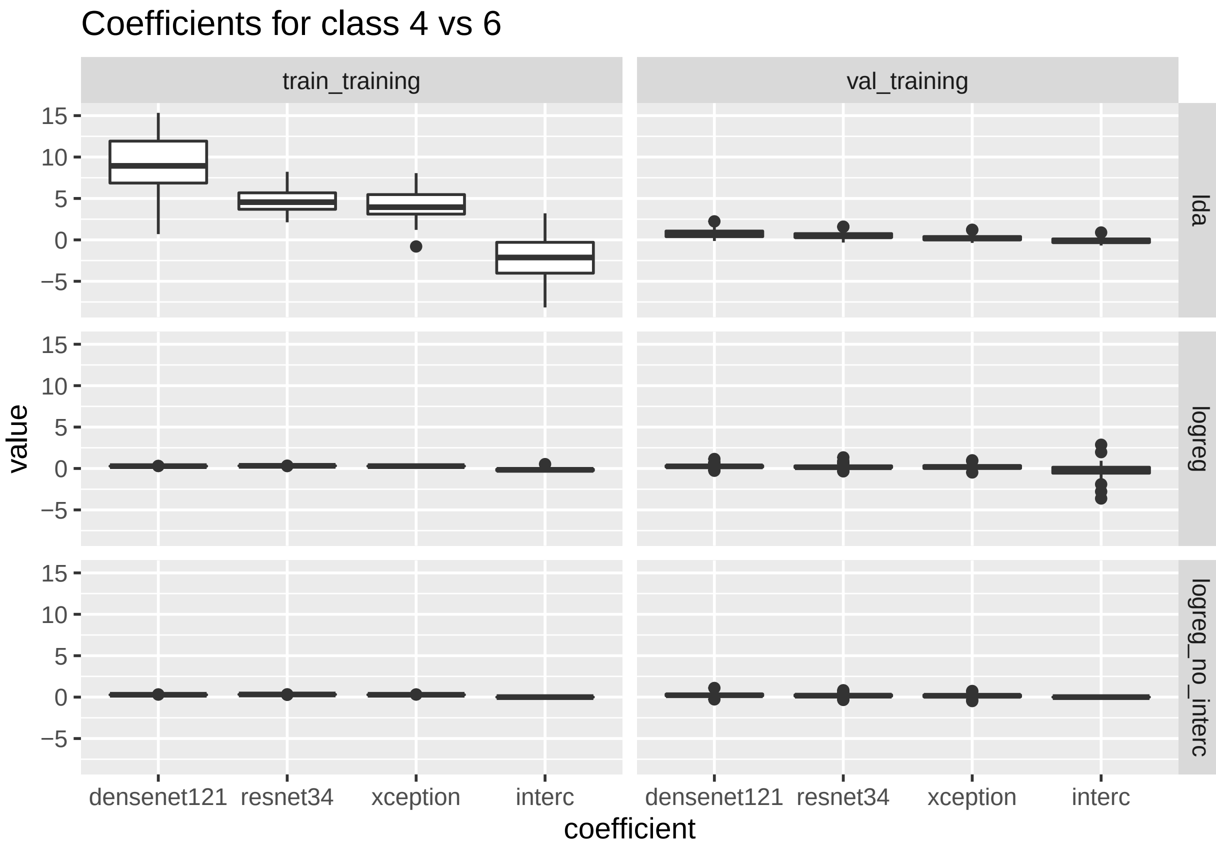

Coefficients for class 3 vs 8
train_training
val_training
lda
logreg
logreg_no_interc
value
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient

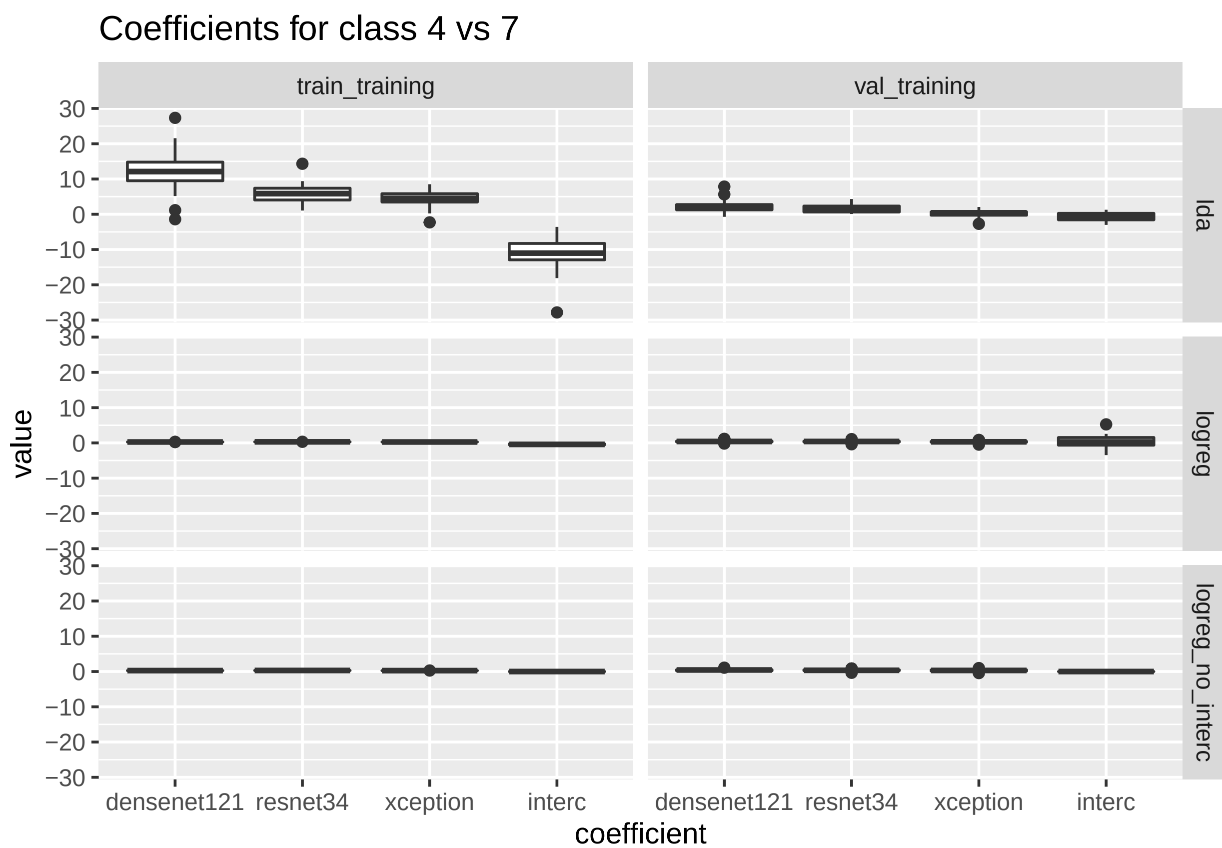

Coefficients for class 3 vs 9
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient



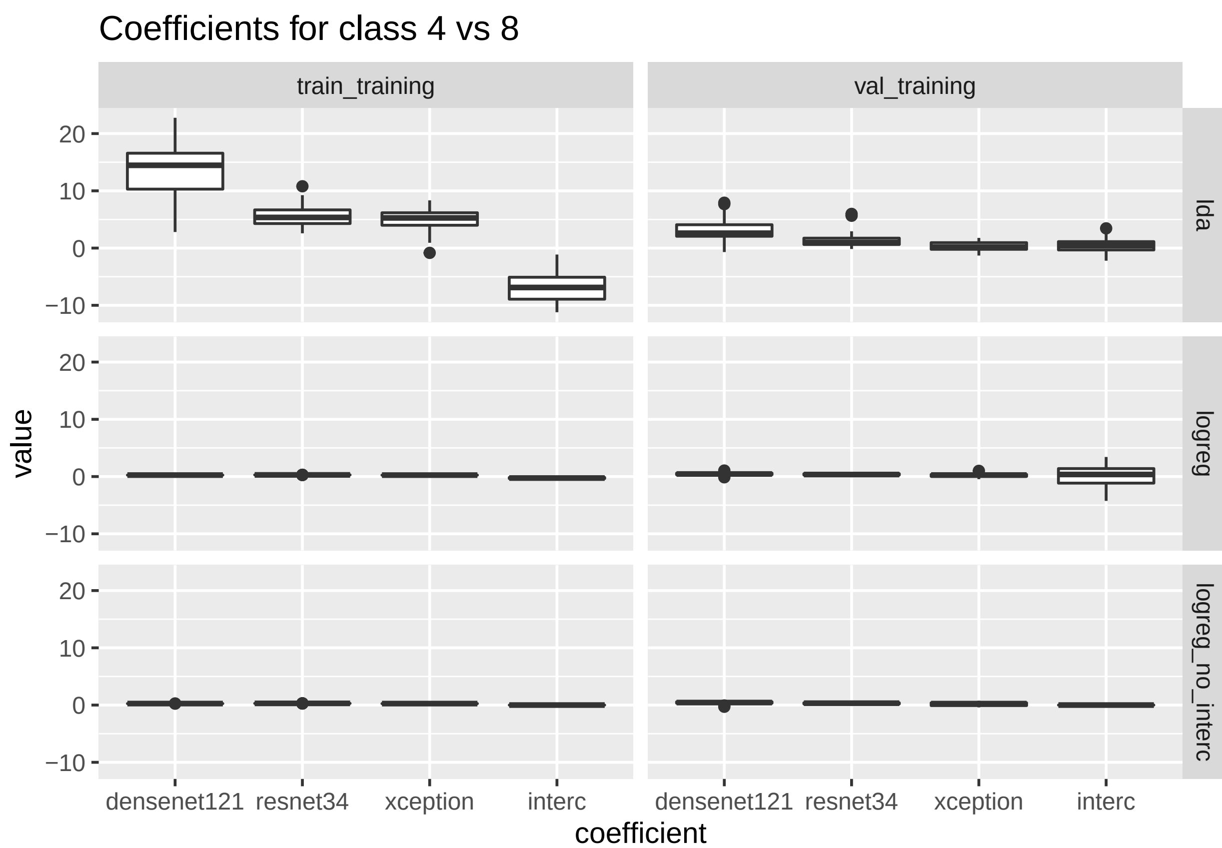
Coefficients for class 3 vs 10
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
−10
−20
densenet121
resnet34
xception
interc
coefficient

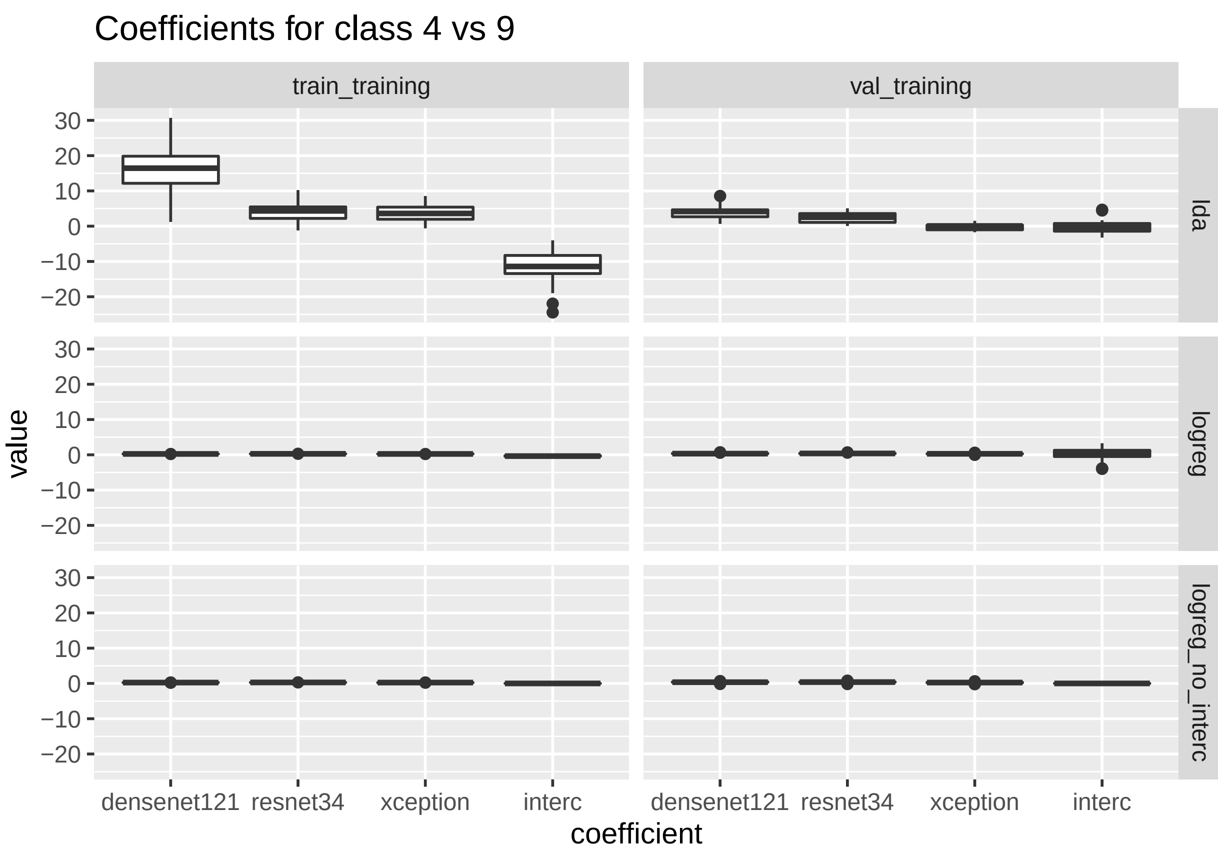

Coefficients for class 4 vs 5
train_training
val_training
lda
logreg
logreg_no_interc
value
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient

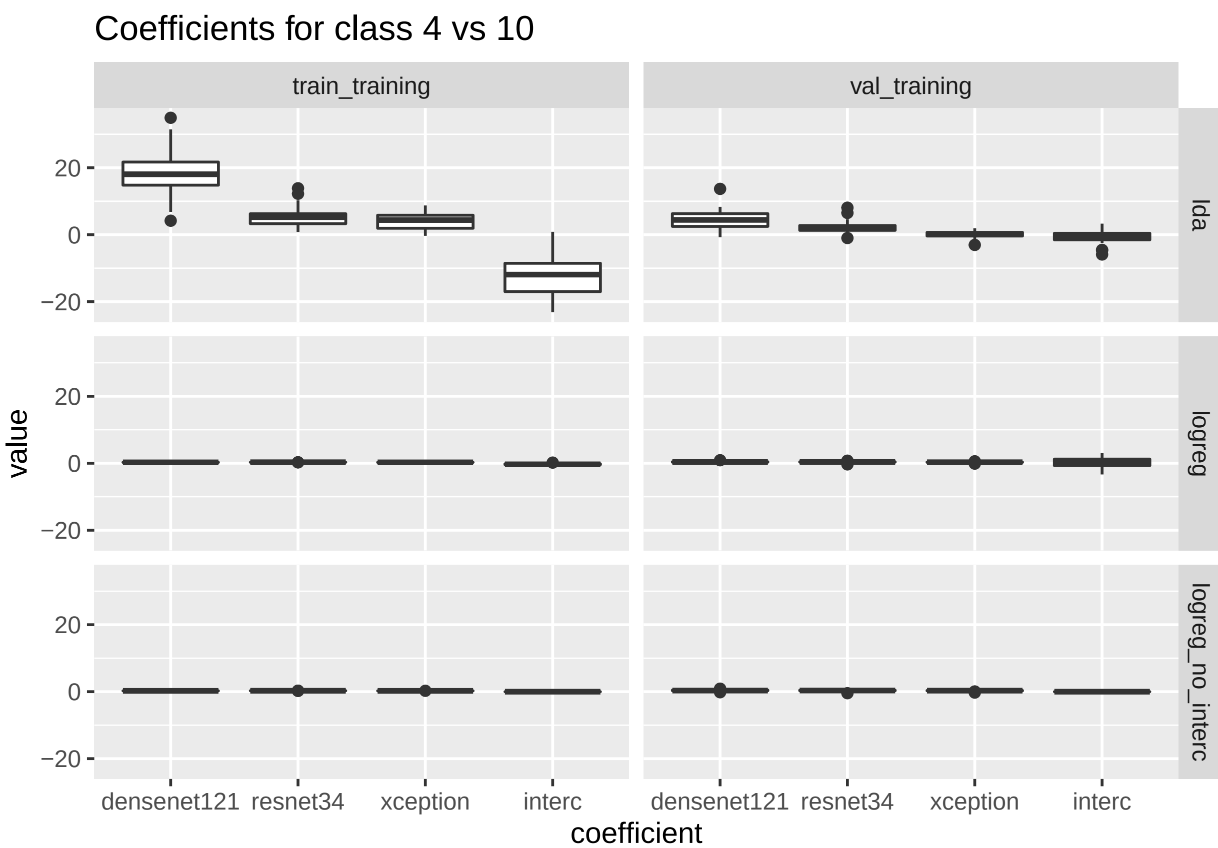

Coefficients for class 4 vs 6
train_training
val_training
lda
logreg
logreg_no_interc
value
15
10
5
0
−5
densenet121
resnet34
xception
interc
coefficient



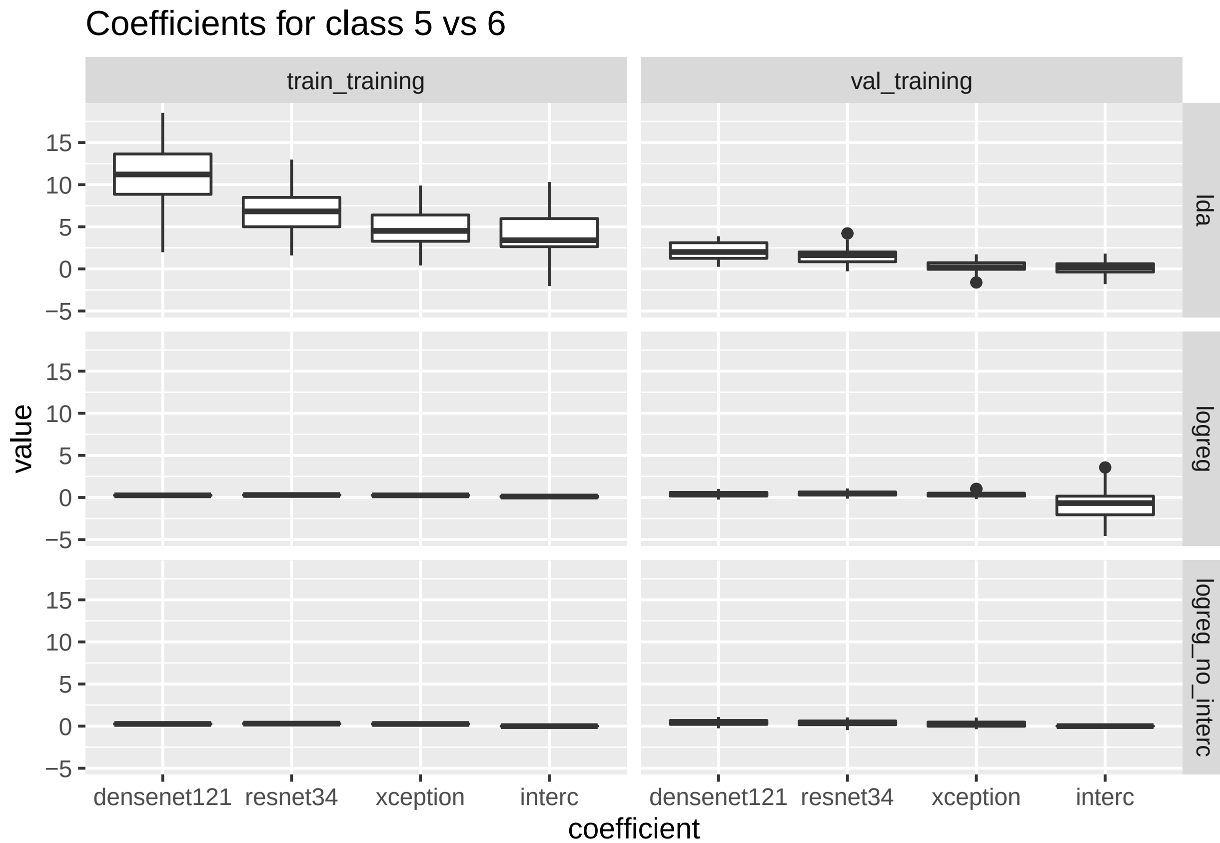
Coefficients for class 4 vs 7
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
−10
−20
−30
densenet121
resnet34
xception
interc
coefficient

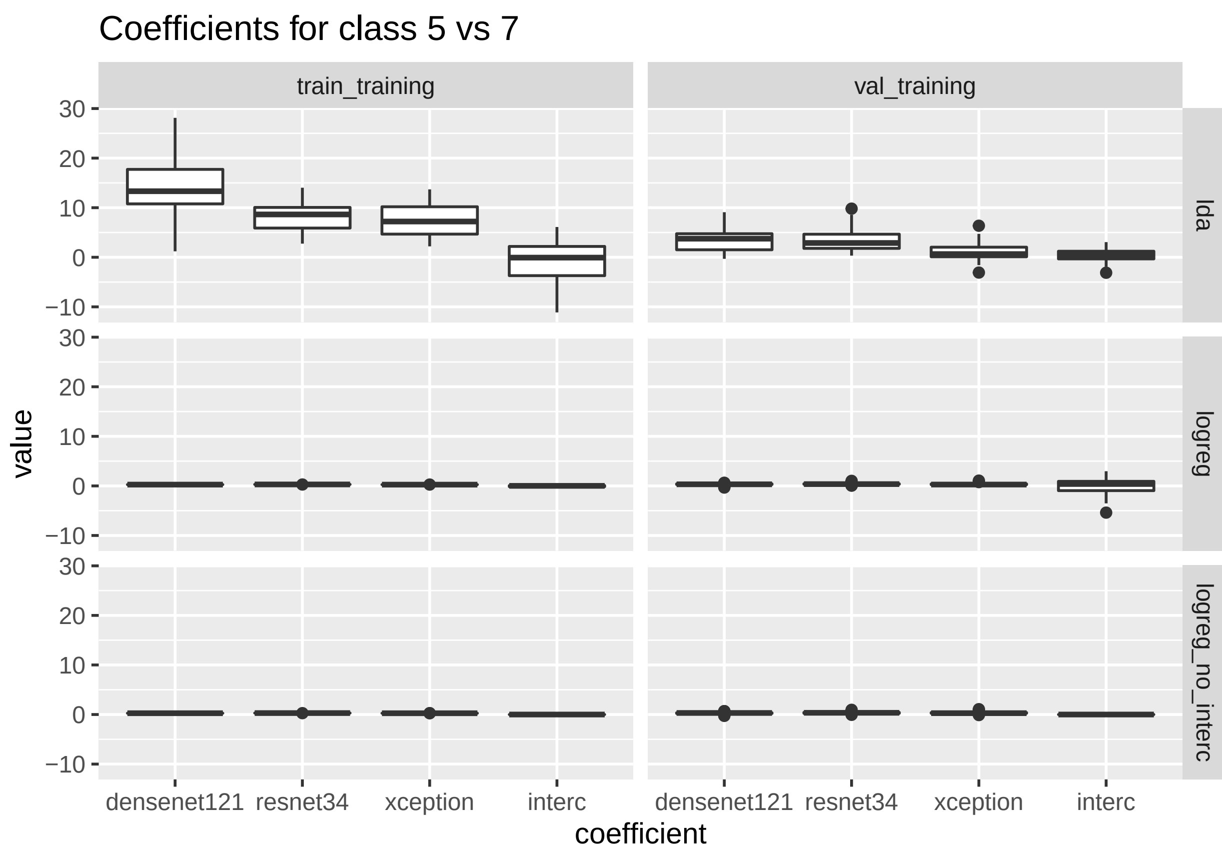

Coefficients for class 4 vs 8
train_training
val_training
lda
logreg
logreg_no_interc
value
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient

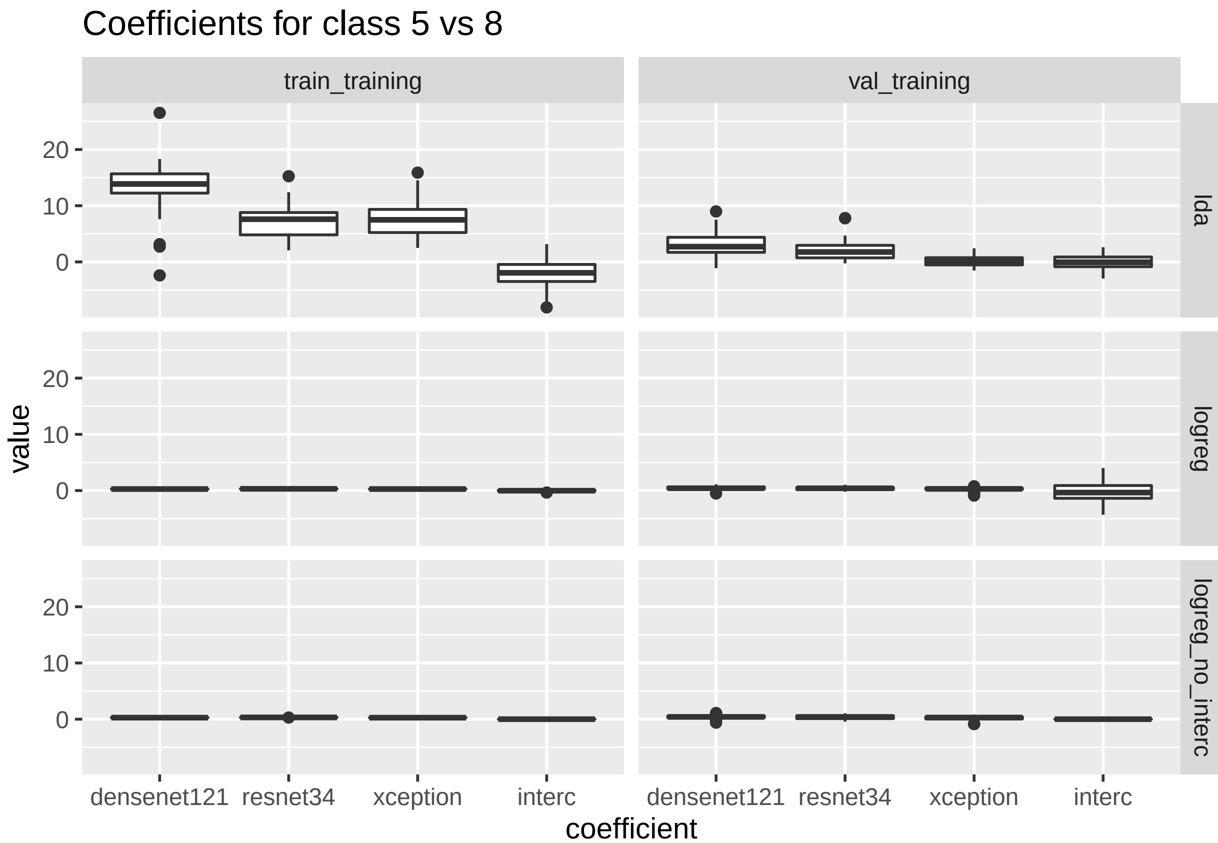

Coefficients for class 4 vs 9
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
−10
−20
densenet121
resnet34
xception
interc
coefficient

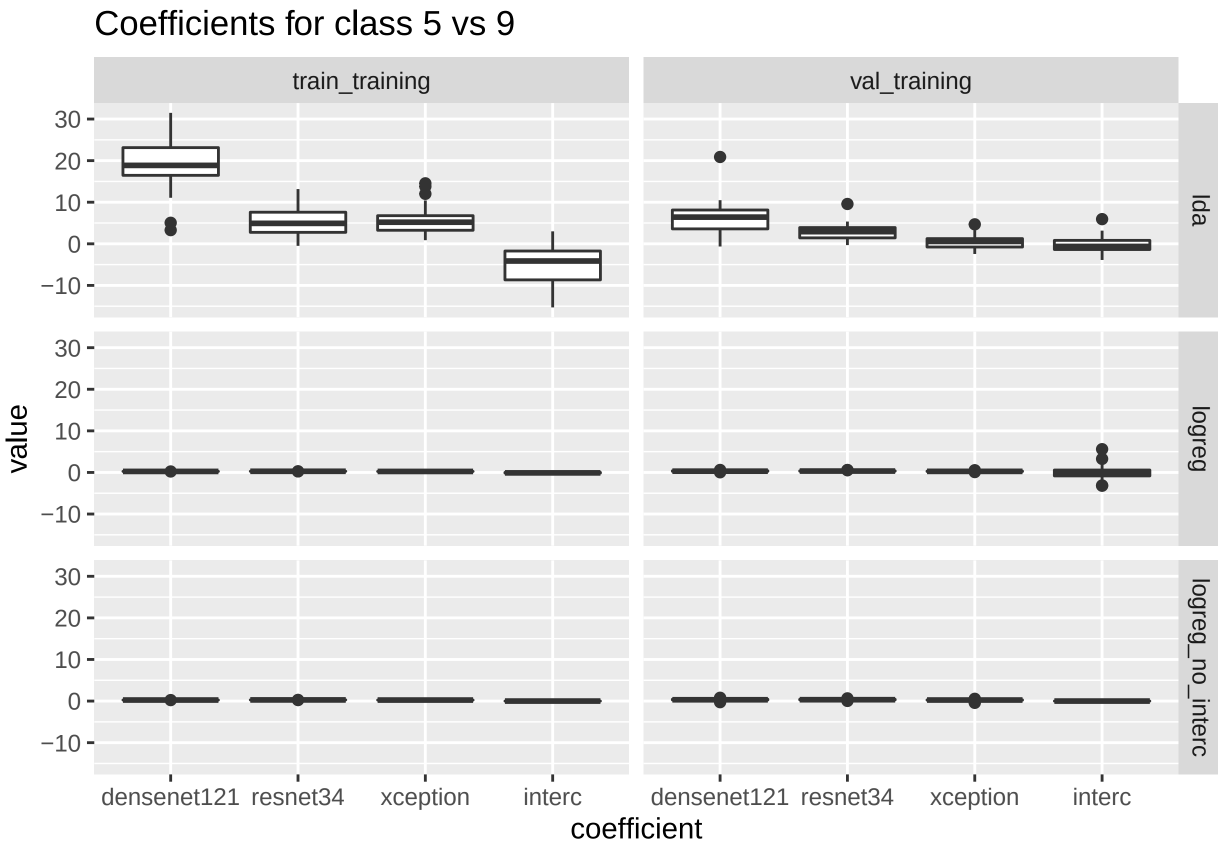


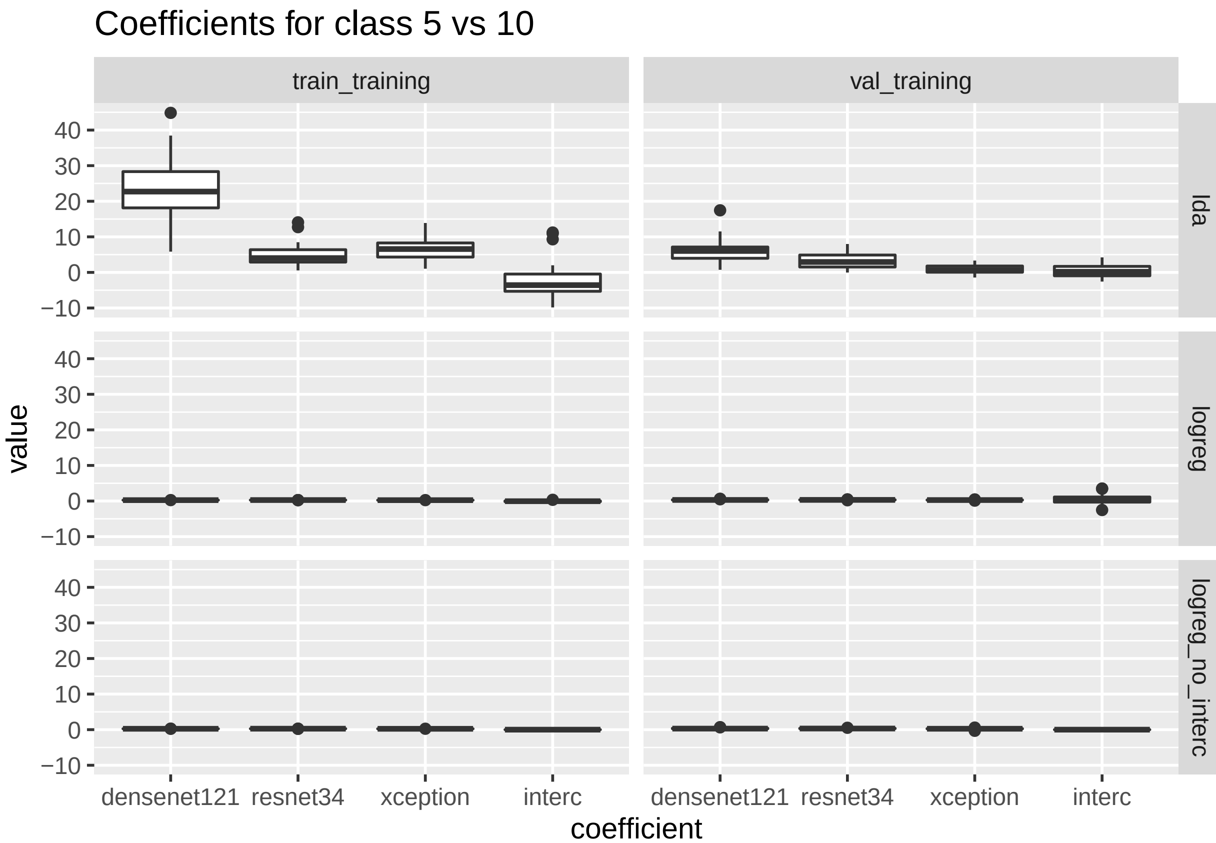
Coefficients for class 4 vs 10
train_training
val_training
lda
logreg
logreg_no_interc
value
20
0
−20
densenet121
resnet34
xception
interc
coefficient

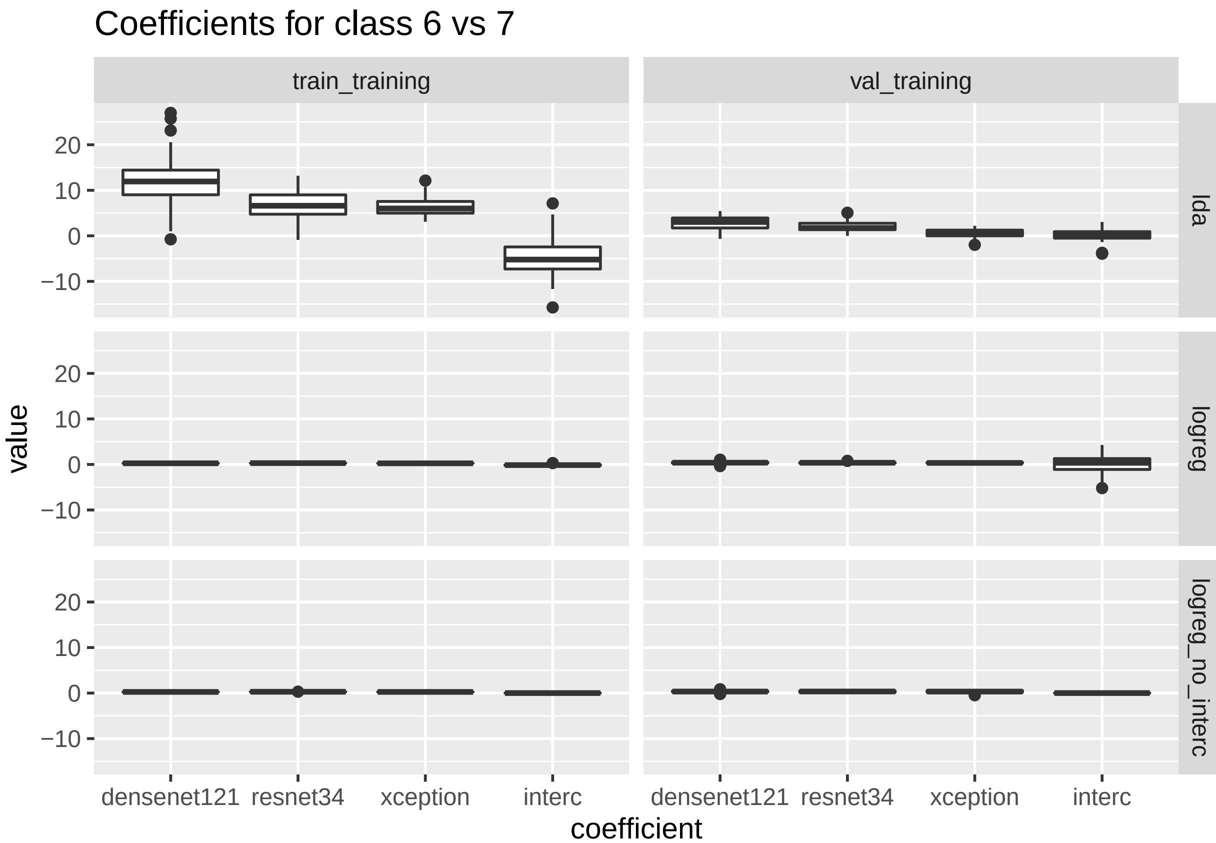

Coefficients for class 5 vs 6
train_training
val_training
lda
logreg
logreg_no_interc
value
15
10
5
0
−5
densenet121
resnet34
xception
interc
coefficient

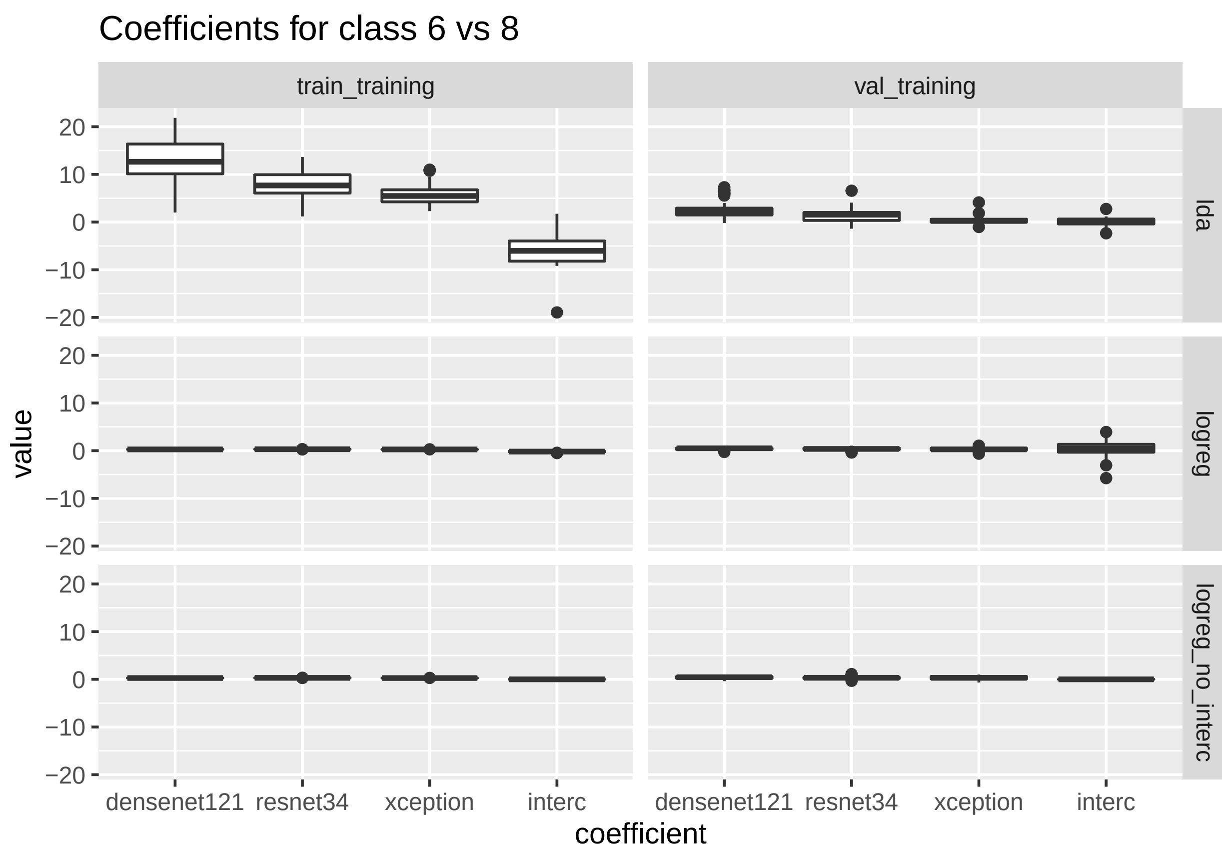

Coefficients for class 5 vs 7
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient

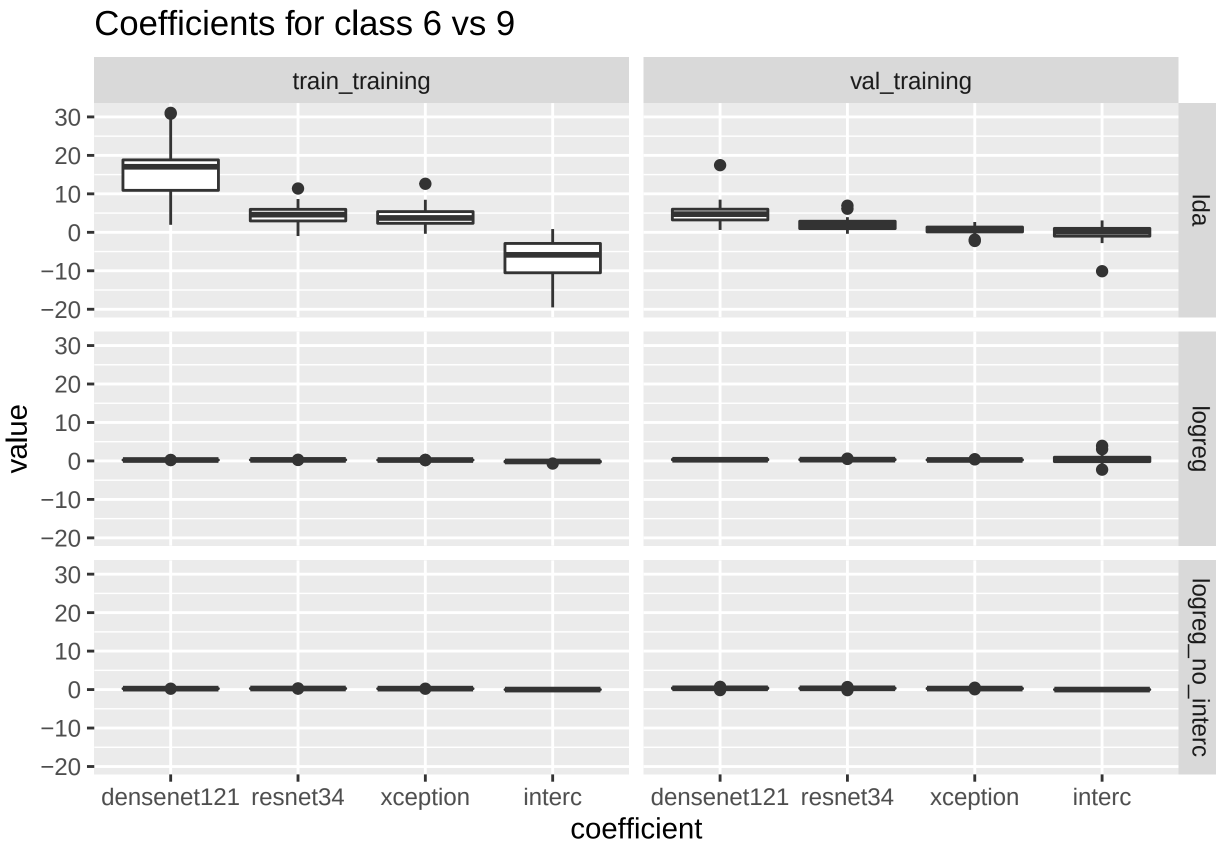

Coefficients for class 5 vs 8
train_training
val_training
lda
logreg
logreg_no_interc
value
20
10
0
densenet121
resnet34
xception
interc
coefficient

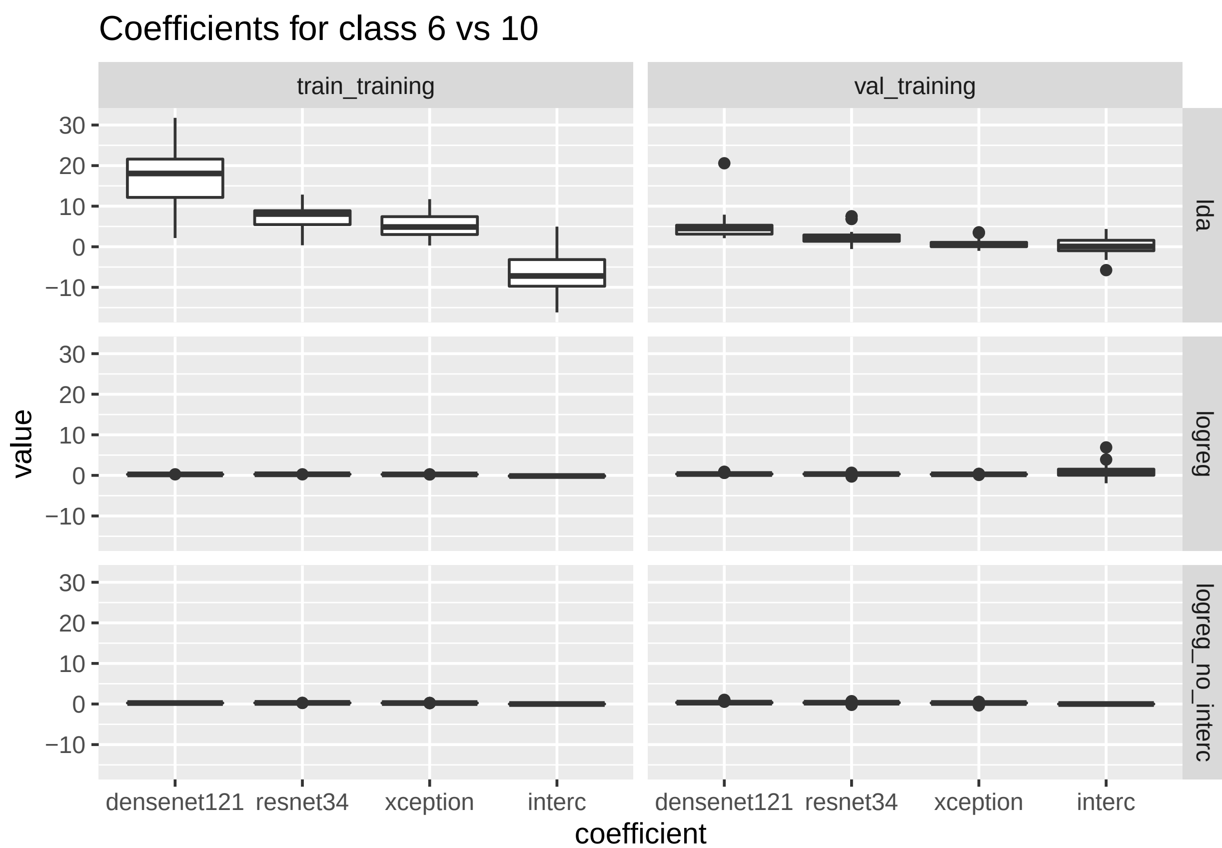

Coefficients for class 5 vs 9
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient


Coefficients for class 5 vs 10
train_training
val_training
lda
logreg
logreg_no_interc
value
40
30
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient

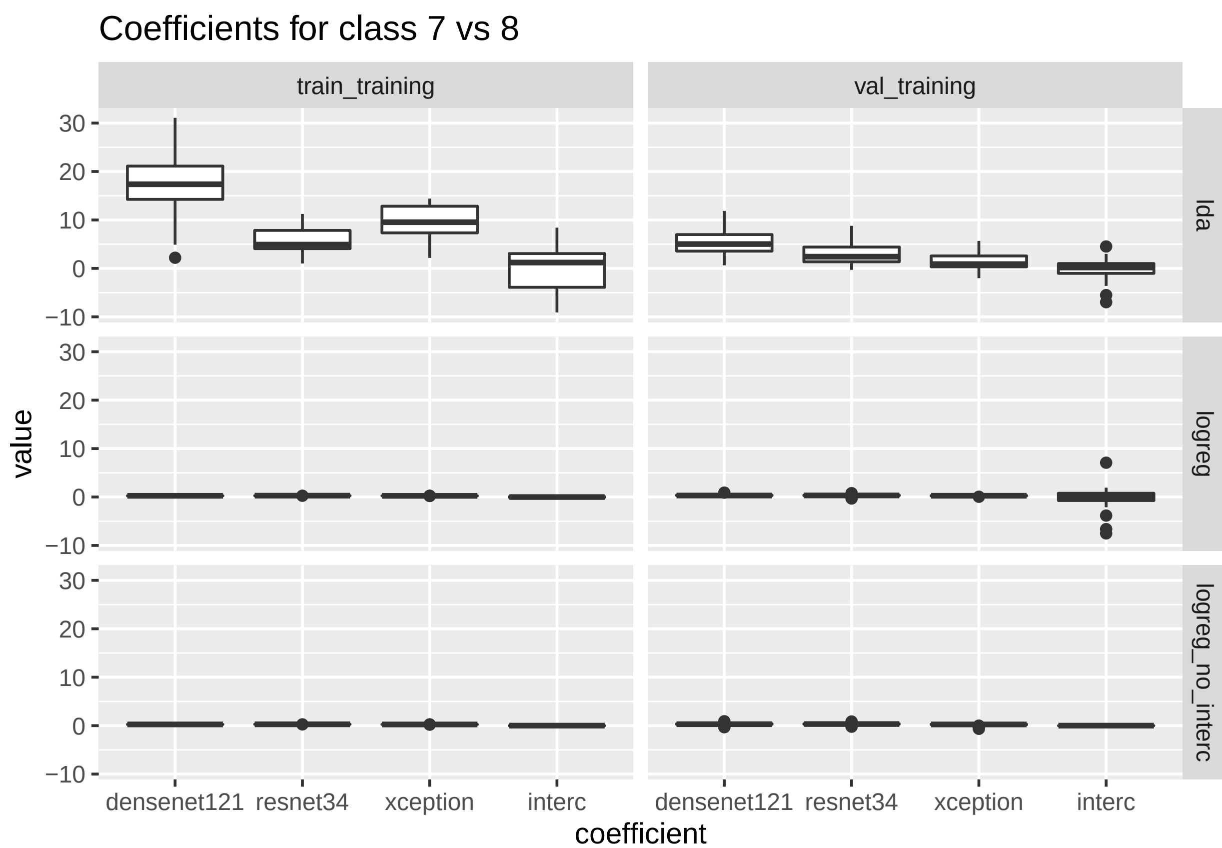

Coefficients for class 6 vs 7
train_training
val_training
lda
logreg
logreg_no_interc
value
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient

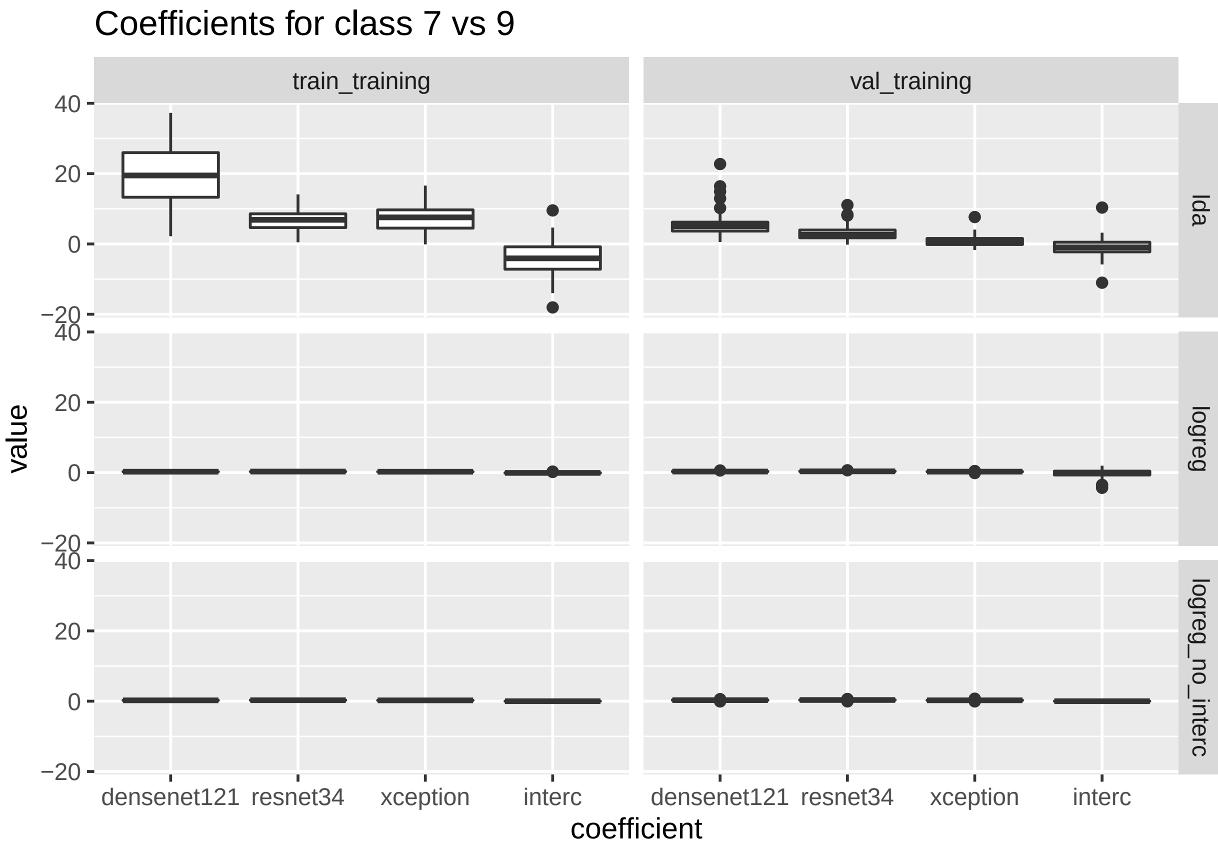

Coefficients for class 6 vs 8
train_training
val_training
lda
logreg
logreg_no_interc
value
20
10
0
−10
−20
densenet121
resnet34
xception
interc
coefficient

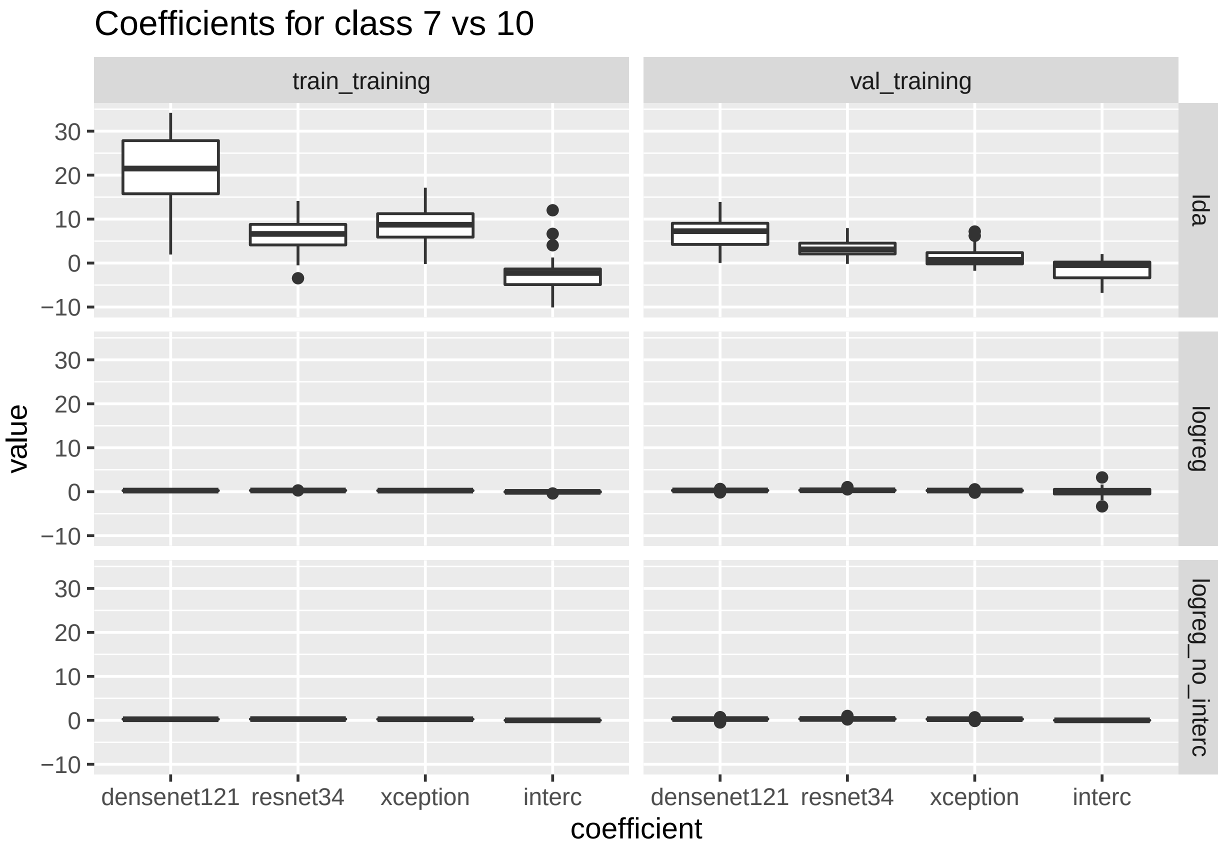

Coefficients for class 6 vs 9
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
−10
−20
densenet121
resnet34
xception
interc
coefficient

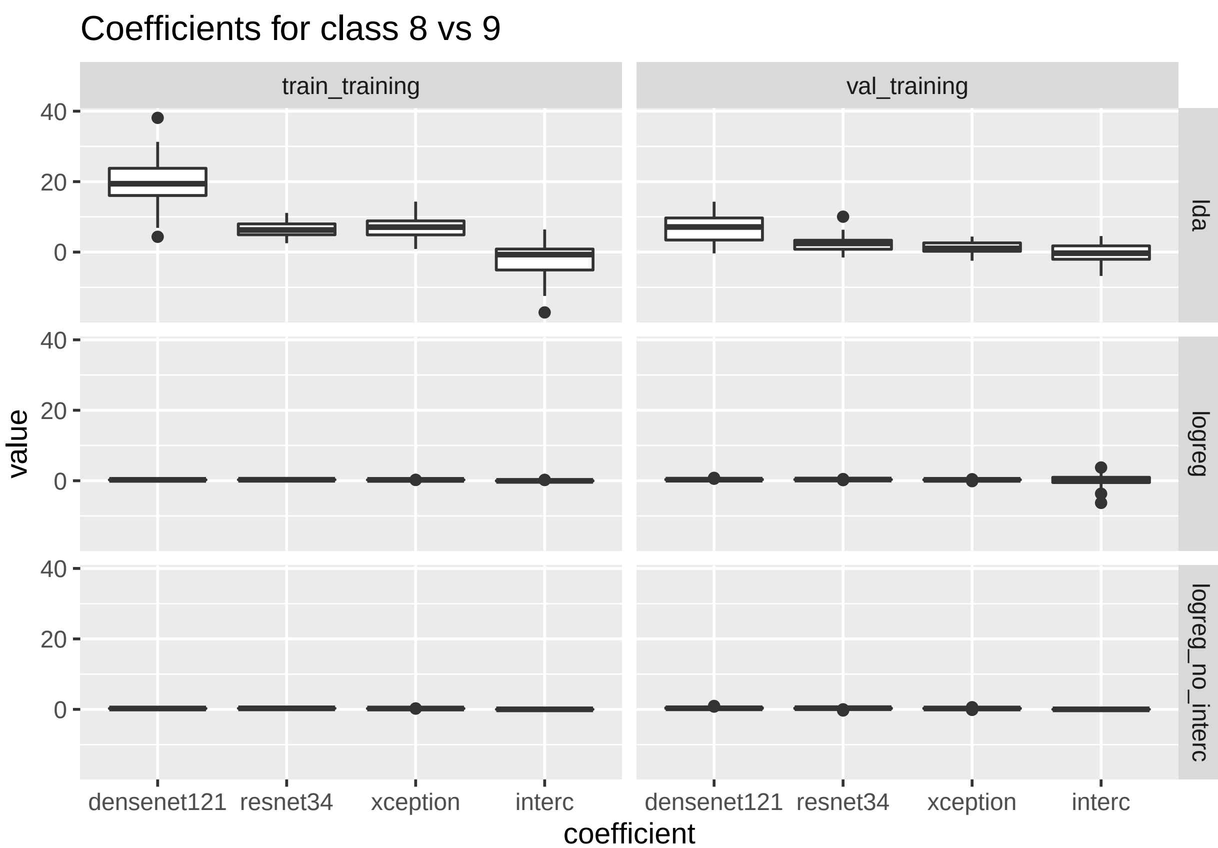


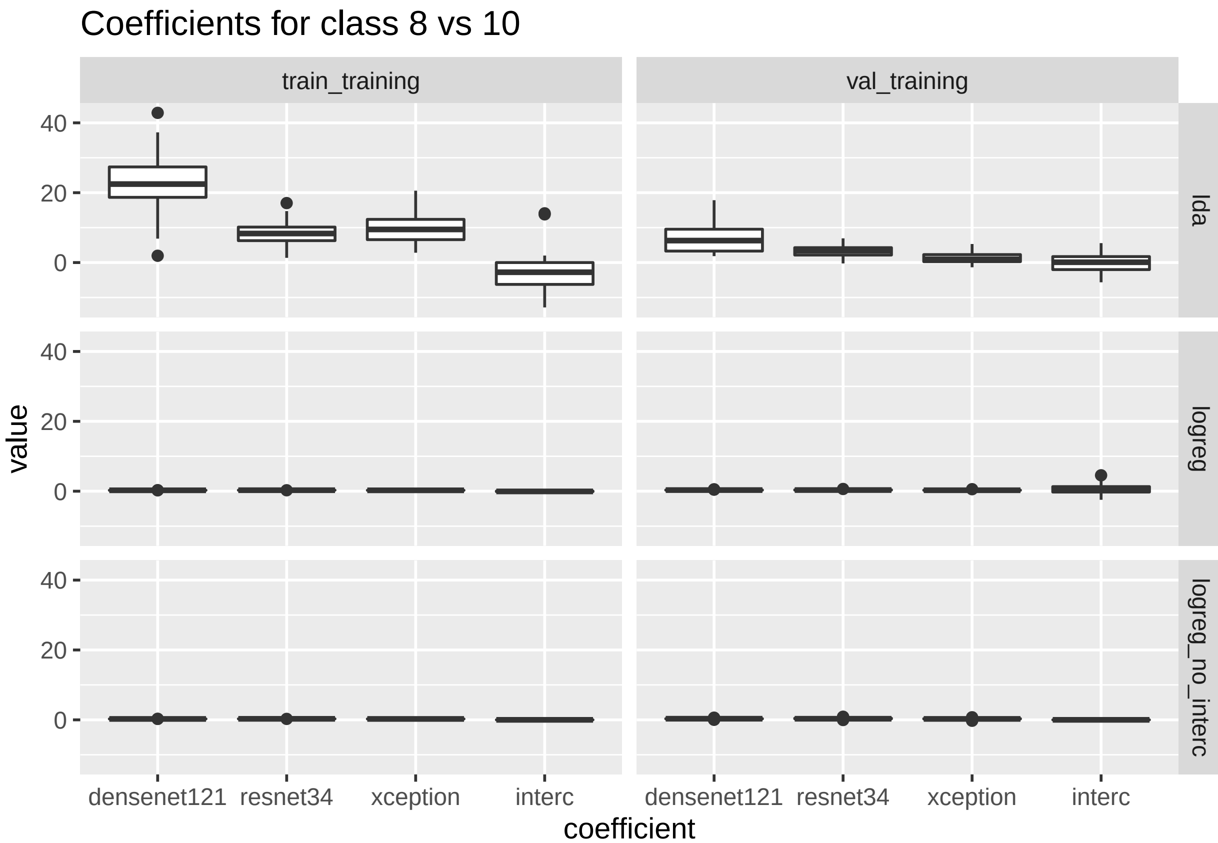
Coefficients for class 6 vs 10
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient

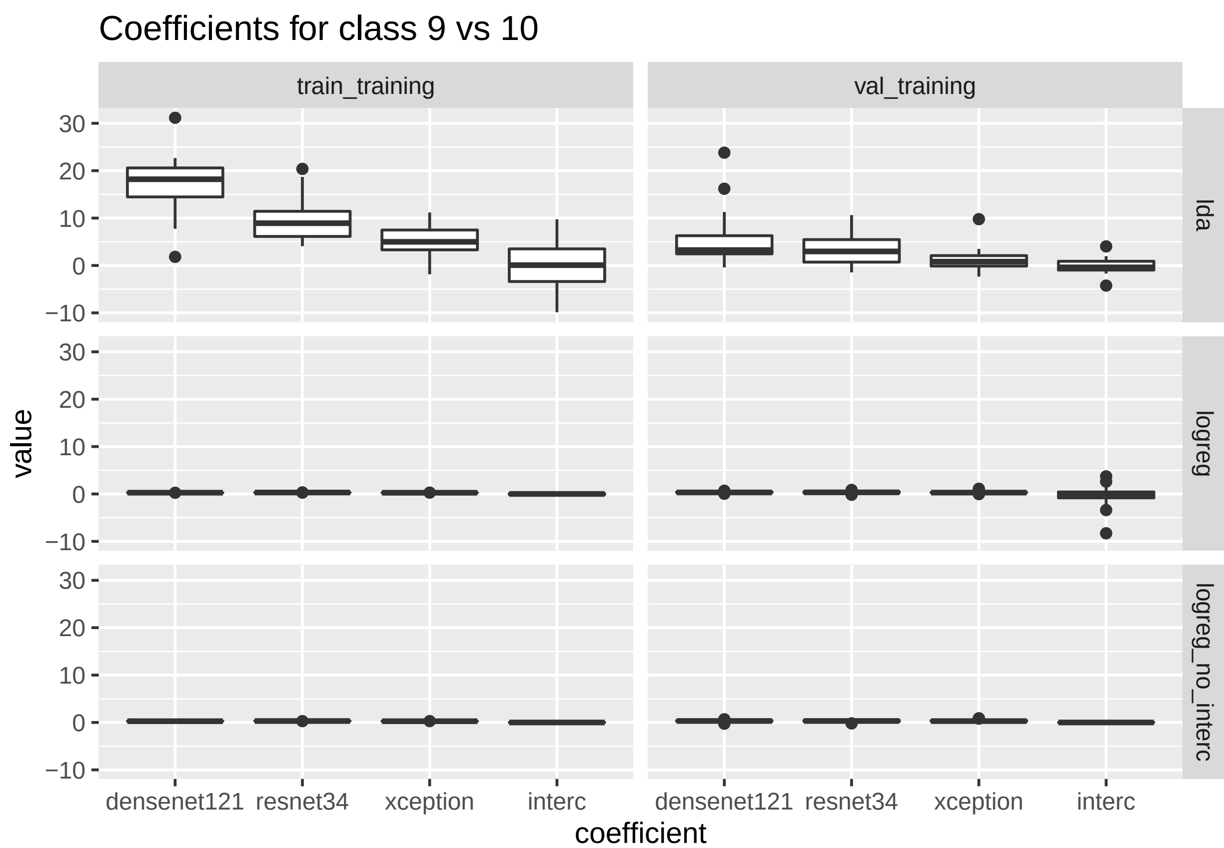

Coefficients for class 7 vs 8
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient


Coefficients for class 7 vs 9
train_training
val_training
lda
logreg
logreg_no_interc
value
40
20
0
−20
densenet121
resnet34
xception
interc
coefficient


Coefficients for class 7 vs 10
train_training
val_training
lda
logreg
logreg_no_interc
value
30
20
10
0
−10
densenet121
resnet34
xception
interc
coefficient

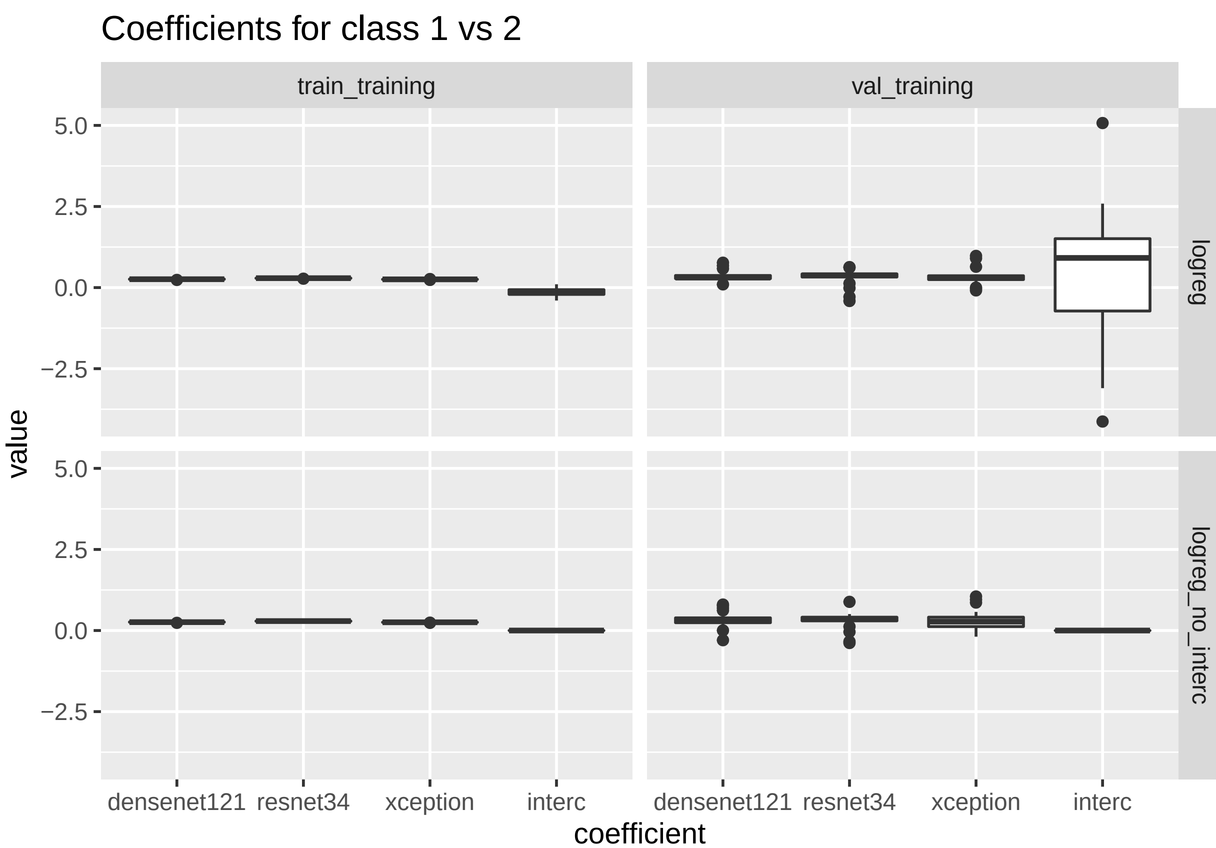

Coefficients for class 8 vs 9
train_training
val_training
lda
logreg
logreg_no_interc
value
40
20
0
densenet121
resnet34
xception
interc
coefficient

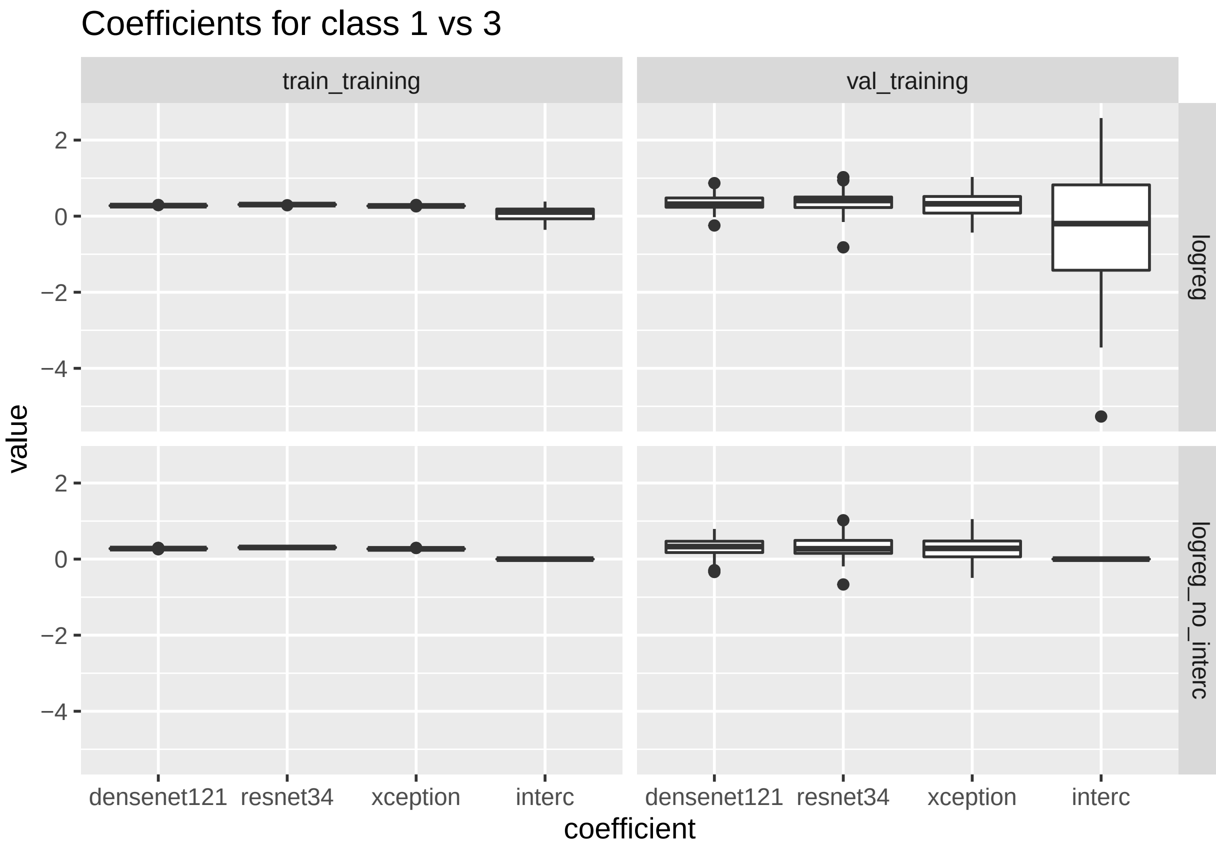

Coefficients for class 8 vs 10
train_training
val_training
lda
logreg
logreg_no_interc
value
40
20
0
densenet121
resnet34
xception
interc
coefficient



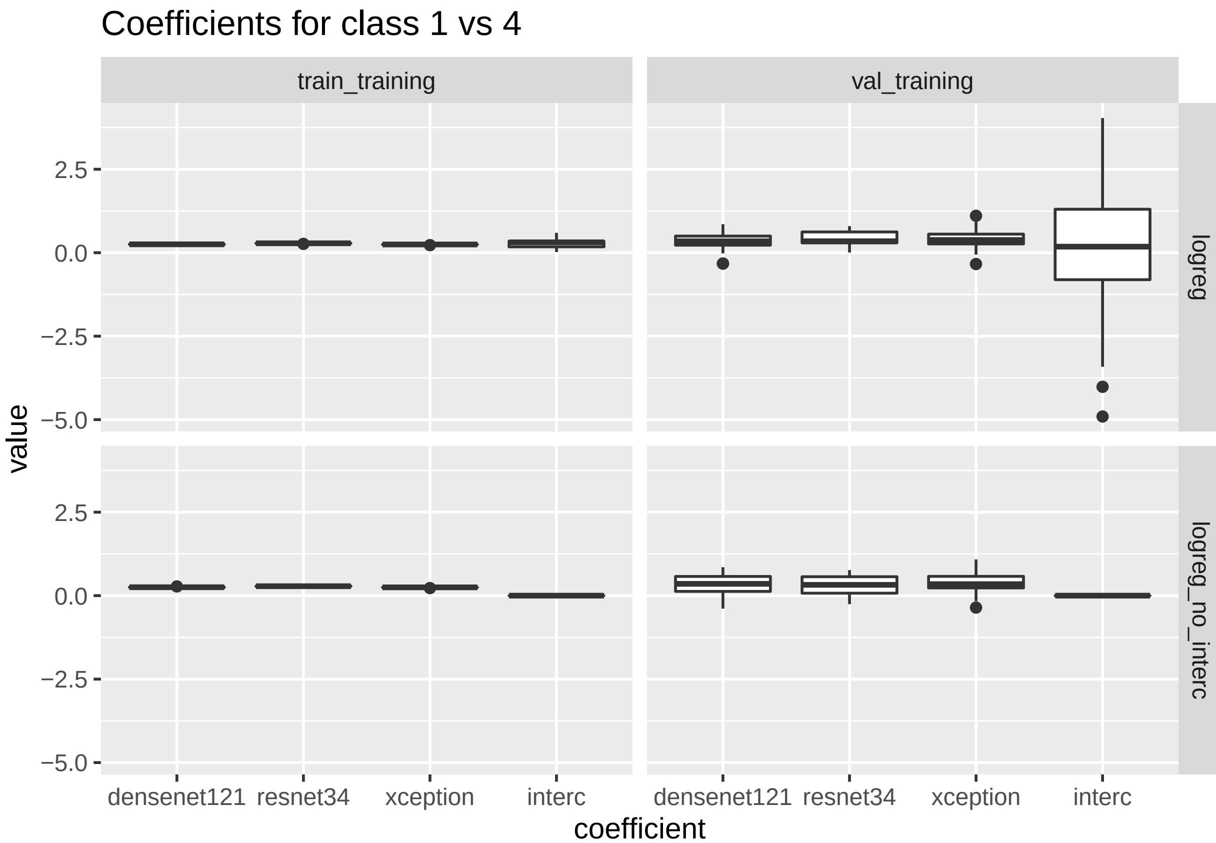
Logistic regression coefficients are much smaller than lda, it is probably due to l2 normalization used during logreg training.

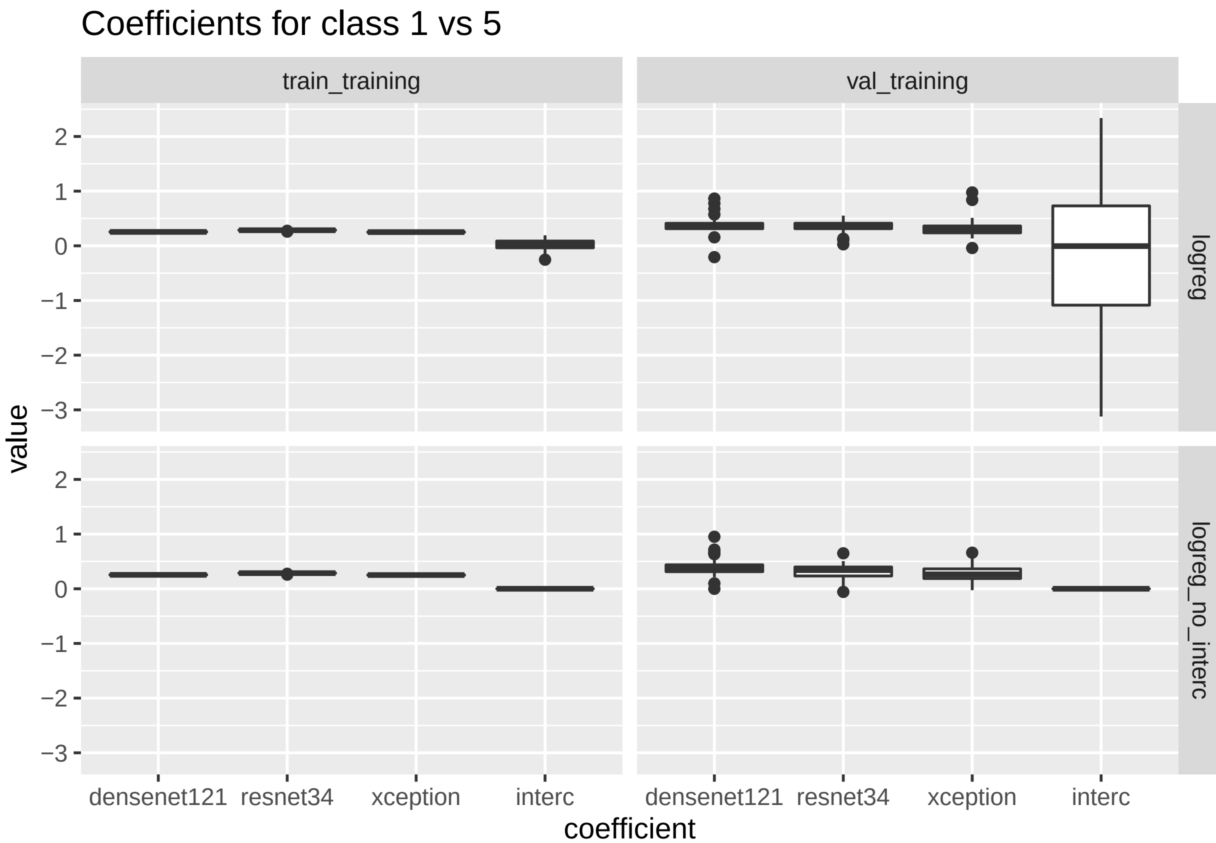
```
for (cl1 in 1:(classes - 1))
{
  for (cl2 in (cl1 + 1):classes)
  {
    cur_plt <- combiner_coefs %>% filter(class1 == cl1 & class2 == cl2 & combining_method!="lda") %>% gg
      facet_grid(cols=vars(train_type), rows=vars(combining_method)) + ggtitle(paste("Coefficients for c
    print(cur_plt)
  }
}
```


Coefficients for class 1 vs 2
train_training
val_training
logreg
logreg_no_interc
value
5.0
2.5
0.0
−2.5
densenet121
resnet34
xception
interc
coefficient

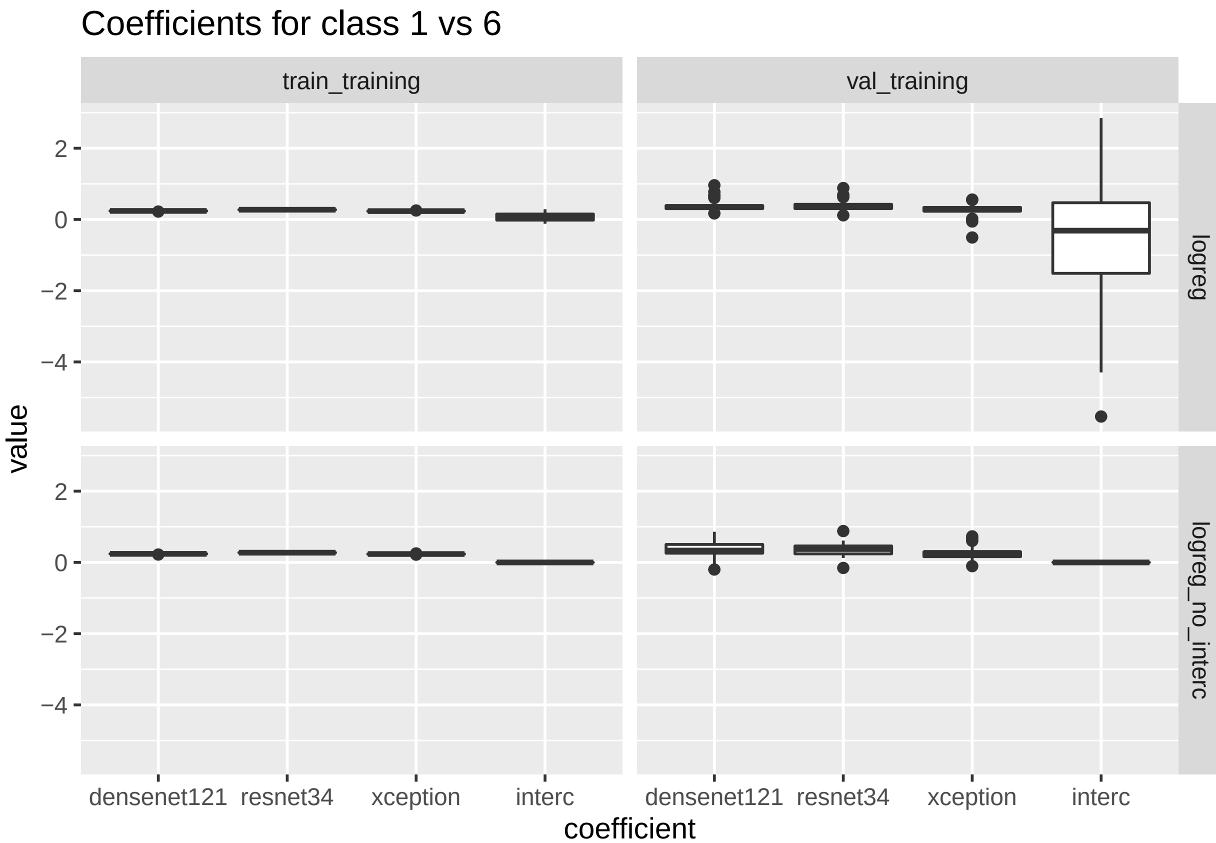


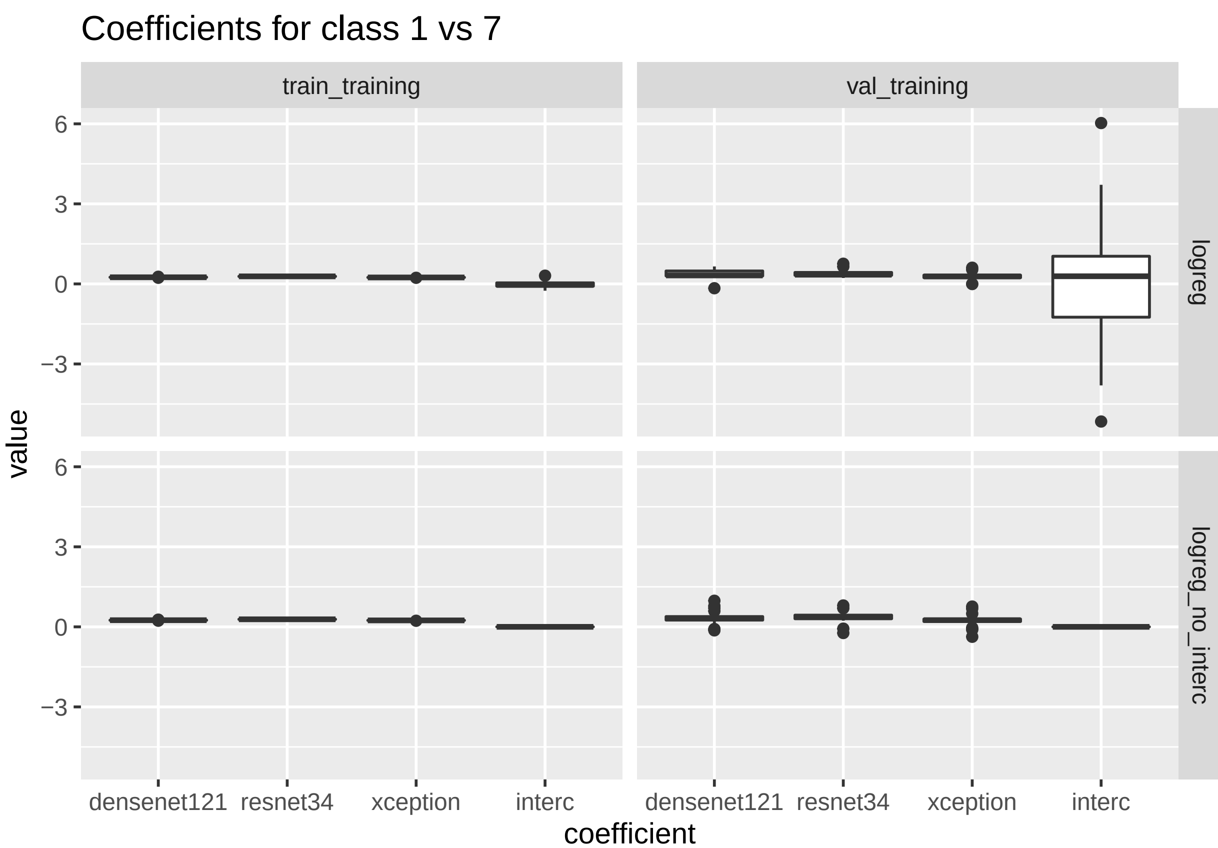
Coefficients for class 1 vs 3
train_training
val_training
logreg
logreg_no_interc
value
2
0
−2
−4
densenet121
resnet34
xception
interc
coefficient

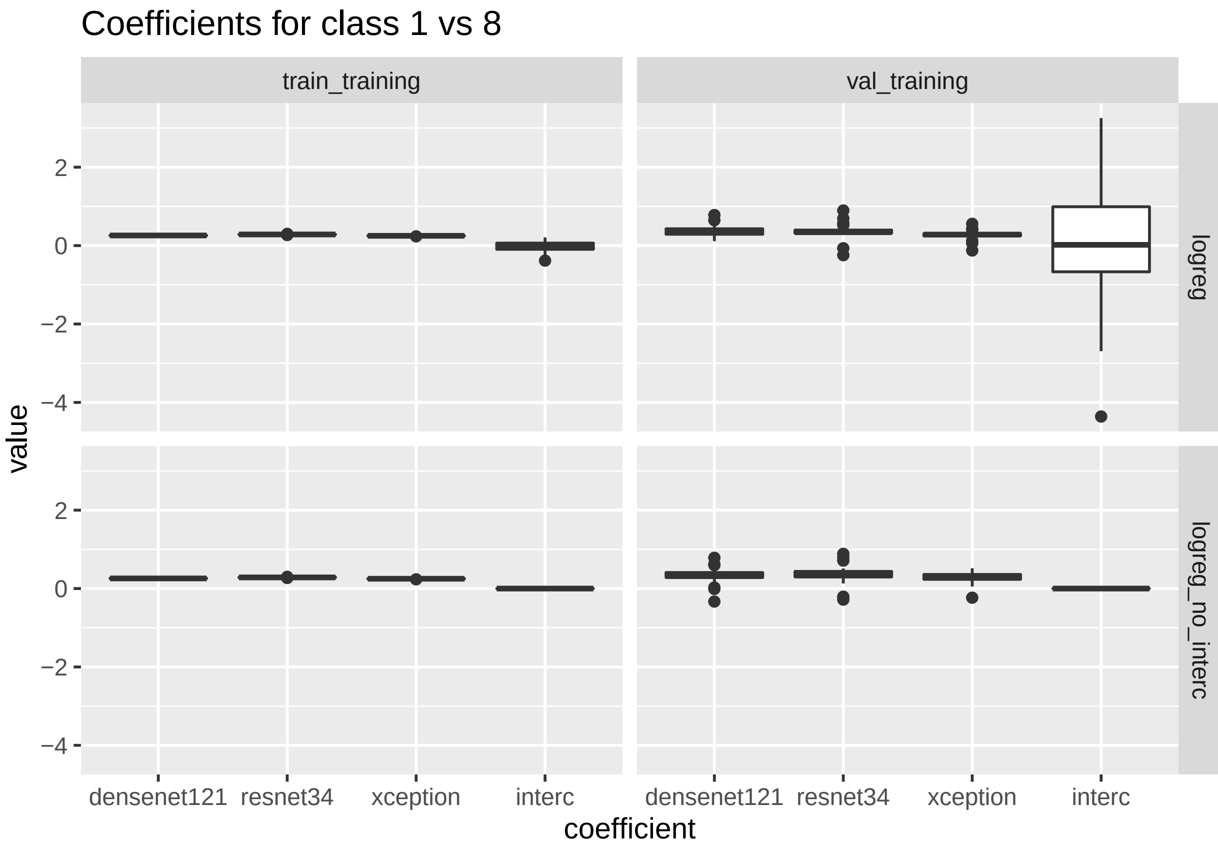

Coefficients for class 1 vs 4
train_training
val_training
logreg
logreg_no_interc
value
2.5
0.0
−2.5
−5.0
densenet121
resnet34
xception
interc
coefficient

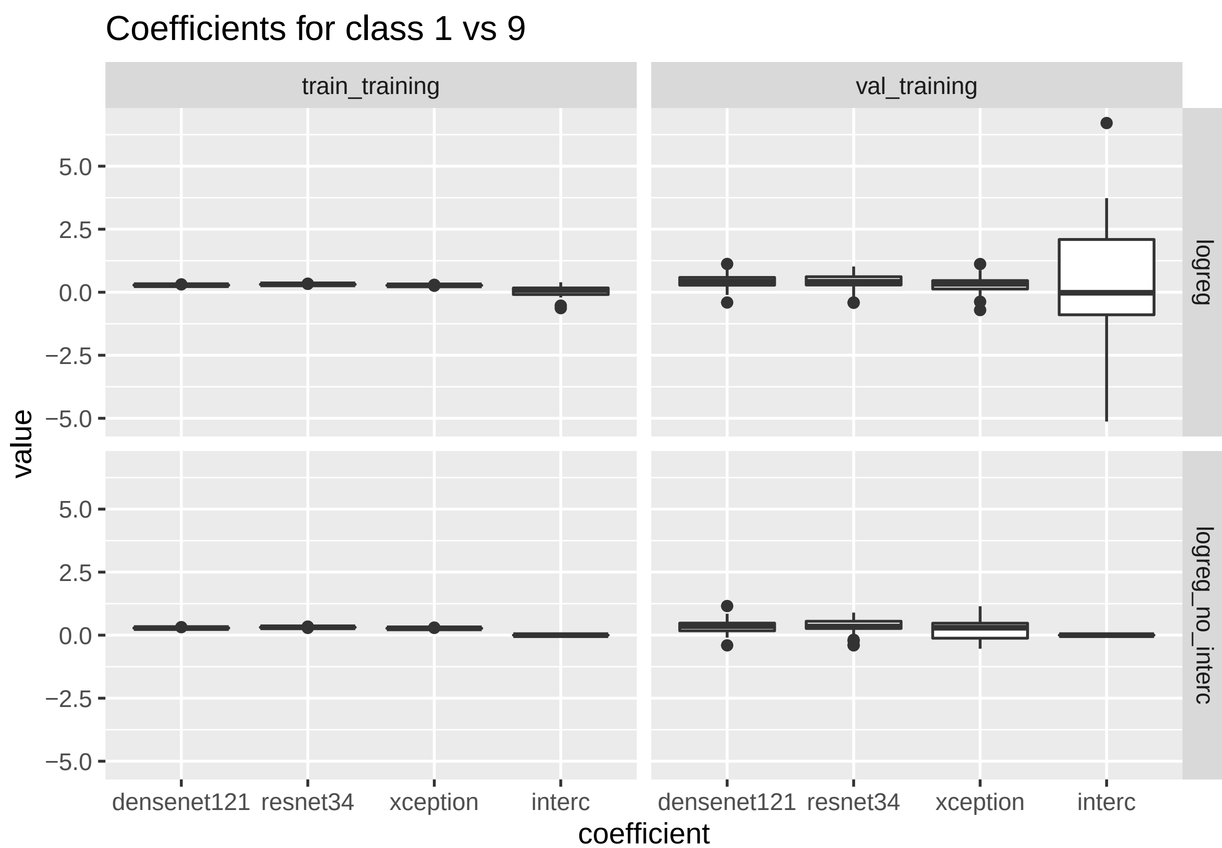


Coefficients for class 1 vs 5
train_training
val_training
logreg
logreg_no_interc
value
2
1
0
−1
−2
−3
densenet121
resnet34
xception
interc
coefficient

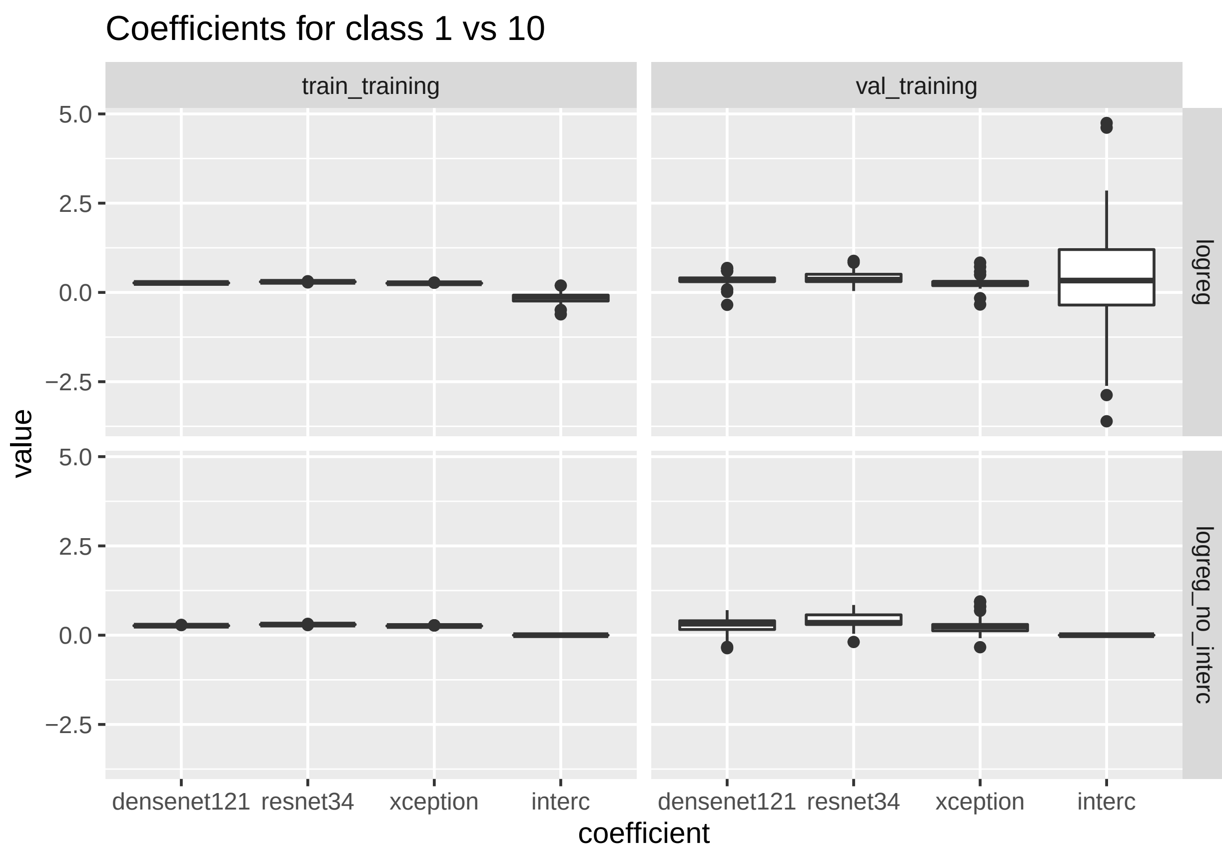

Coefficients for class 1 vs 6
train_training
val_training
logreg
logreg_no_interc
value
2
0
−2
−4
densenet121
resnet34
xception
interc
coefficient

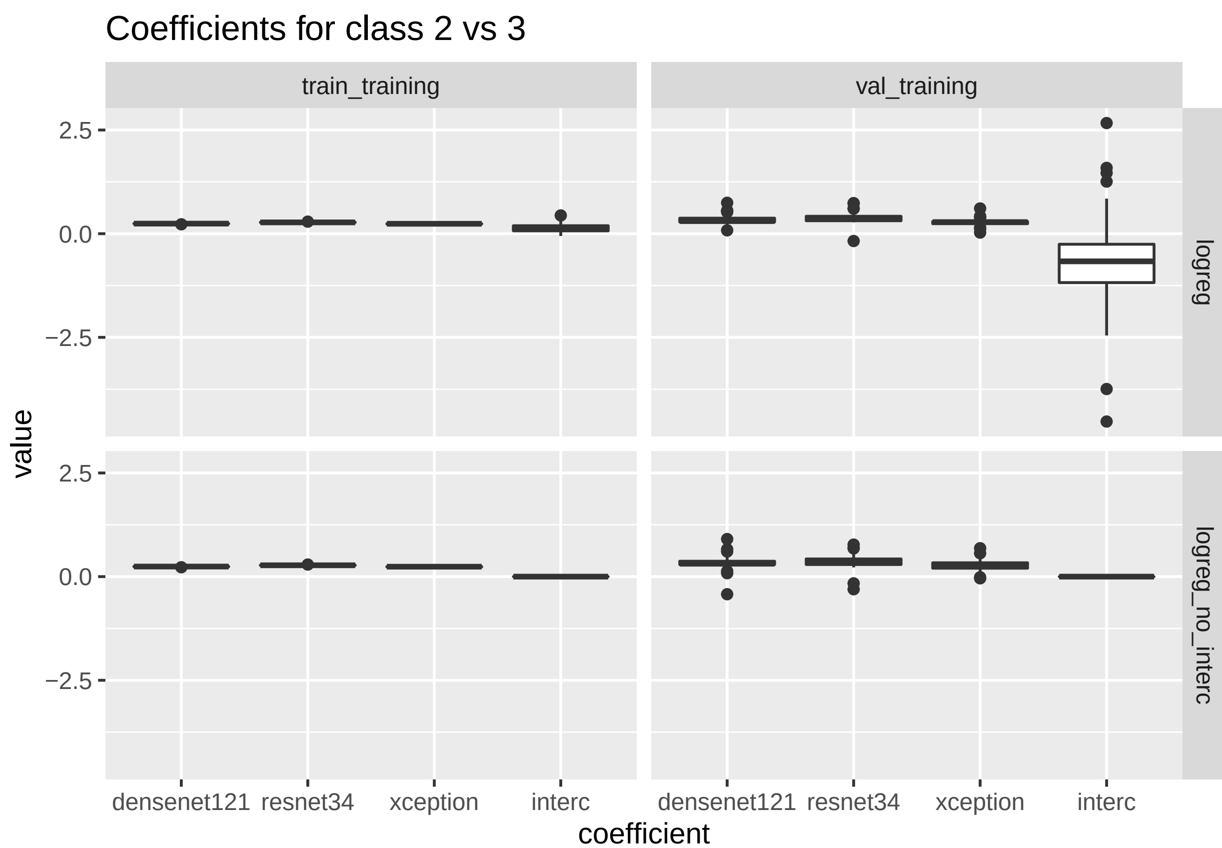

Coefficients for class 1 vs 7
train_training
val_training
logreg
logreg_no_interc
value
6
3
0
−3
densenet121
resnet34
xception
interc
coefficient


Coefficients for class 1 vs 8
train_training
val_training
logreg
logreg_no_interc
value
2
0
−2
−4
densenet121
resnet34
xception
interc
coefficient

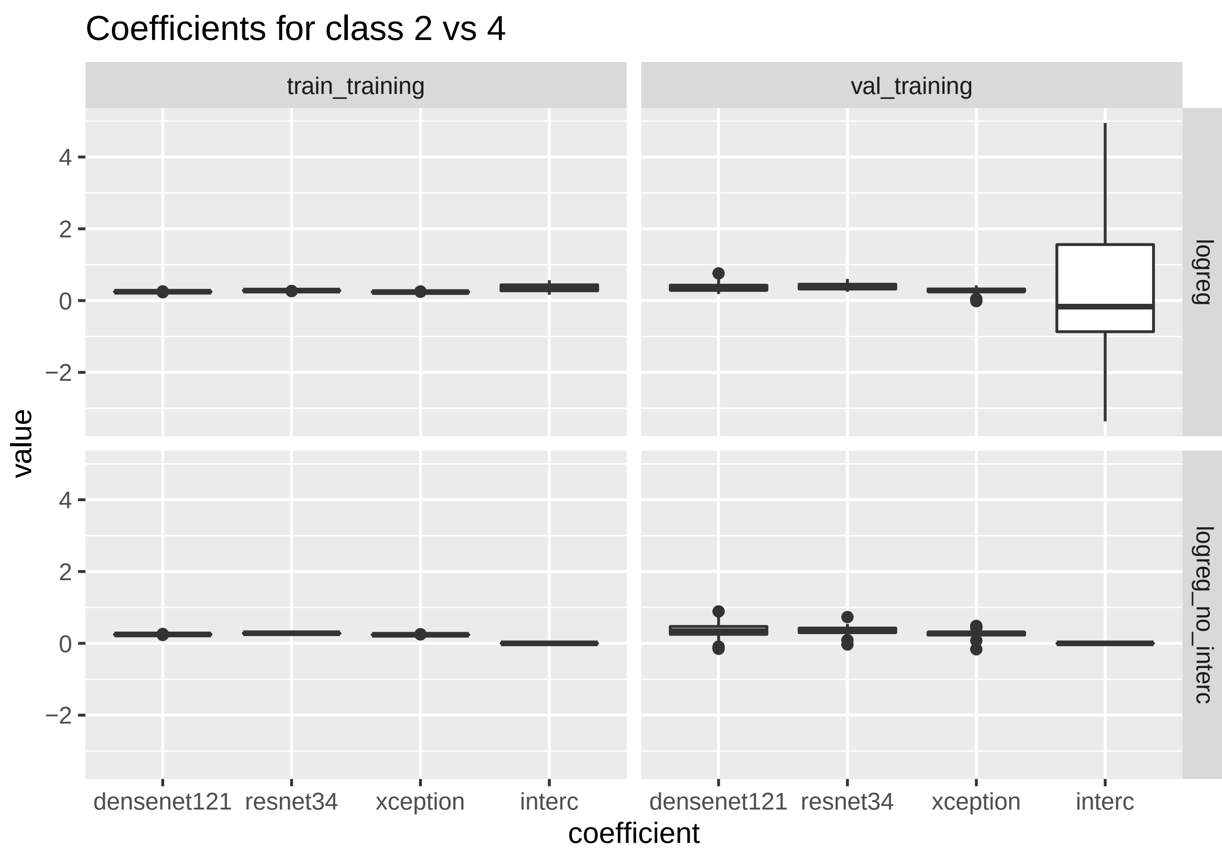


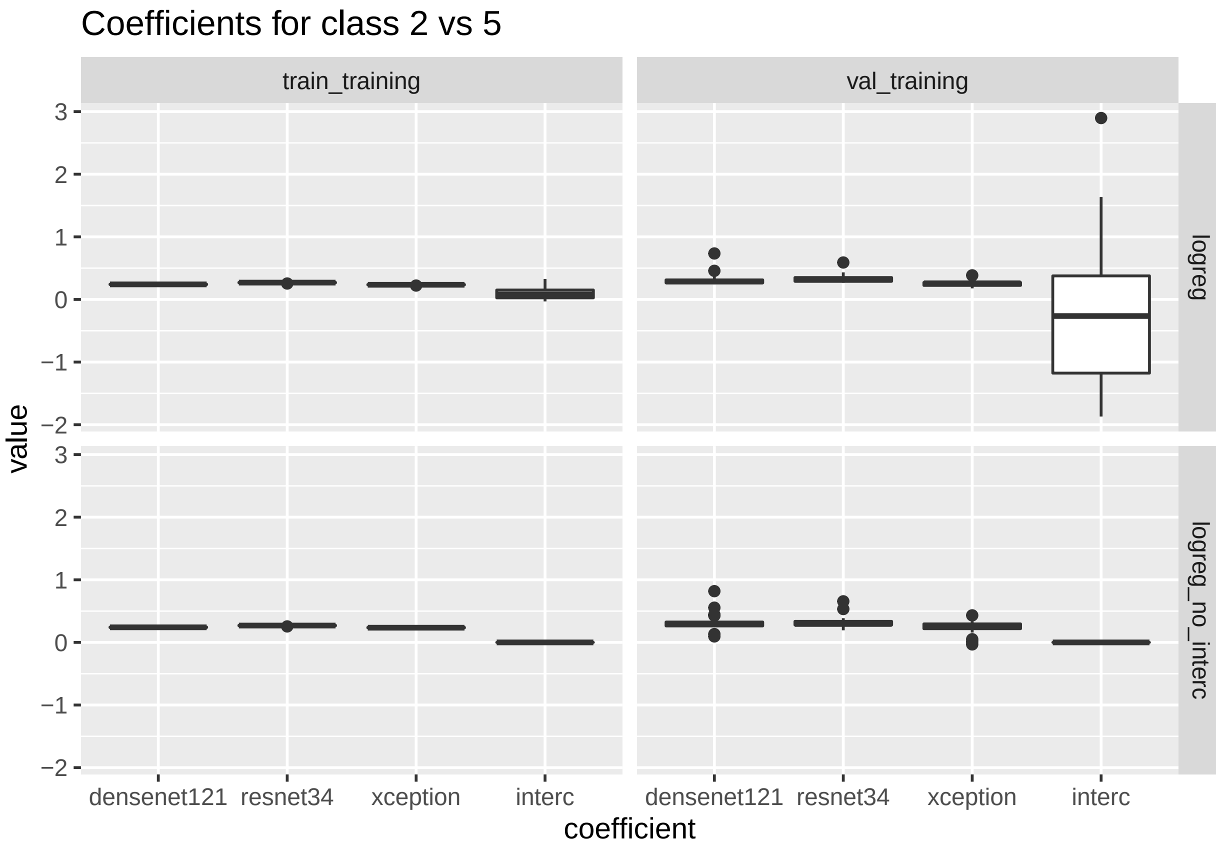
Coefficients for class 1 vs 9
train_training
val_training
logreg
logreg_no_interc
value
5.0
2.5
0.0
−2.5
−5.0
densenet121
resnet34
xception
interc
coefficient


Coefficients for class 1 vs 10
train_training
val_training
logreg
logreg_no_interc
value
5.0
2.5
0.0
−2.5
densenet121
resnet34
xception
interc
coefficient

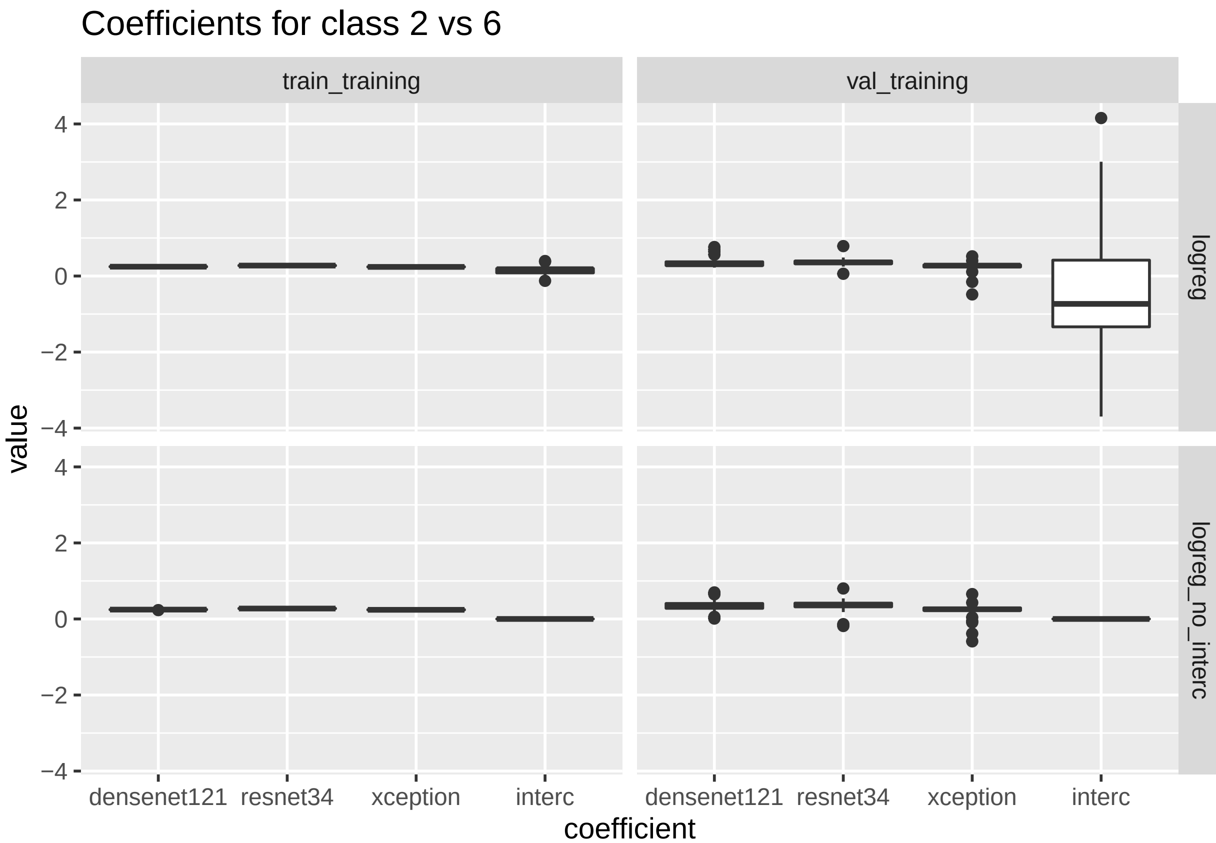


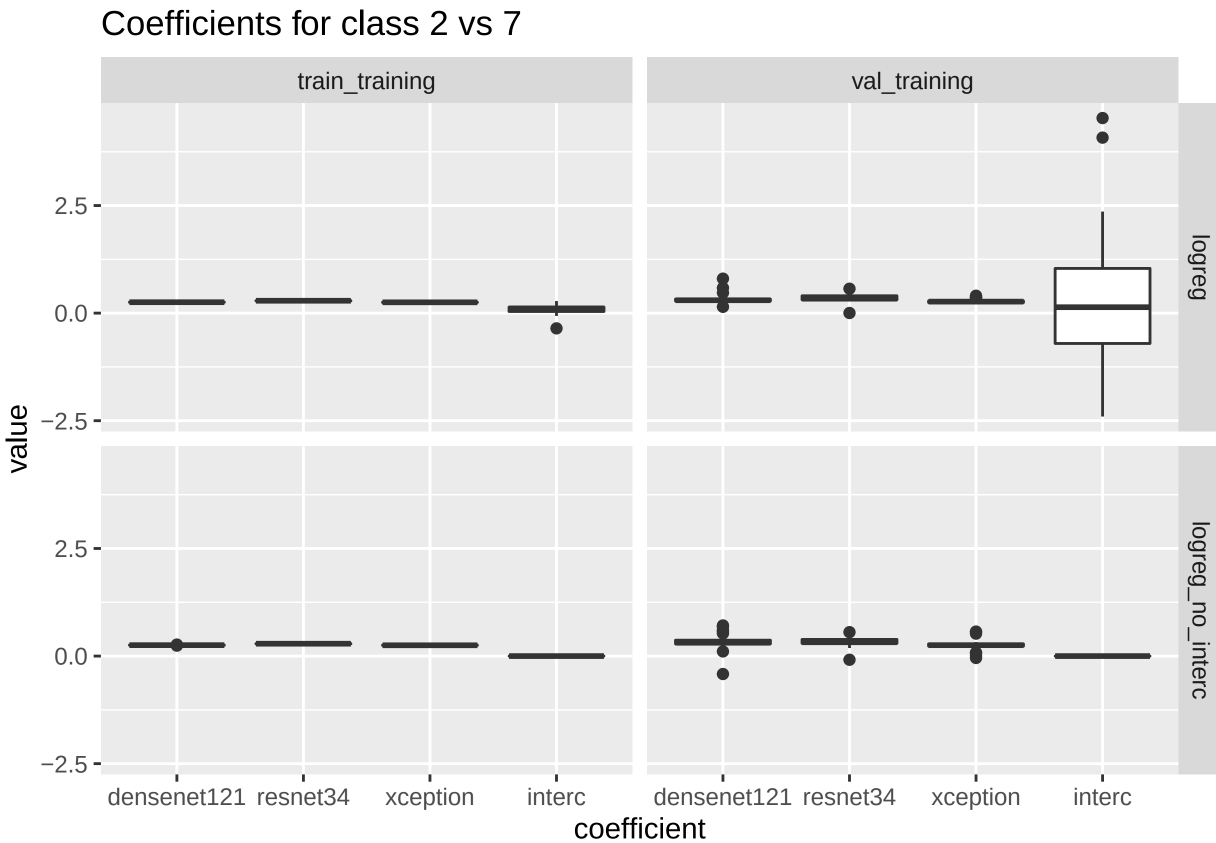
Coefficients for class 2 vs 3
train_training
val_training
logreg
logreg_no_interc
value
2.5
0.0
−2.5
densenet121
resnet34
xception
interc
coefficient

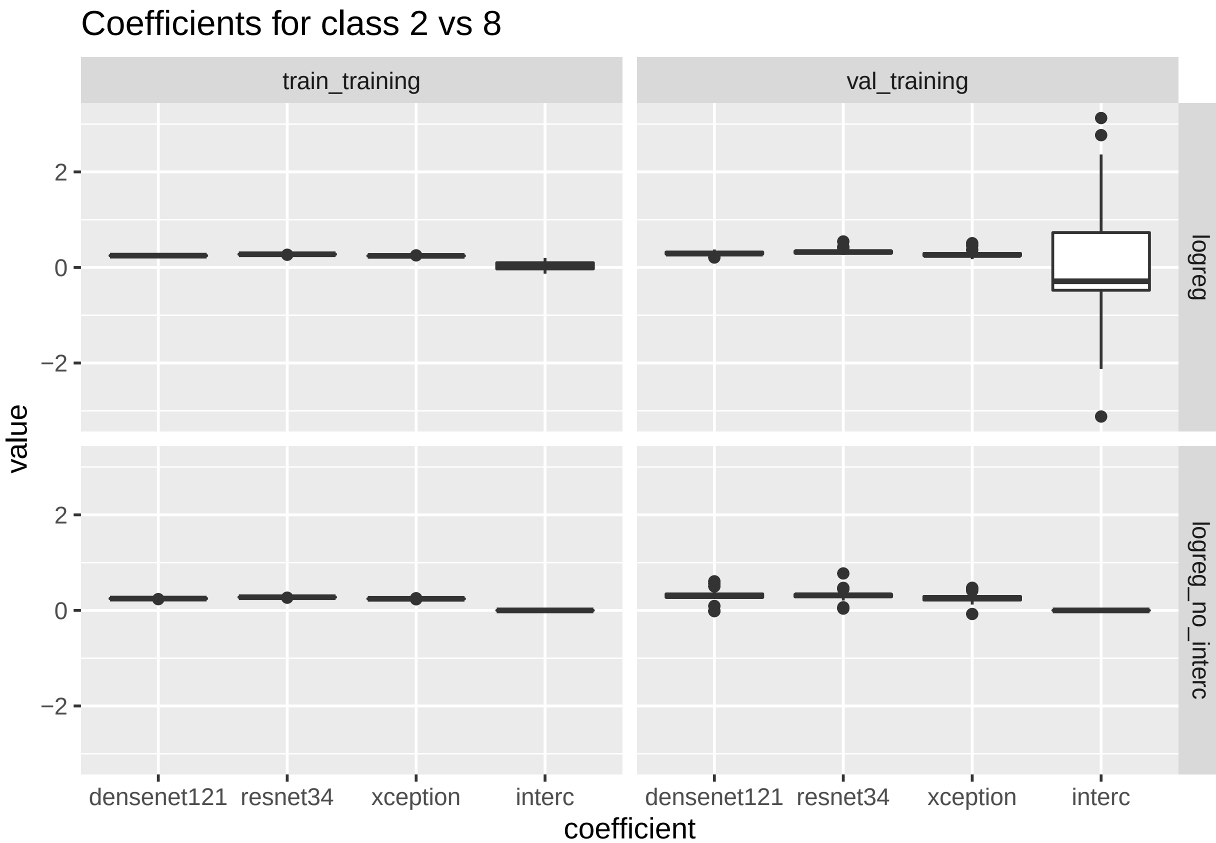

Coefficients for class 2 vs 4
train_training
val_training
logreg
logreg_no_interc
value
4
2
0
−2
densenet121
resnet34
xception
interc
coefficient

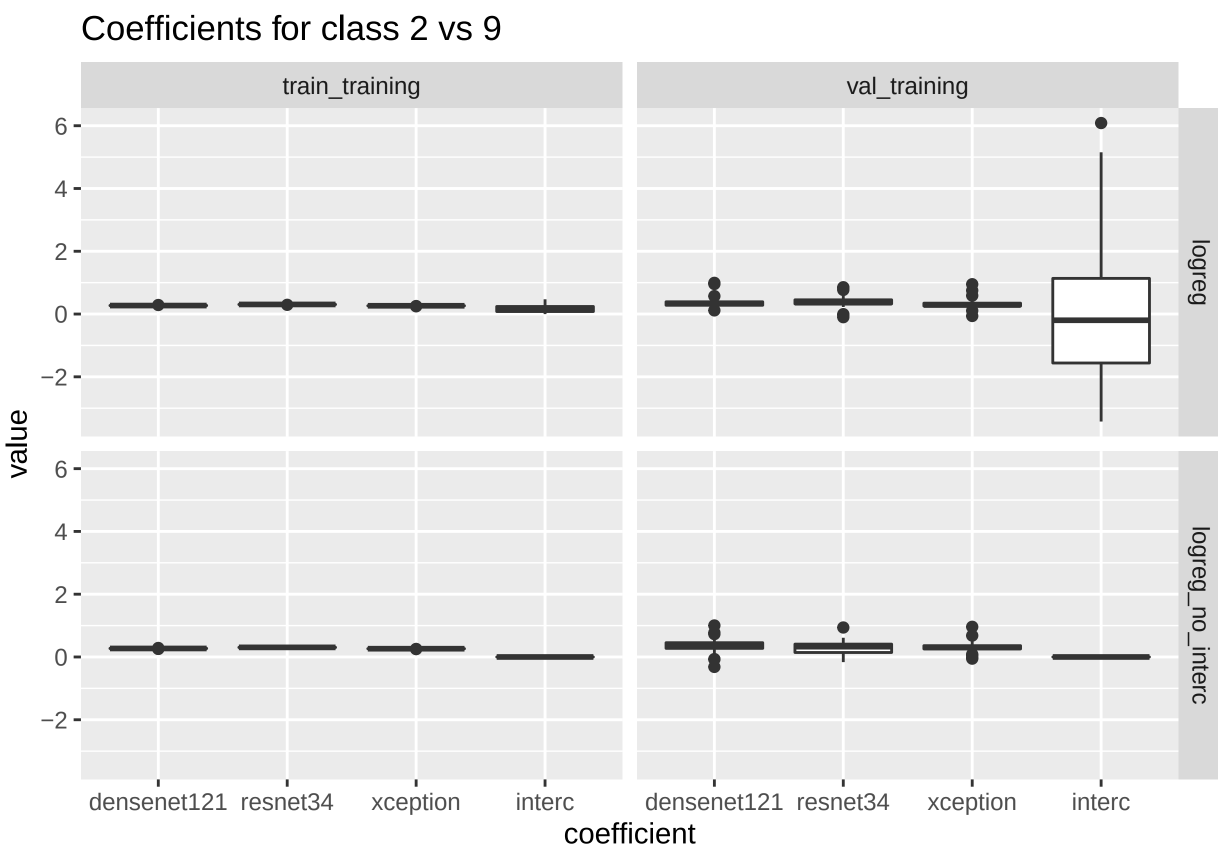


Coefficients for class 2 vs 5
train_training
val_training
logreg
logreg_no_interc
value
3
2
1
0
−1
−2
densenet121
resnet34
xception
interc
coefficient

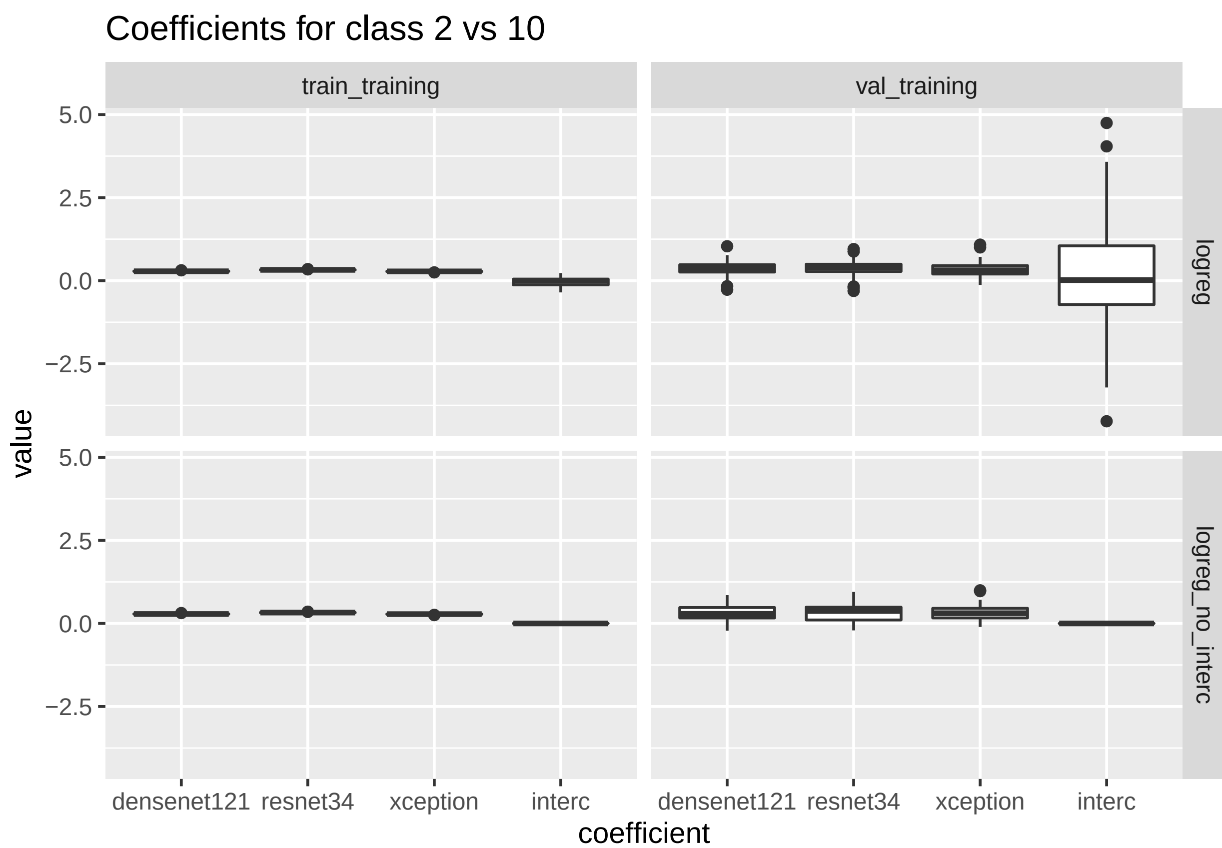

Coefficients for class 2 vs 6
train_training
val_training
logreg
logreg_no_interc
value
4
2
0
−2
−4
densenet121
resnet34
xception
interc
coefficient

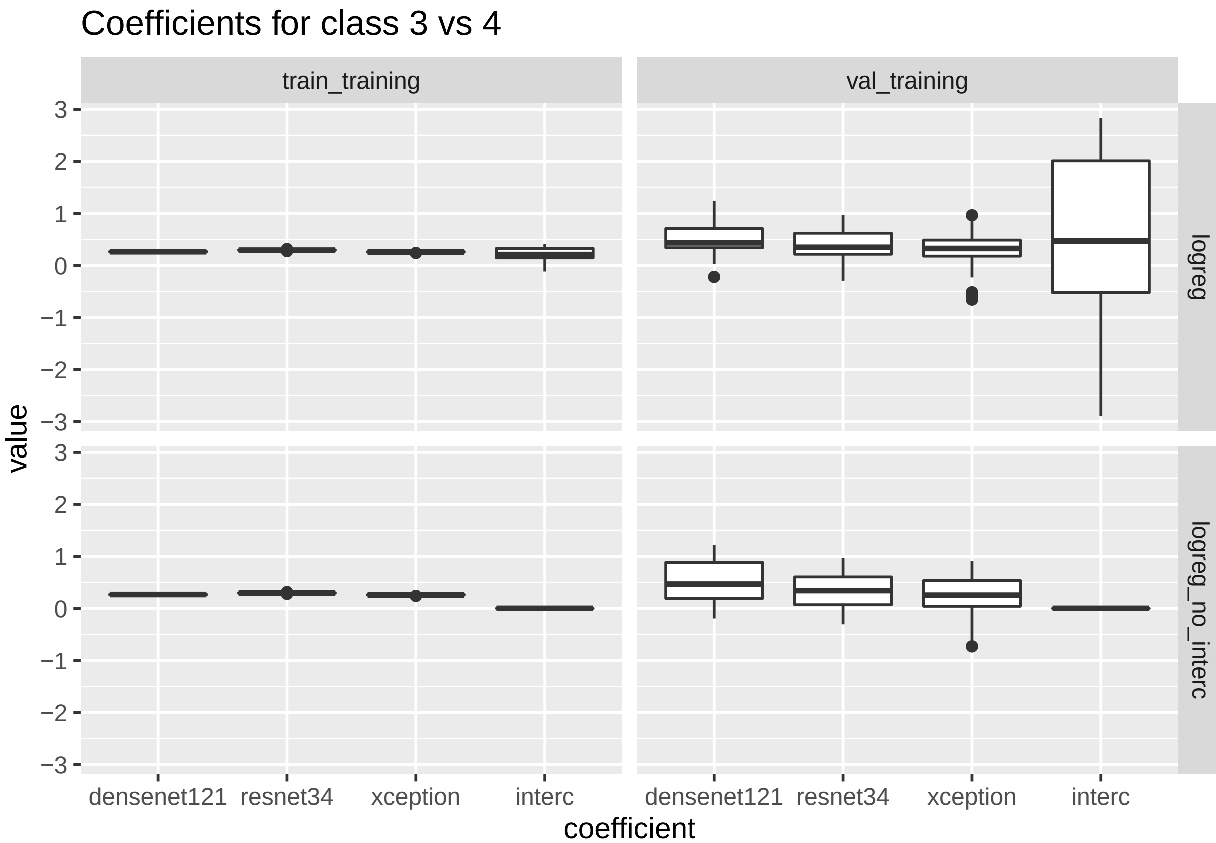

Coefficients for class 2 vs 7
train_training
val_training
logreg
logreg_no_interc
value
2.5
0.0
−2.5
2.5
0.0
−2.5
densenet121
resnet34
xception
interc
densenet121
resnet34
xception
interc
coefficient

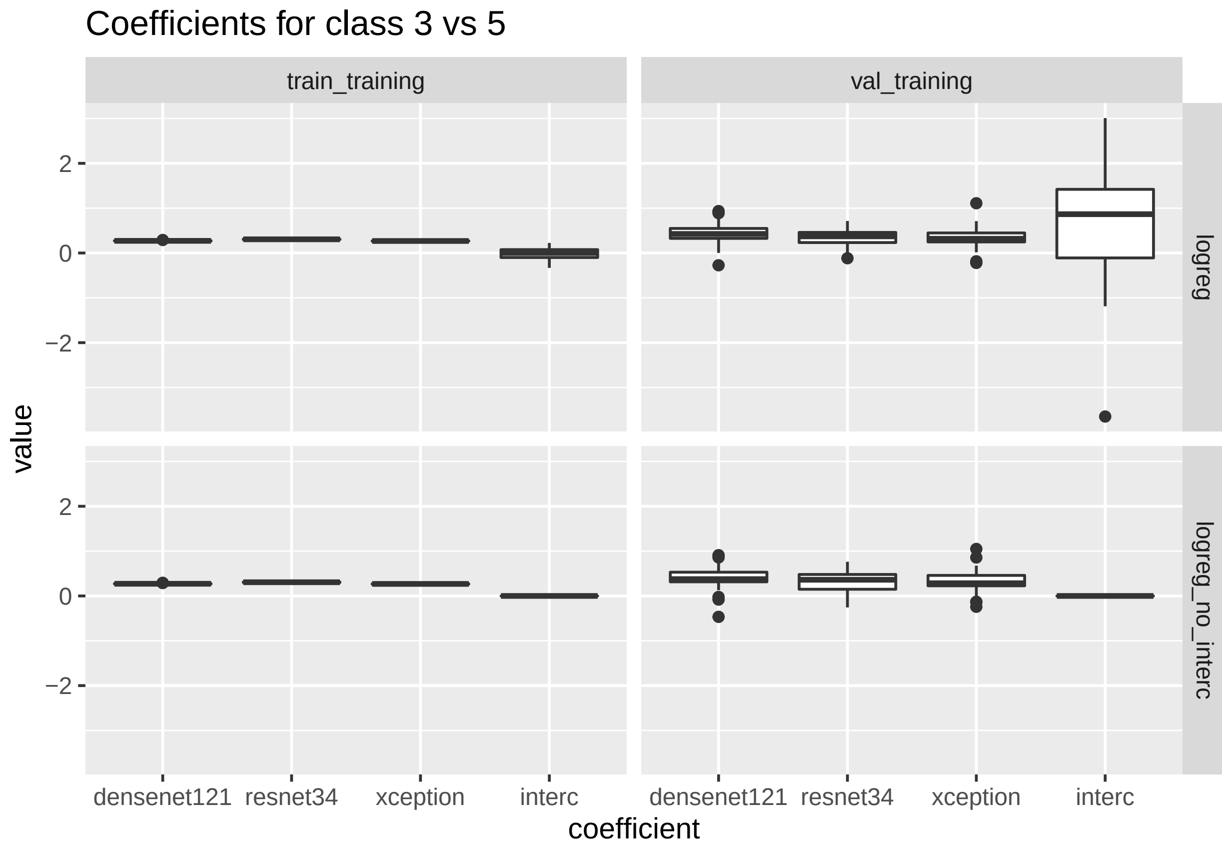

Coefficients for class 2 vs 8
train_training
val_training
logreg
logreg_no_interc
value
2
0
−2
2
0
−2
densenet121
resnet34
xception
interc
densenet121
resnet34
xception
interc
coefficient



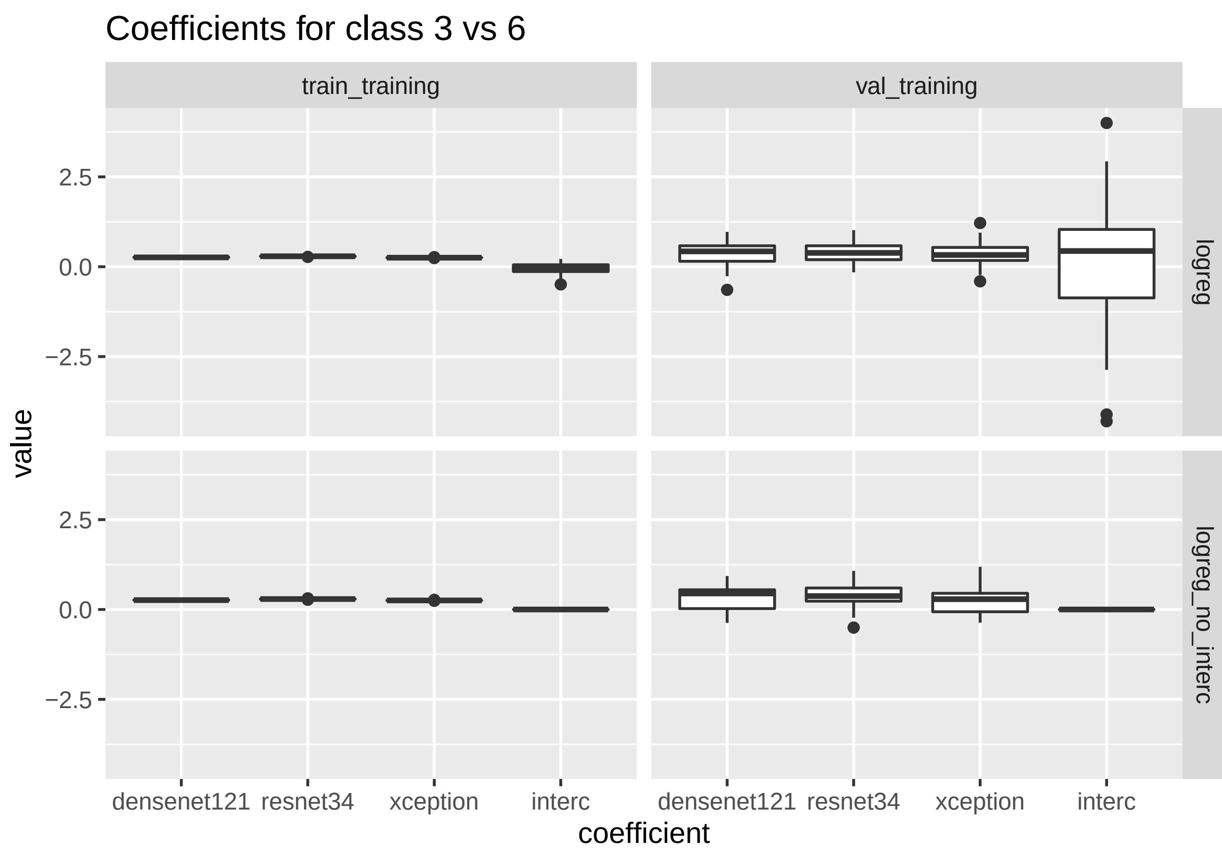
Coefficients for class 2 vs 9
train_training
val_training
logreg
logreg_no_interc
value
6
4
2
0
−2
densenet121
resnet34
xception
interc
coefficient

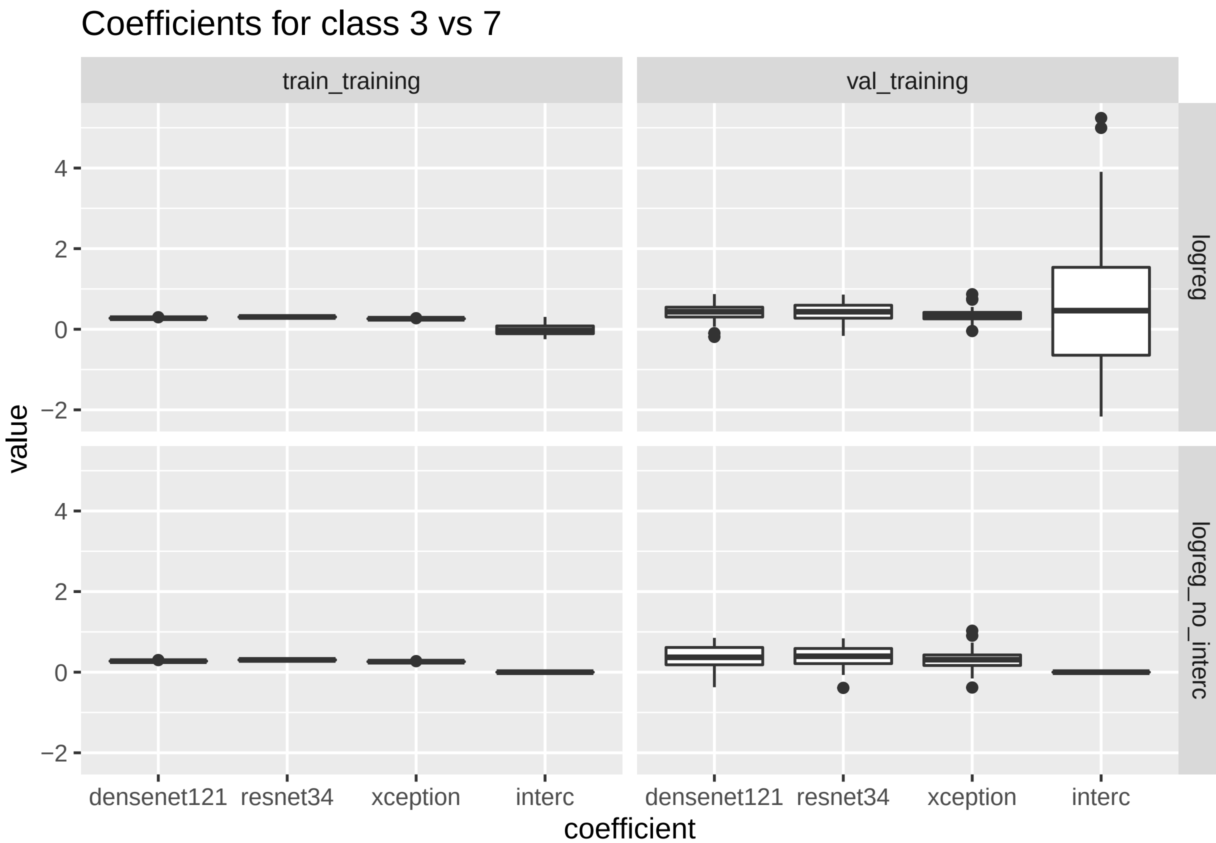

Coefficients for class 2 vs 10
train_training
val_training
logreg
logreg_no_interc
value
5.0
2.5
0.0
−2.5
densenet121
resnet34
xception
interc
coefficient

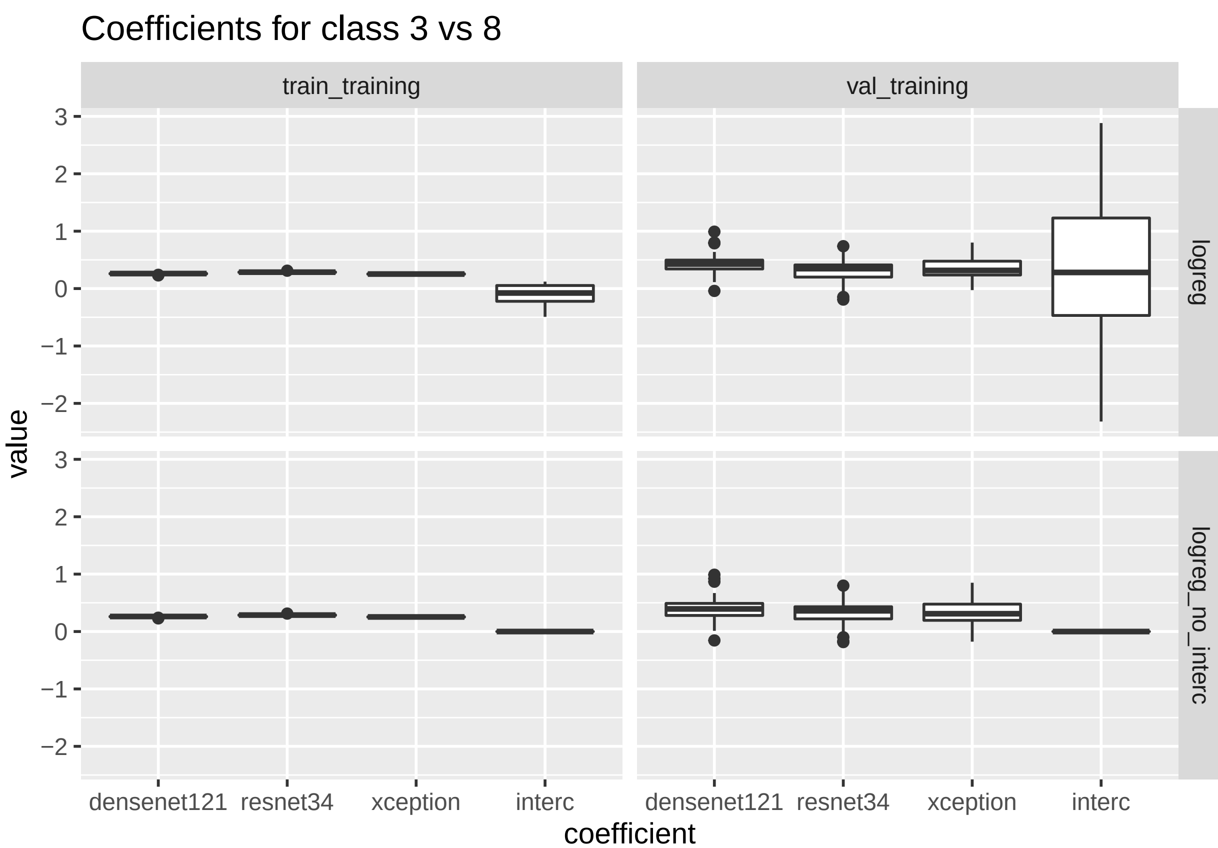

Coefficients for class 3 vs 4
train_training
val_training
logreg
logreg_no_interc
value
3
2
1
0
−1
−2
−3
densenet121
resnet34
xception
interc
coefficient

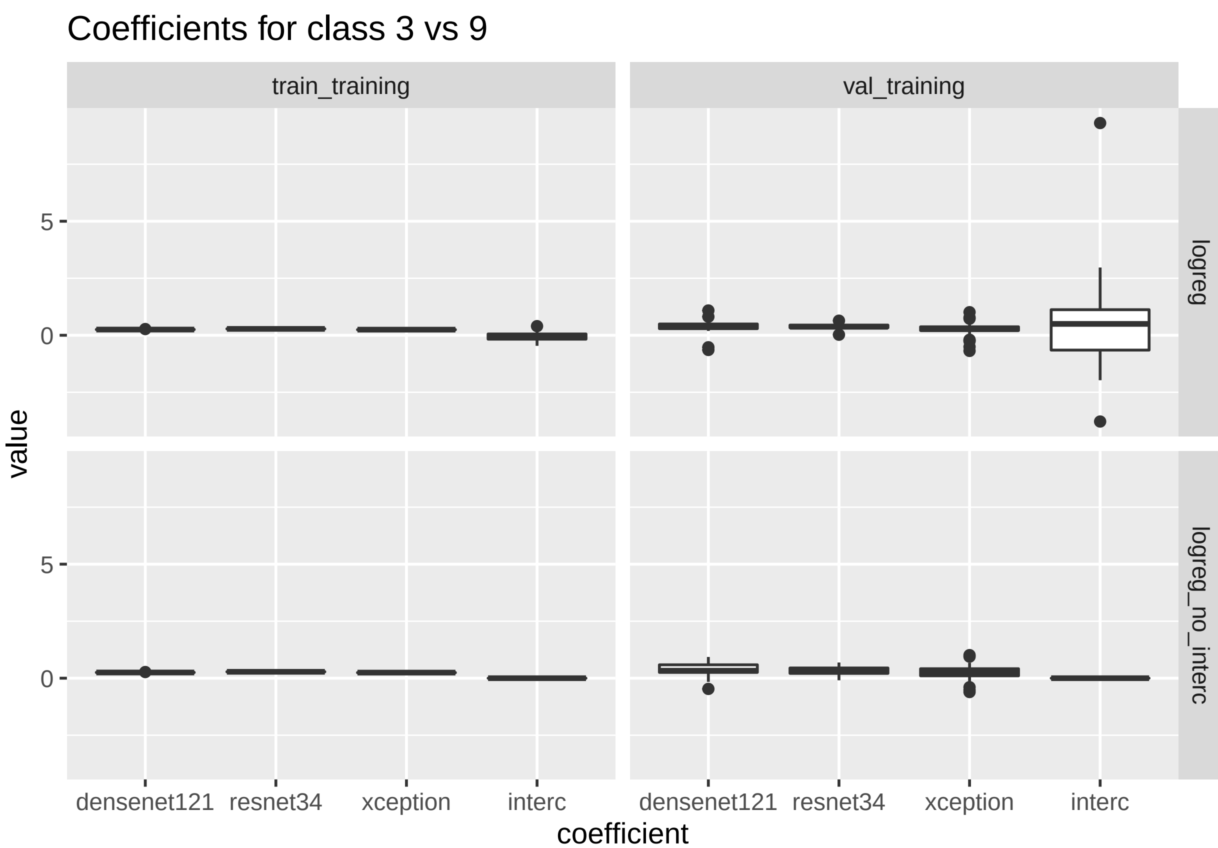


Coefficients for class 3 vs 5
train_training
val_training
logreg
logreg_no_interc
value
2
0
−2
2
0
−2
densenet121
resnet34
xception
interc
densenet121
resnet34
xception
interc
coefficient

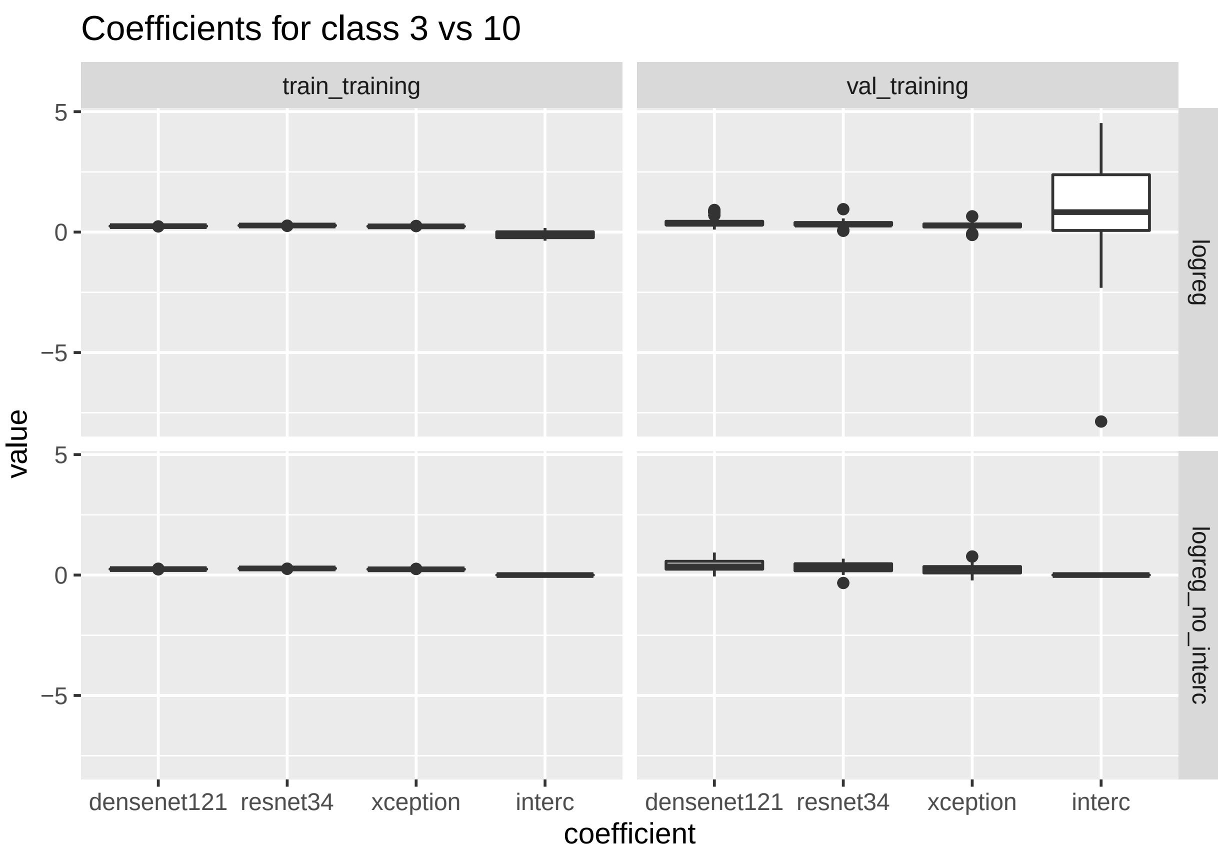

Coefficients for class 3 vs 6
train_training
val_training
logreg
logreg_no_interc
value
2.5
0.0
−2.5
2.5
0.0
−2.5
densenet121
resnet34
xception
interc
densenet121
resnet34
xception
interc
coefficient

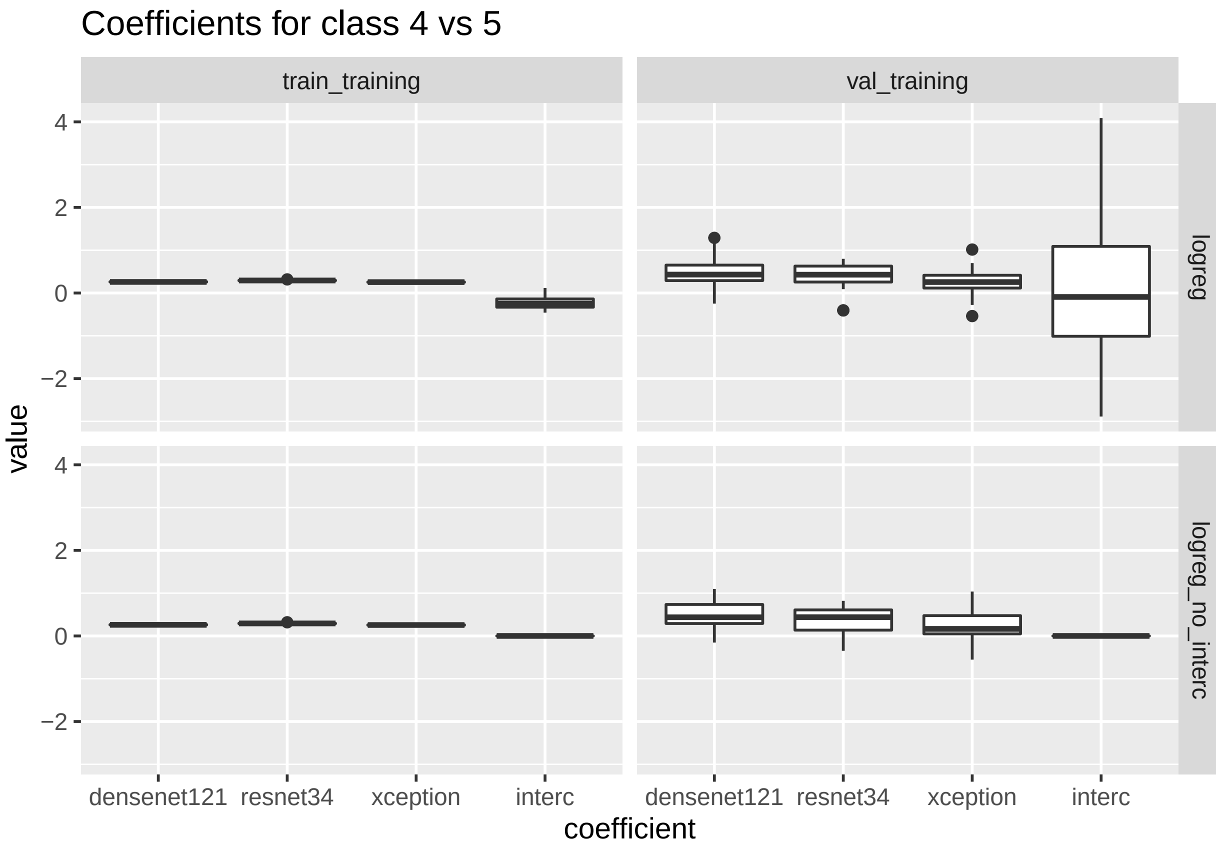

Coefficients for class 3 vs 7
train_training
val_training
logreg
logreg_no_interc
value
4
2
0
−2
densenet121
resnet34
xception
interc
coefficient


Coefficients for class 3 vs 8
train_training
val_training
logreg
logreg_no_interc
value
3
2
1
0
−1
−2
densenet121
resnet34
xception
interc
coefficient

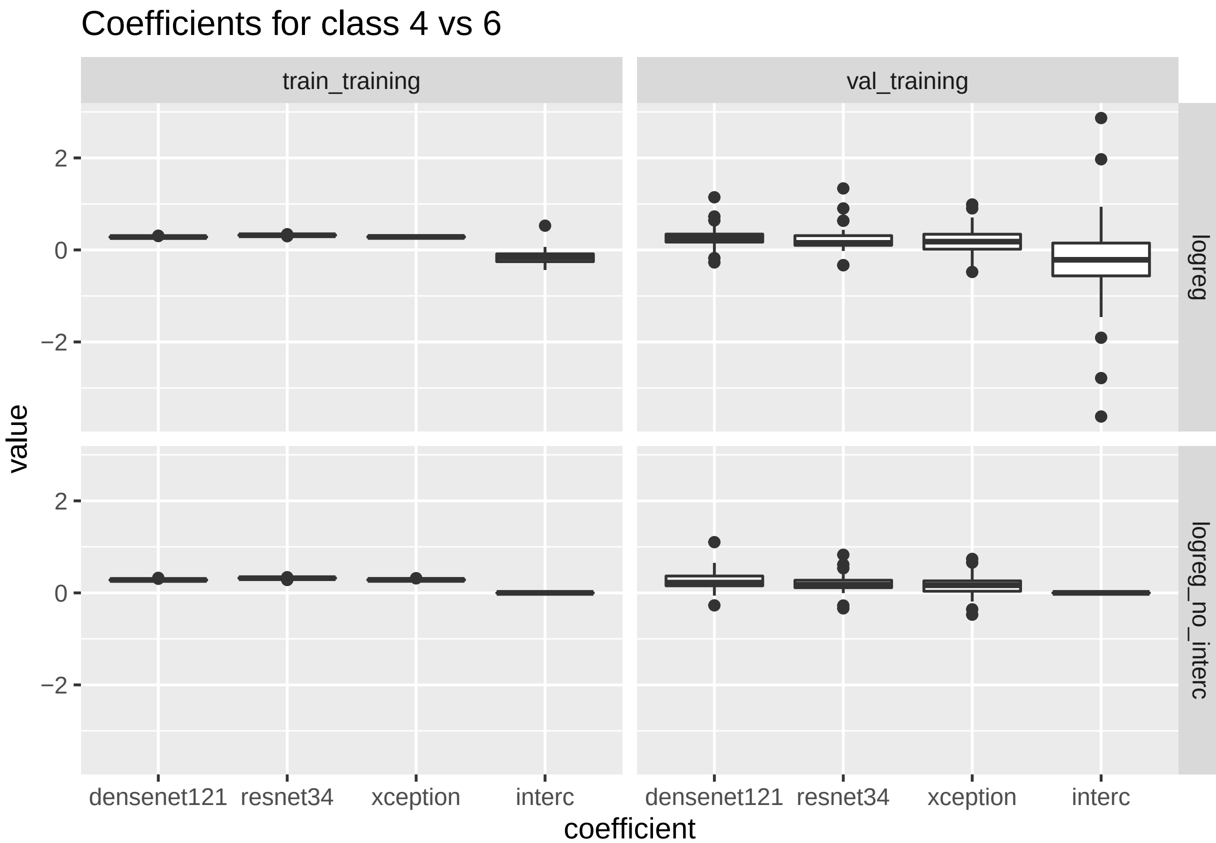


Coefficients for class 3 vs 9
train_training
val_training
logreg
logreg_no_interc
value
5
0
5
0
densenet121
resnet34
xception
interc
densenet121
resnet34
xception
interc
coefficient

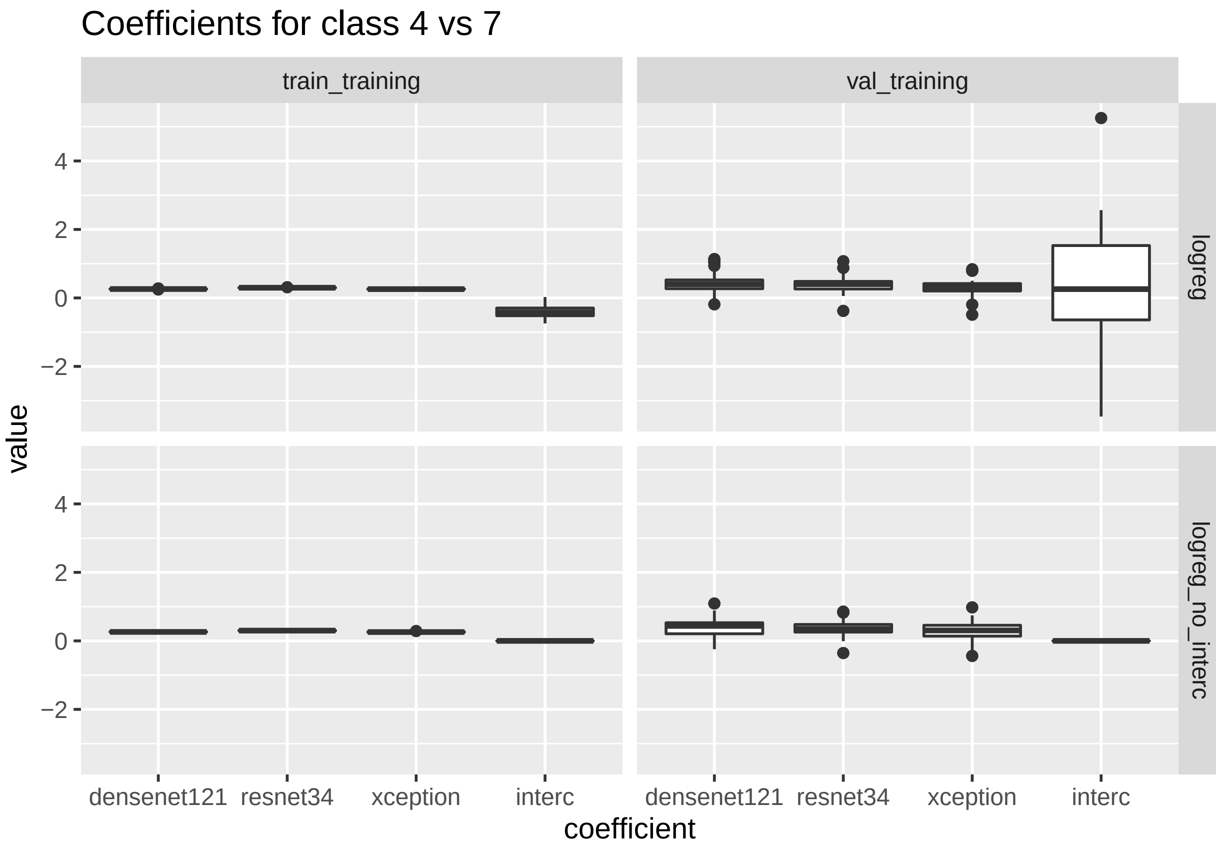

Coefficients for class 3 vs 10
train_training
val_training
logreg
logreg_no_interc
value
5
0
−5
5
0
−5
densenet121
resnet34
xception
interc
densenet121
resnet34
xception
interc
coefficient



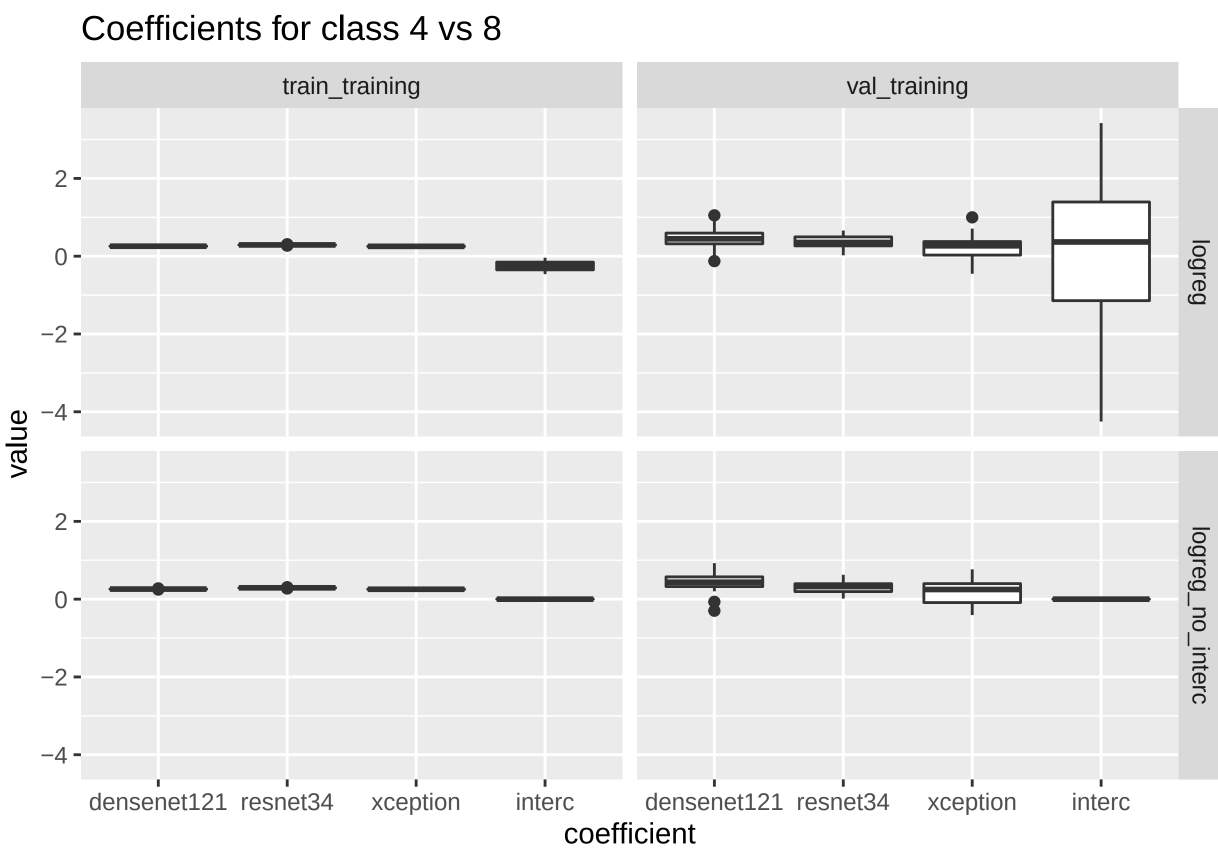
Coefficients for class 4 vs 5
train_training
val_training
logreg
logreg_no_interc
value
4
2
0
−2
densenet121
resnet34
xception
interc
coefficient



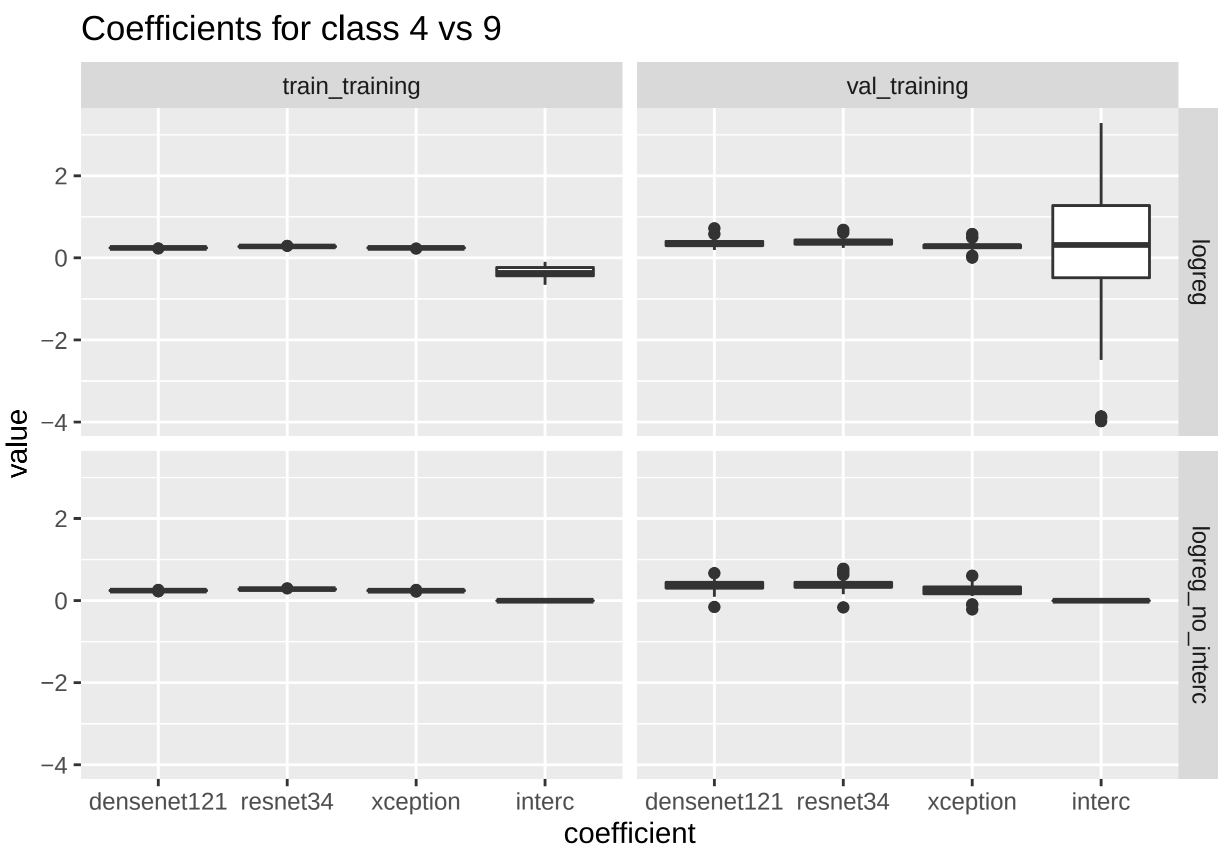
Coefficients for class 4 vs 6
train_training
val_training
logreg
logreg_no_interc
value
2
0
−2
densenet121
resnet34
xception
interc
coefficient

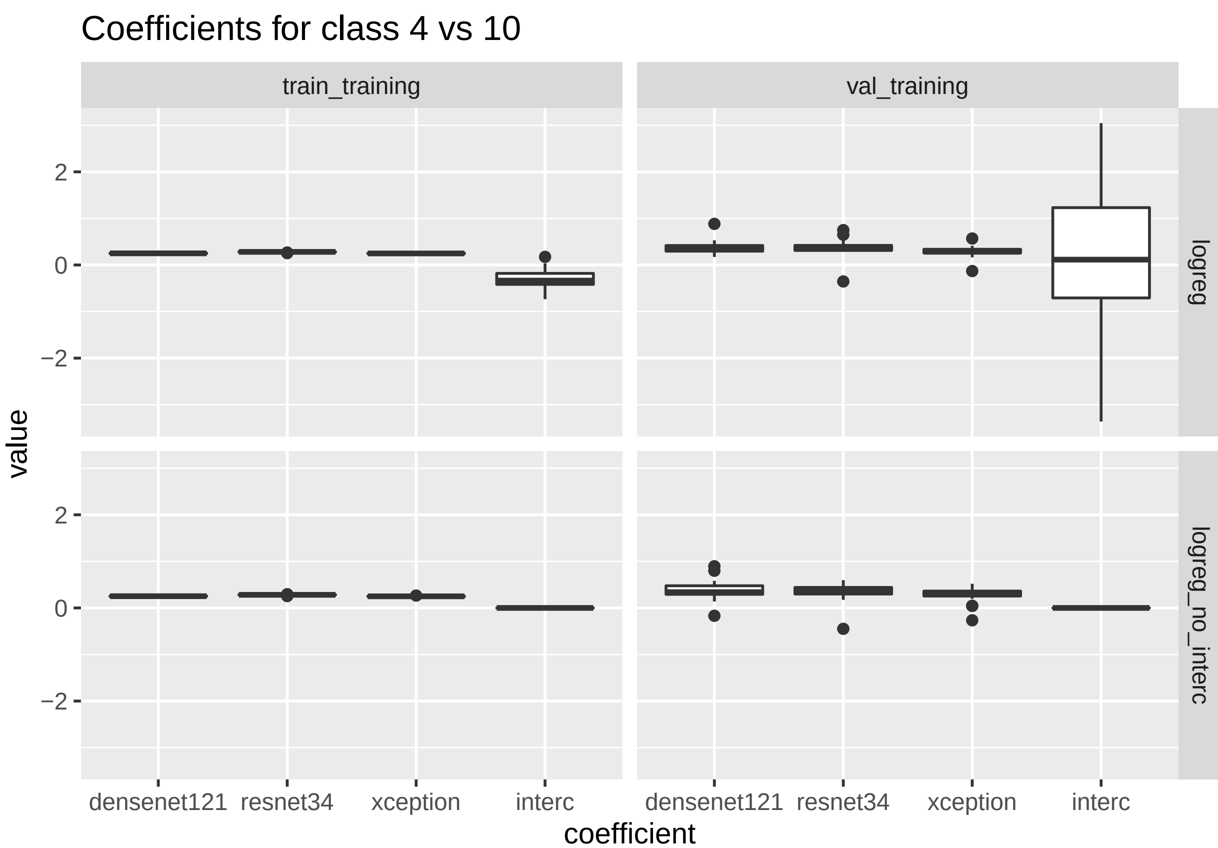


Coefficients for class 4 vs 7
train_training
val_training
logreg
logreg_no_interc
value
4
2
0
−2
4
2
0
−2
densenet121
resnet34
xception
interc
densenet121
resnet34
xception
interc
coefficient

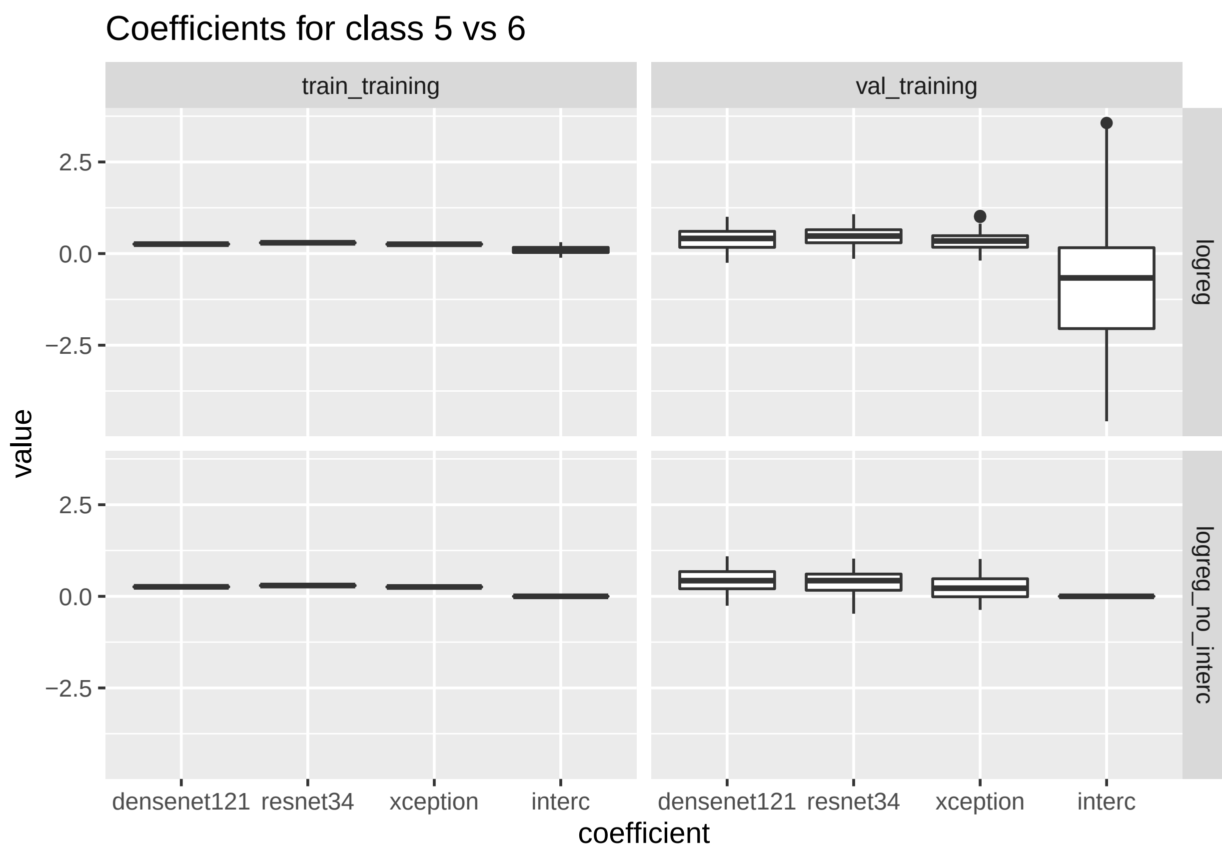


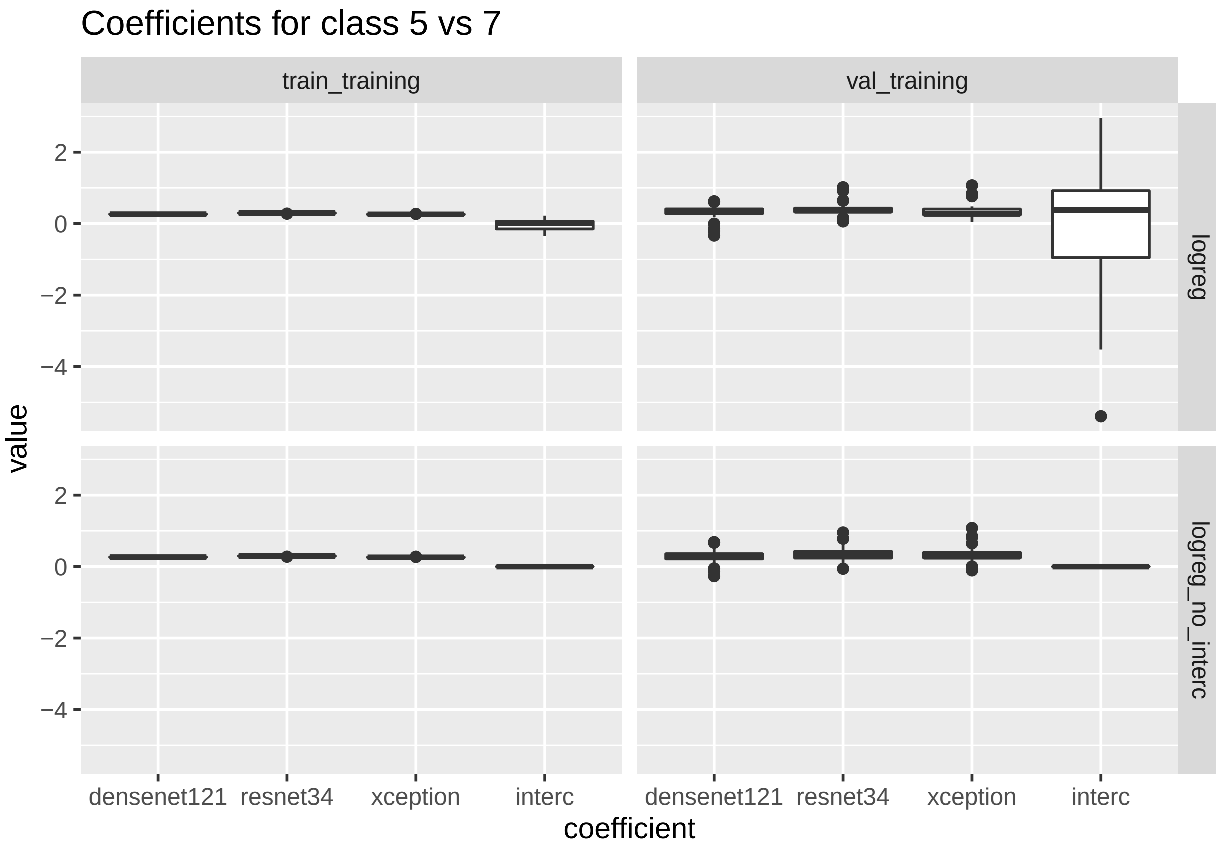
Coefficients for class 4 vs 8
train_training
val_training
logreg
logreg_no_interc
value
2
0
−2
−4
densenet121
resnet34
xception
interc
coefficient

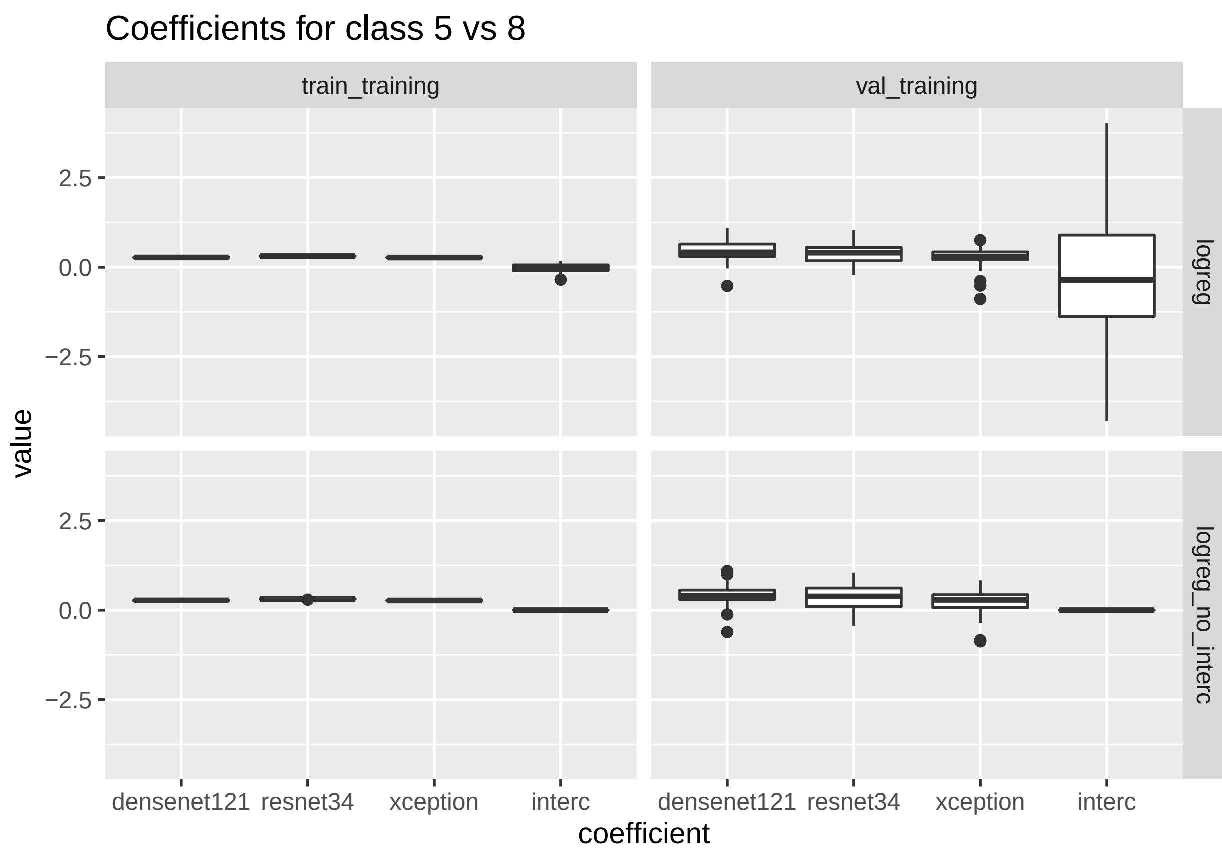

Coefficients for class 4 vs 9
train_training
val_training
logreg
logreg_no_interc
value
2
0
−2
−4
densenet121
resnet34
xception
interc
coefficient

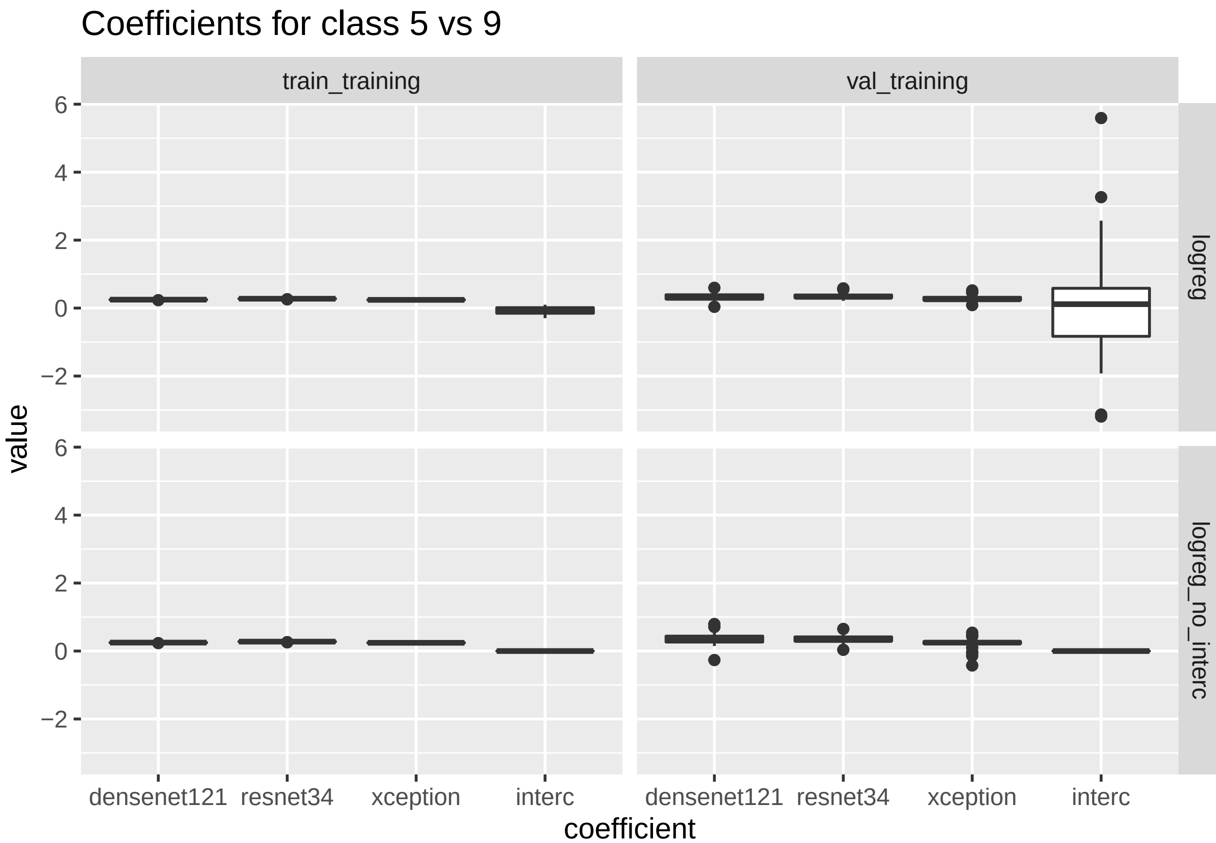


Coefficients for class 4 vs 10
train_training
val_training
logreg
logreg_no_interc
value
2
0
−2
2
0
−2
densenet121
resnet34
xception
interc
densenet121
resnet34
xception
interc
coefficient

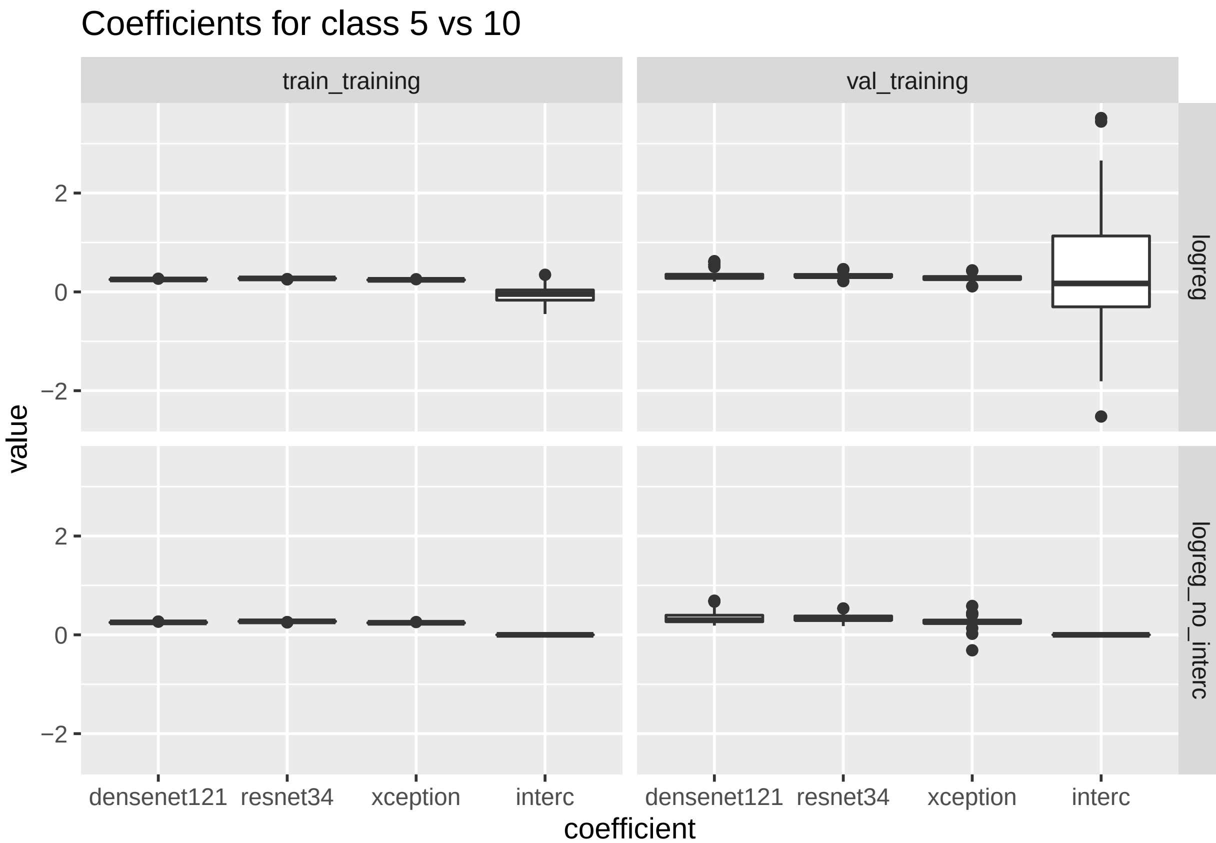

Coefficients for class 5 vs 6
train_training
val_training
logreg
logreg_no_interc
value
2.5
0.0
−2.5
densenet121
resnet34
xception
interc
coefficient

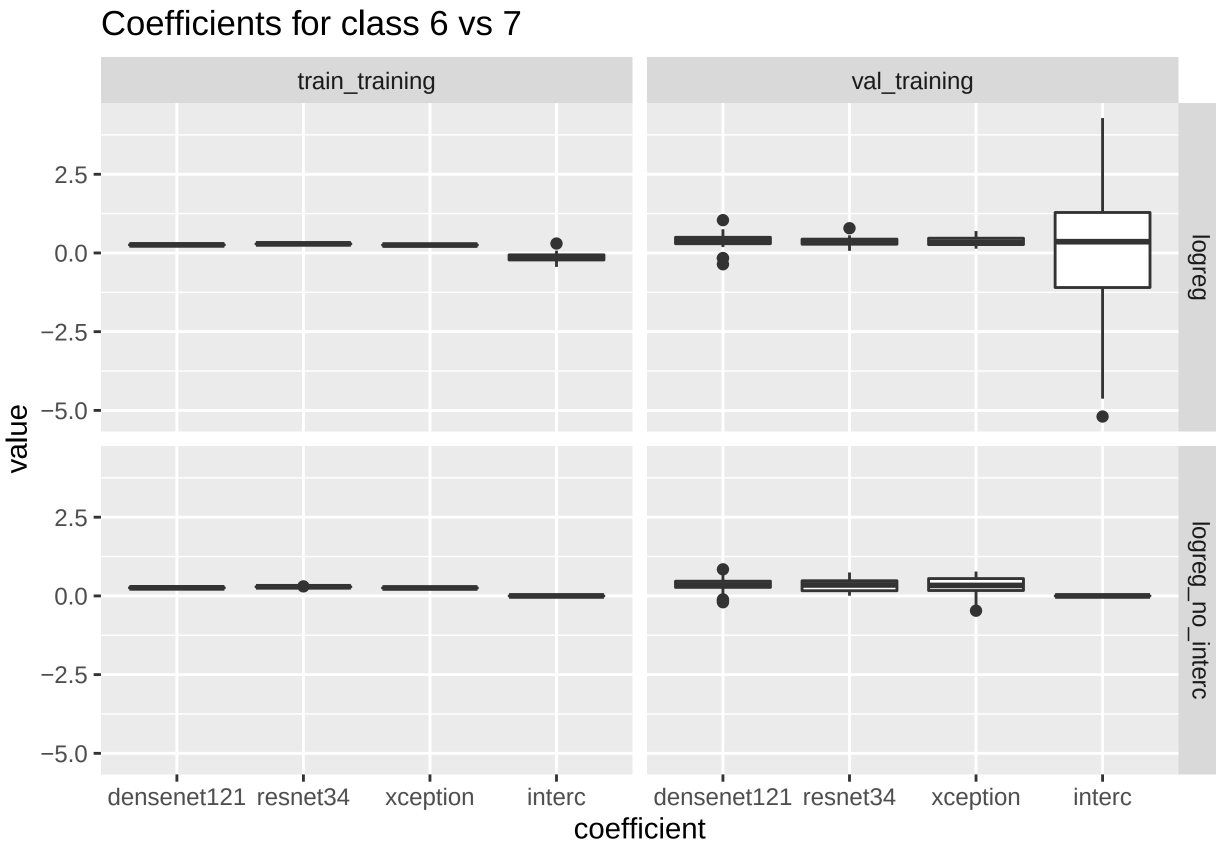

Coefficients for class 5 vs 7
train_training
val_training
logreg
logreg_no_interc
value
2
0
−2
−4
densenet121
resnet34
xception
interc
coefficient


Coefficients for class 5 vs 8
train_training
val_training
logreg
logreg_no_interc
value
2.5
0.0
−2.5
densenet121
resnet34
xception
interc
coefficient

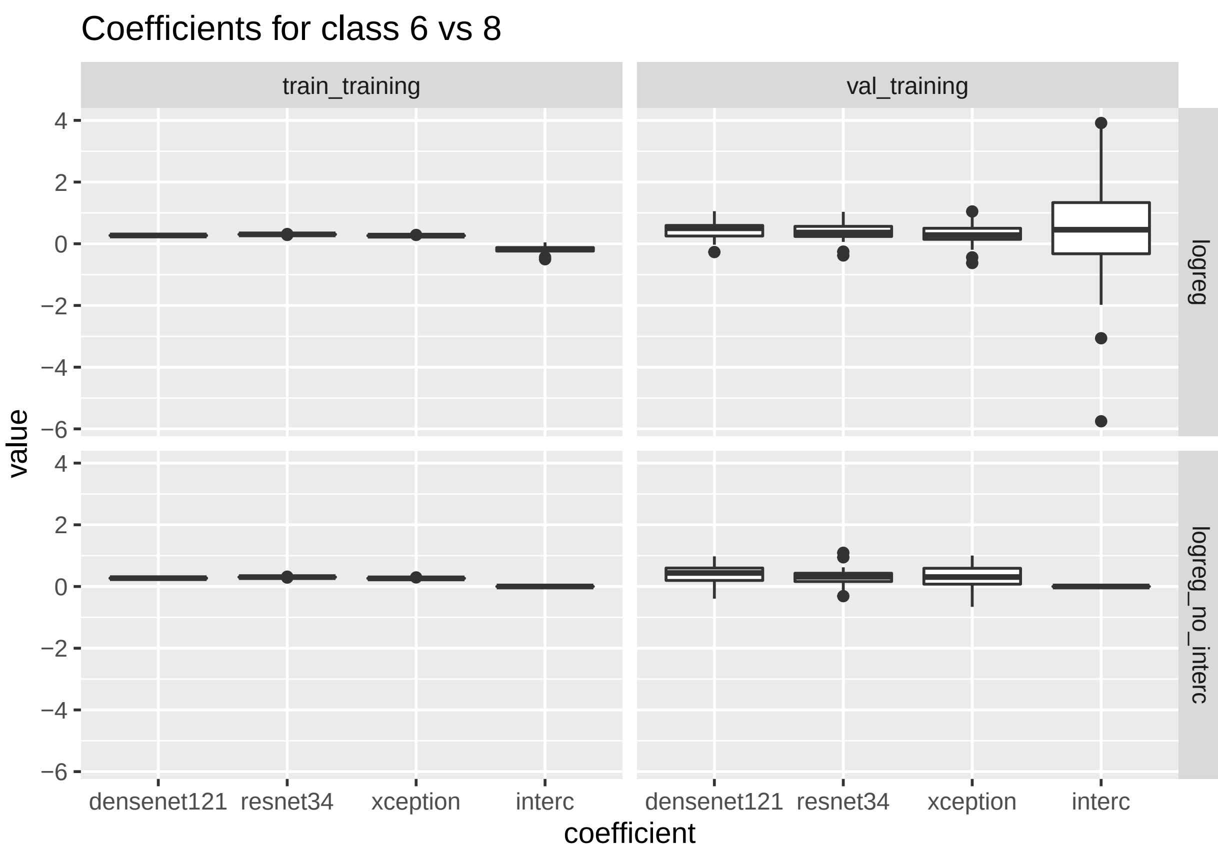


Coefficients for class 5 vs 9
train_training
val_training
logreg
logreg_no_interc
value
6
4
2
0
−2
densenet121
resnet34
xception
interc
coefficient

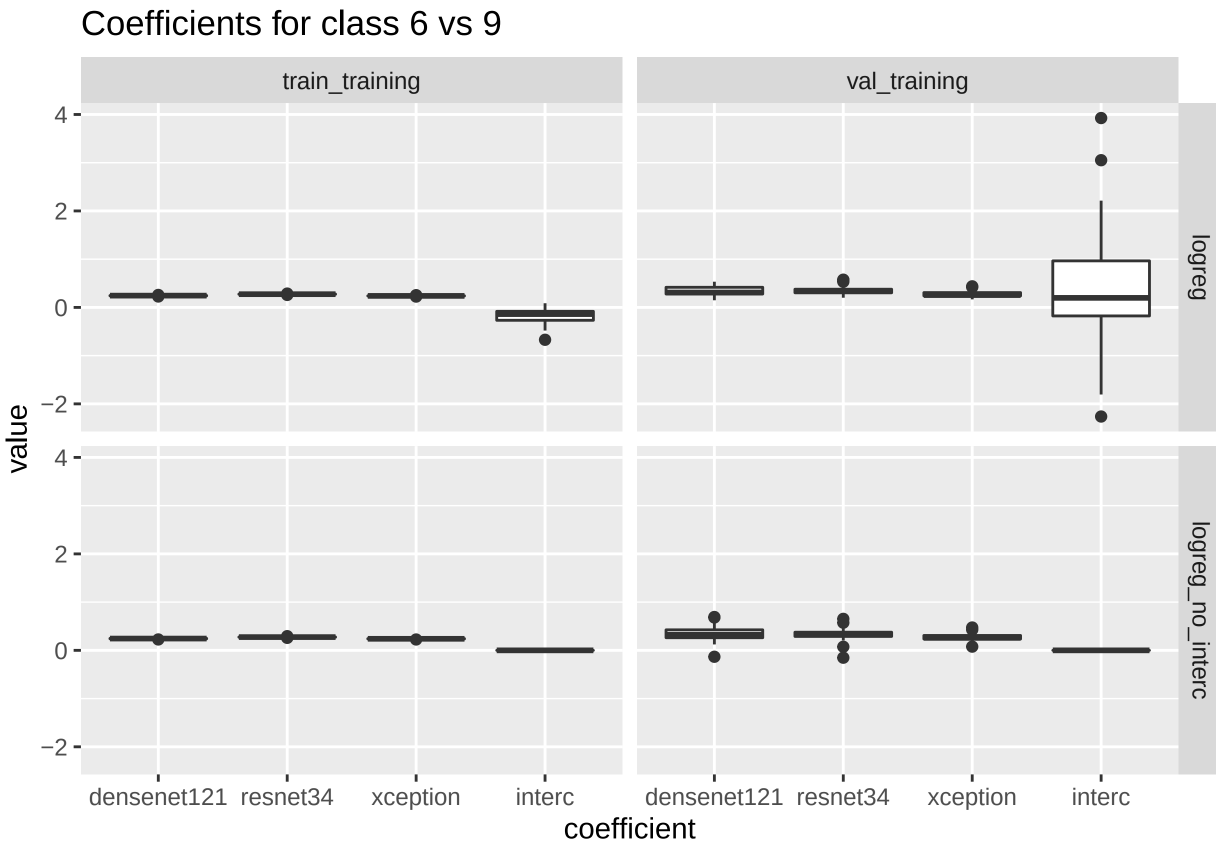

Coefficients for class 5 vs 10
train_training
val_training
logreg
logreg_no_interc
value
2
0
−2
2
0
−2
densenet121
resnet34
xception
interc
densenet121
resnet34
xception
interc
coefficient



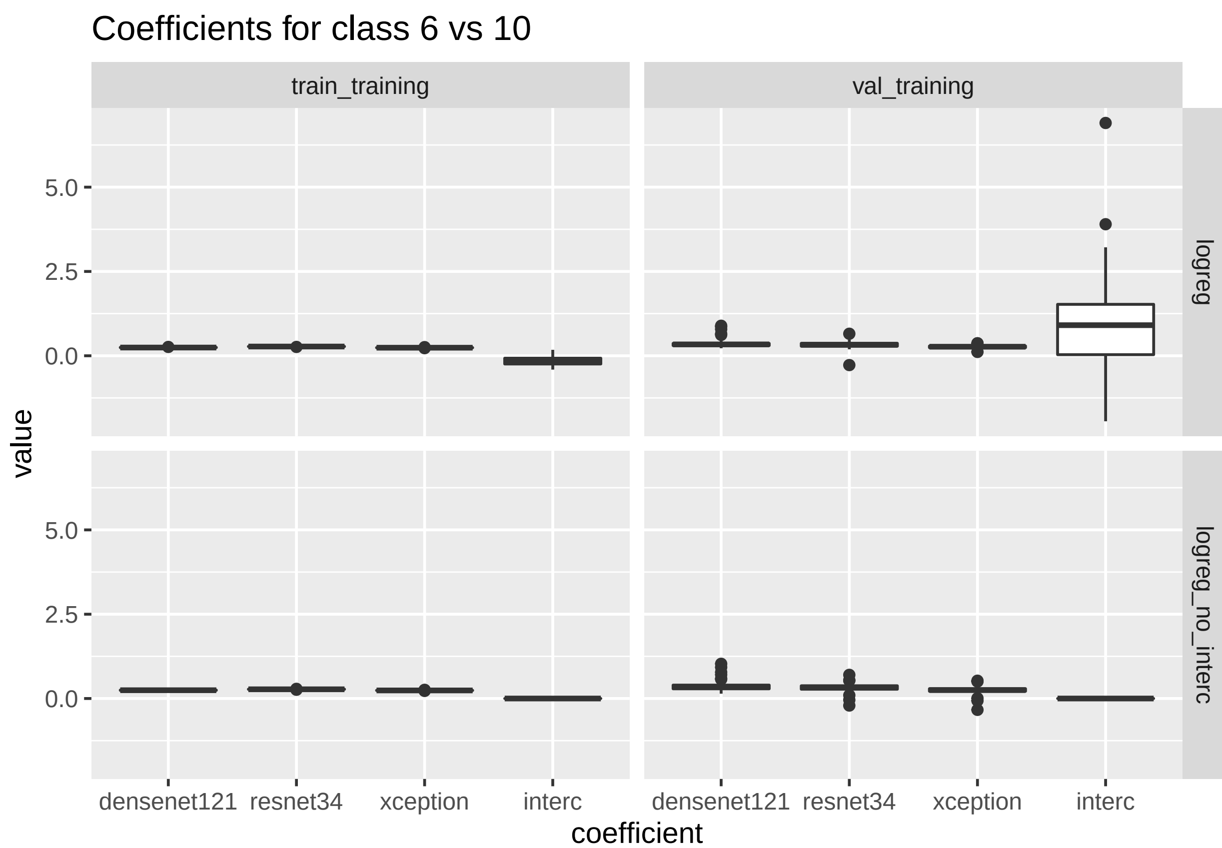
Coefficients for class 6 vs 7
train_training
val_training
logreg
logreg_no_interc
2.5
0.0
−2.5
−5.0
value
densenet121
resnet34
xception
interc
coefficient

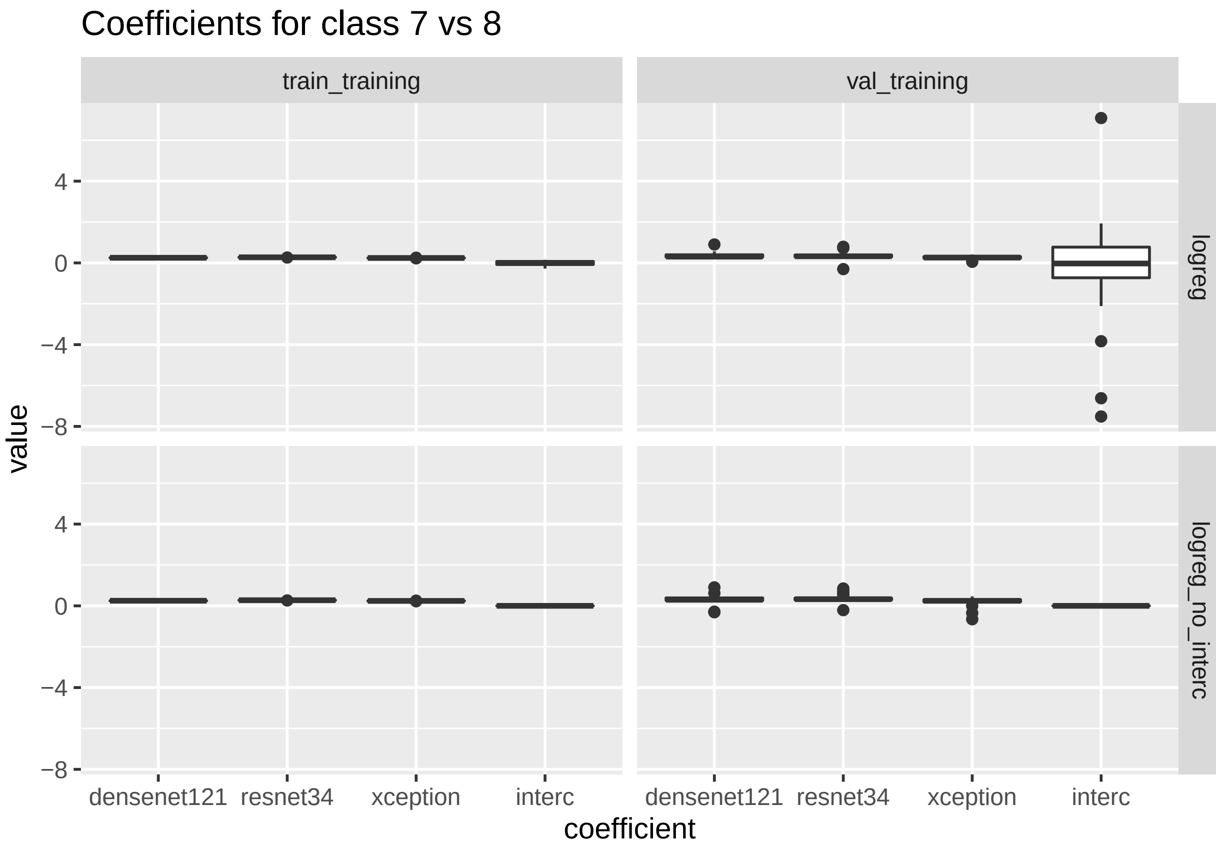


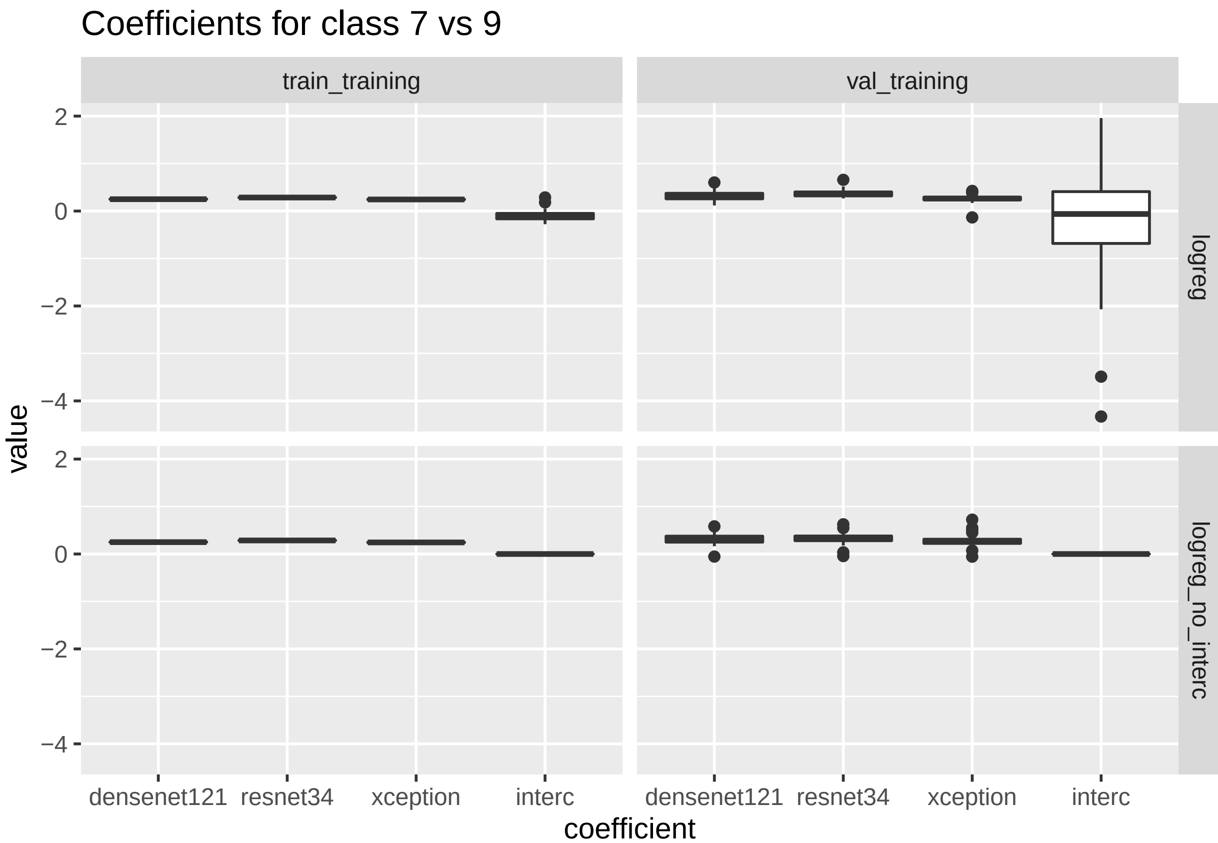
Coefficients for class 6 vs 8
train_training
val_training
logreg
logreg_no_interc
value
4
2
0
−2
−4
−6
densenet121
resnet34
xception
interc
coefficient

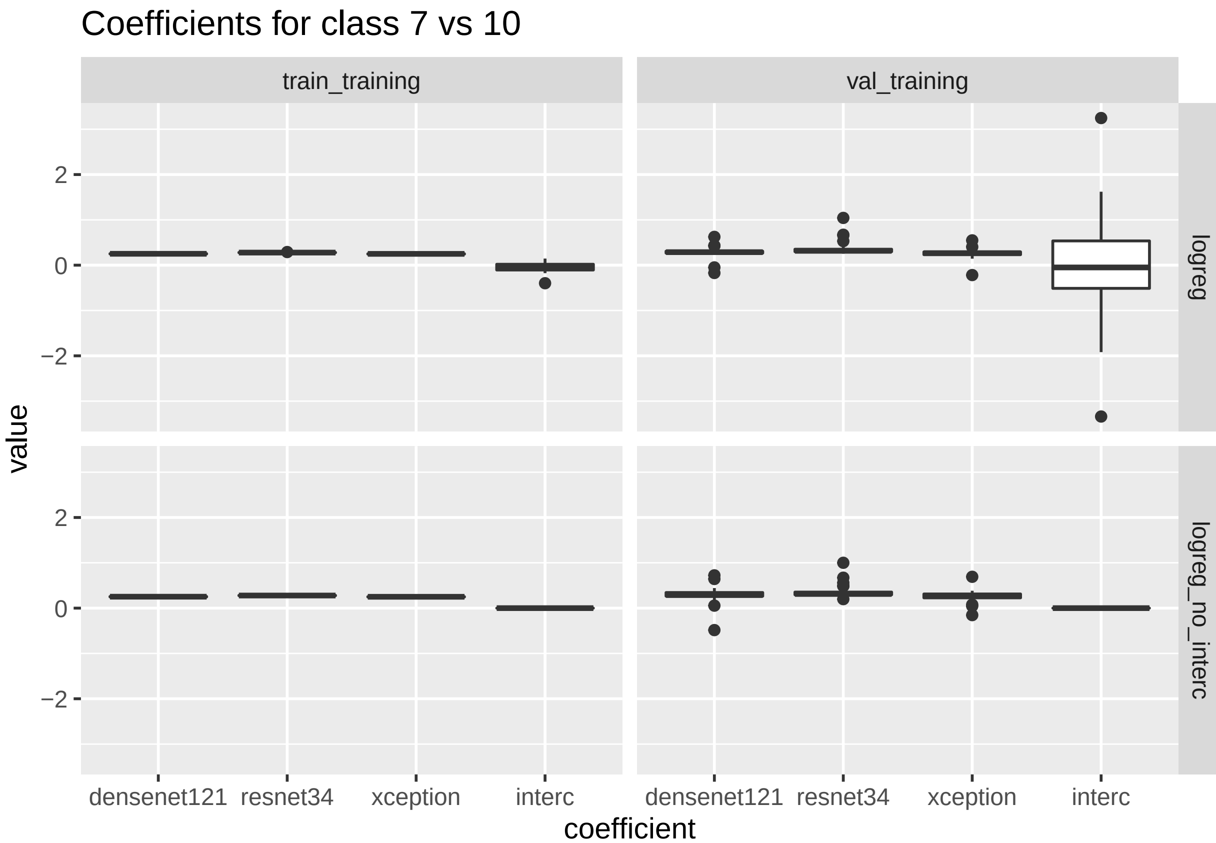


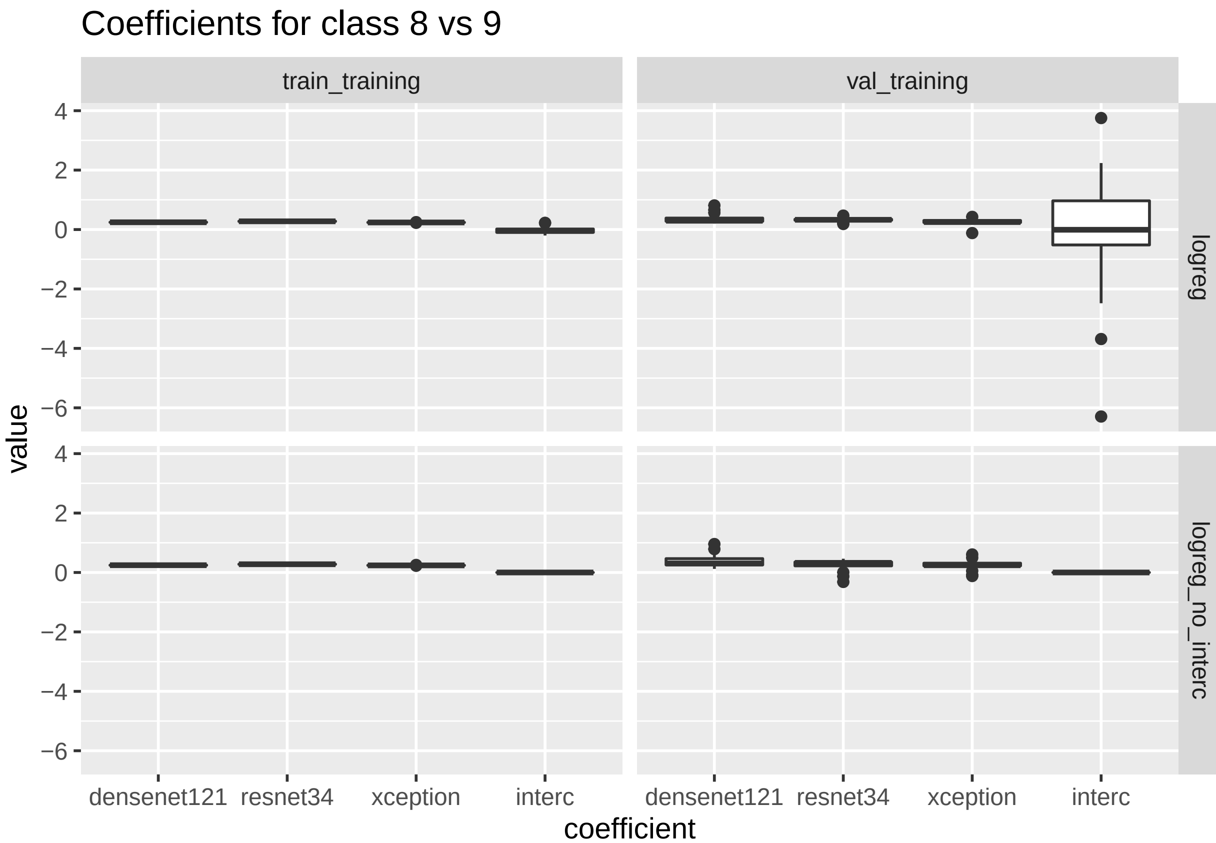
Coefficients for class 6 vs 9
train_training
val_training
logreg
logreg_no_interc
value
4
2
0
−2
densenet121
resnet34
xception
interc
coefficient

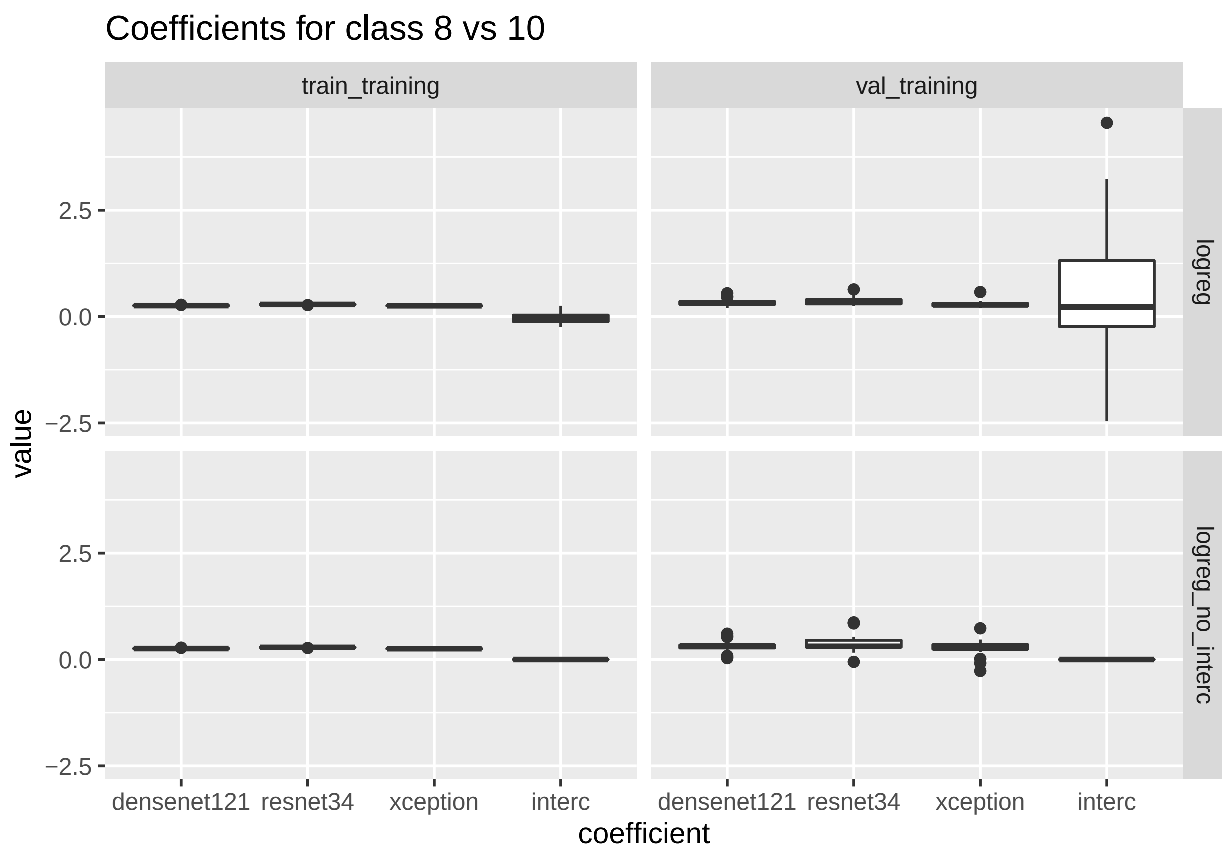

Coefficients for class 6 vs 10
train_training
val_training
logreg
logreg_no_interc
value
5.0
2.5
0.0
densenet121
resnet34
xception
interc
coefficient

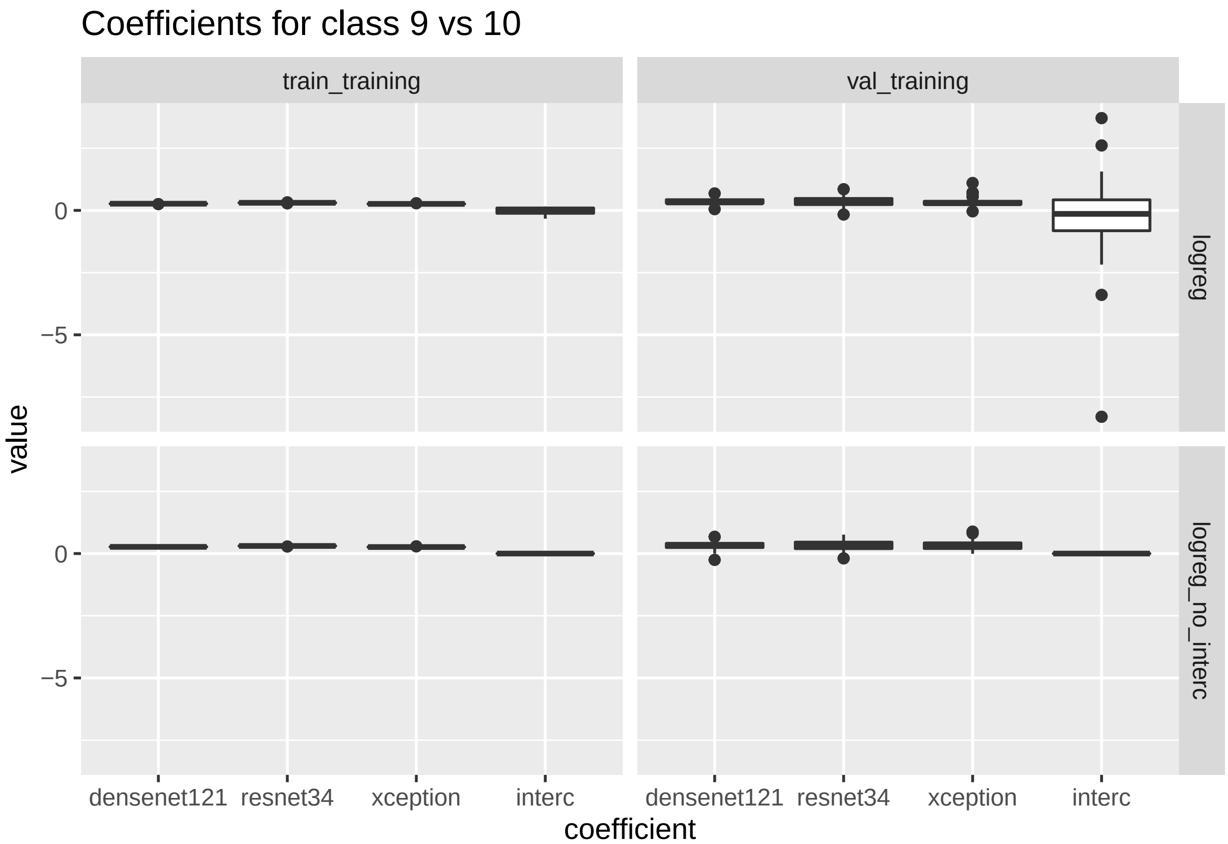

Coefficients for class 7 vs 8
train_training
val_training
logreg
logreg_no_interc
value
4
0
−4
−8
densenet121
resnet34
xception
interc
coefficient


Coefficients for class 7 vs 9
train_training
val_training
logreg
logreg_no_interc
value
2
0
−2
−4
densenet121
resnet34
xception
interc
coefficient


Coefficients for class 7 vs 10
train_training
val_training
logreg
logreg_no_interc
value
2
0
−2
2
0
−2
densenet121
resnet34
xception
interc
densenet121
resnet34
xception
interc
coefficient


Coefficients for class 8 vs 9
train_training
val_training
logreg
logreg_no_interc
value
4
2
0
−2
−4
−6
densenet121
resnet34
xception
interc
coefficient


Coefficients for class 8 vs 10
train_training
val_training
logreg
logreg_no_interc
value
2.5
0.0
−2.5
2.5
0.0
−2.5
densenet121
resnet34
xception
interc
densenet121
resnet34
xception
interc
coefficient



Validation training has higher variance in all coefficients. The difference is largest in intercept.

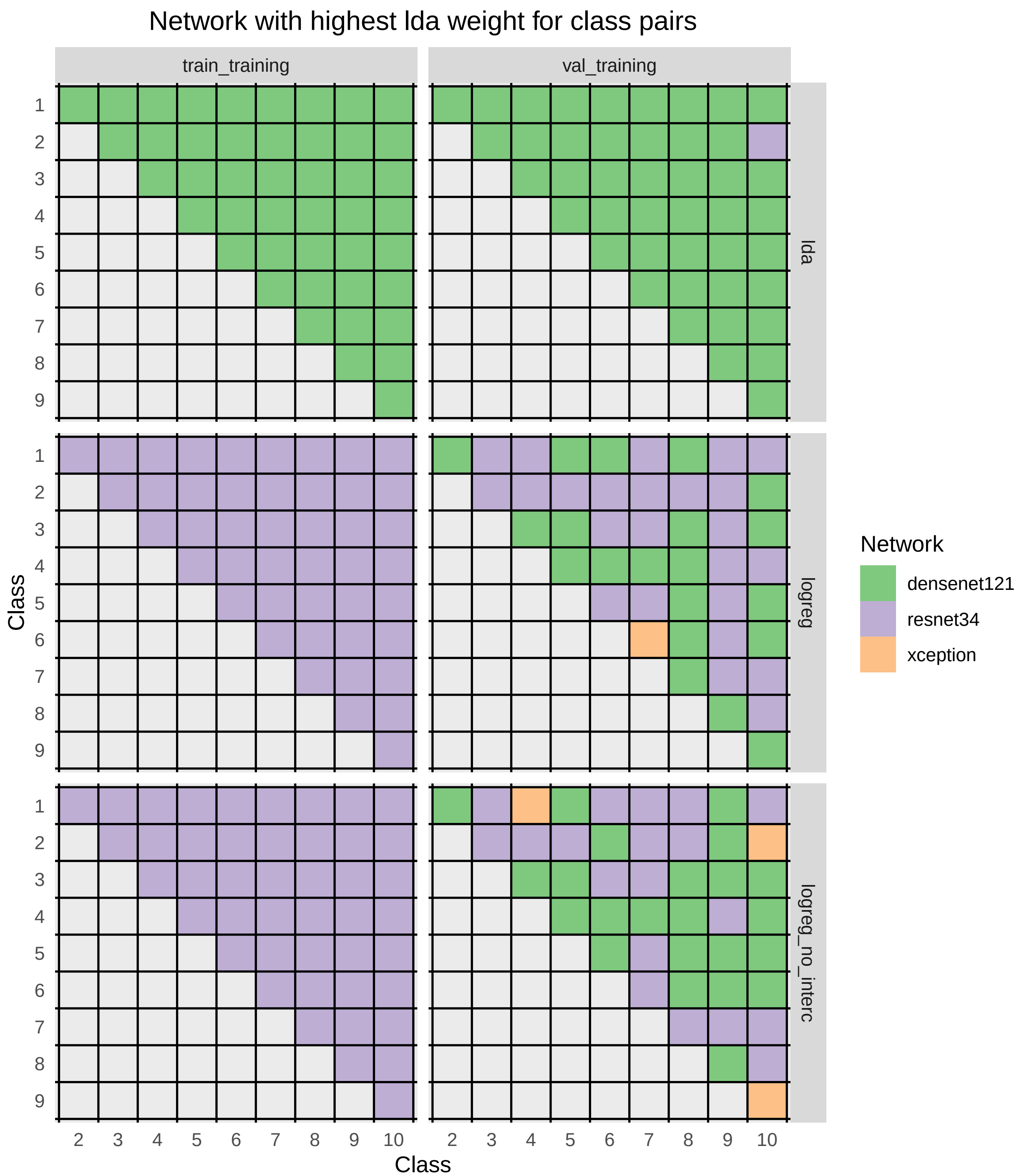
```
avg_combiner_coefs <- combiner_coefs %>% filter(coefficient != "interc") %>% group_by(class1, class2, pr
```

```
## `summarise()` has grouped output by 'class1', 'class2', 'precision', 'train_type', 'coefficient'. You
```

```
avg_combiner_c_w <- pivot_wider(avg_combiner_coefs, names_from = coefficient, values_from = value)
avg_combiner_c_w[, c("class1", "class2")] <- lapply(avg_combiner_c_w[, c("class1", "class2")], as.factor
avg_combiner_c_w$top_net <- factor(c("densenet121", "resnet34", "xception")[max.col(as.matrix(avg_combin
```

```
coefs_grid <- ggplot(avg_combiner_c_w, aes(x=class2, y=class1, fill=top_net)) +
  geom_raster() +
  scale_fill_brewer(type="qual") +
  facet_grid(cols=vars(train_type), rows=vars(combining_method)) +
  scale_y_discrete(limits=rev) +
  geom_vline(xintercept=seq(-0.5, 9.5, 1.0)) +
  geom_hline(yintercept=seq(-0.5, 9.5, 1.0)) +
  guides(fill=guide_legend(title="Network")) +
  xlab("Class") +
  ylab("Class") +
  ggtitle("Network with highest lda weight for class pairs") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.ticks = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())
```

```
coefs_grid
```



Densenet is clearly dominant for both LDA training methodologies. Surprisingly for logreg is dominant resnet in the tt case and mix of densenet and resnet in vt case. For all combining methods, training methodology tt provided better accuracy.