Multinomial Model

Daniel Polites

Setup

```
# Loads the MNIST dataset, saves as an .RData file if not in WD
if (!(file.exists("mnist_data.RData"))) {

  # ## installs older python version
  # reticulate::install_python("3.10:latest")
  # keras::install_keras(python_version = "3.10")
  # ## re-loads keras
  # library(keras)

  ## get MNIST data
  mnist <- dataset_mnist()
  ## save to WD as .RData
  save(mnist, file = "mnist_data.RData")

} else {
  ## read-in MNIST data
  load(file = "mnist_data.RData")
}

# Access the training and testing sets
x_train <- mnist$train$x
y_train <- mnist$train$y
x_test <- mnist$test$x
y_test <- mnist$test$y

rm(mnist)
```

```
## plot function
plot_mnist_array <- function(plt, main_label = NA, color = FALSE, dim_n = 28) {

  if (color == TRUE) {
      colfunc <- colorRampPalette(c("red", "white", "blue"))

      min_abs <- -max(abs(range(plt)))
      max_abs <- max(abs(range(plt)))

      col <- colfunc(256)
  } else {
    col <- gray((255:0)/255)
    min_abs <- 0
```

```
    max_abs <- 255
    }



  ## create image
  image(x = 1:dim_n,
        y = 1:dim_n,
        ## image is oriented incorrectly, this fixes it
        z = t(apply(plt, 2, rev)),
        col = col,
        zlim = c(min_abs, max_abs),
        axes = FALSE,
        xlab = NA,
        ylab = NA,
        main = ifelse(is.na(main_label),
                      "",
                      main_label))

  ## create plot border
  rect(xleft = 0.5,
       ybottom = 0.5,
       xright = 28 + 0.5,
       ytop = 28 + 0.5,
       border = "black",
       lwd = 1)

}
```

```
## train data

# initialize matrix
x_train_2 <- matrix(nrow = nrow(x_train),
                    ncol = 28*28)

## likely a faster way to do this in the future
for (i in 1:nrow(x_train)) {
  ## get each layer's matrix image, stretch to 28^2 x 1
  x_train_2[i, ] <- matrix(x_train[i, , ], 1, 28*28)
}

x_train_2 <- x_train_2 %>%
  as.data.frame()

## test data
x_test_2 <- matrix(nrow = nrow(x_test),
                   ncol = 28*28)

for (i in 1:nrow(x_test)) {
  x_test_2[i, ] <- matrix(x_test[i, , ], 1, 28*28)
}

x_test_2 <- x_test_2 %>%
```

```
  as.data.frame()


## re-scale data
x_train_2 <- x_train_2 / 256
x_test_2 <- x_test_2 / 256

## response
# x_test_2$y <- y_test
# x_train_2$y <- y_train
```

Model

train

```
## set training data size
# n <- nrow(x_train_2)
n <- 100

indices <- sample(x = 1:nrow(x_train_2),
                  size = n)

## init data
x_multi <- x_train_2[indices, ]
y_multi <- y_train[indices]

## drop cols with all 0s
#x_multi <- x_multi[, (colSums(x_multi) > 0)]
```

```
## for the sake of the coefficients viz, setting alpha = 0
init_model <- cv.glmnet(x = x_multi %>% as.matrix,
                        y = y_multi %>% factor,
                        family = "multinomial",
                        alpha = 0)
```

```
## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

```
## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

```
multi_model <- predict(init_model,
                       x_multi %>% as.matrix,
                       s = init_model$lambda.min,
                       type = "response")
```

```
## format results
preds_init <- multi_model[, , 1] %>%
  as.data.frame()

preds <- apply(X = preds_init,
               MARGIN = 1,
               FUN = function(x) names(which.max(x)) %>% as.numeric)

## TRAIN confusion matrix
table(y_multi, preds)
```

```
##         preds
## y_multi  0  1  2  3  4  5  6  7  8  9
##        0 12  0  0  0  0  0  0  0  1  0
##        1  0  9  0  0  0  0  0  0  0  0
##        2  0  0  7  1  0  0  0  0  0  0
##        3  0  0  0 15  0  0  0  0  0  0
##        4  0  0  0  0 11  0  0  0  0  0
##        5  0  0  0  1  0  9  0  0  0  0
##        6  0  0  0  0  0  0  5  0  0  0
##        7  0  1  0  0  0  0  0  9  0  0
##        8  0  0  0  1  0  0  0  0 13  0
##        9  0  0  0  0  0  0  0  1  0  4
```

```
## TRAIN misclassification rate
mean(!(y_multi == preds))
```

```
## [1] 0.06
```

test

```
## pre-process data
x_multi_test <- x_test_2 %>%
  select(all_of(names(x_multi)))

## get preds
multi_model_test <- predict(init_model,
                            x_multi_test %>% as.matrix,
                            s = init_model$lambda.min,
                            type = "response")
```

```
## format results
preds_init_test <- multi_model_test[, , 1] %>%
  as.data.frame()

preds_test <- apply(X = preds_init_test,
                    MARGIN = 1,
                    FUN = function(x) names(which.max(x)) %>% as.numeric)

## TEST confusion matrix
table(y_test, preds_test)
```

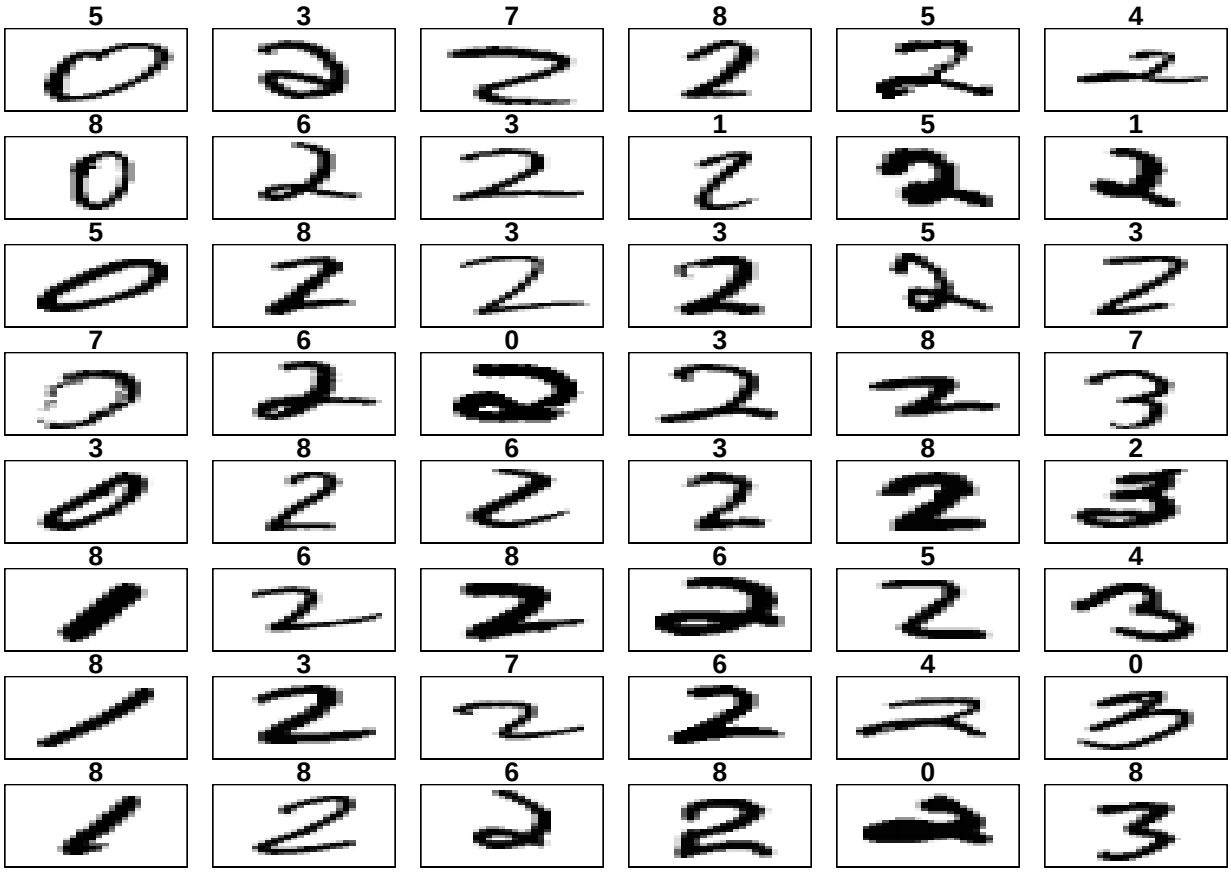
```
##       preds_test
## y_test    0    1    2    3    4    5    6    7    8    9
##      0  903    0    4    9    1   25   16    1   21    0
##      1    0 1006    1    9    0   16    3    0  100    0
##      2   35   43  371  172   24   34  164   18  171    0
##      3   10   20   18  821   10   21   11   18   80    1
##      4   10   38    2    7  797   17   11   12   74   14
##      5   32    8   16  168   28  432    6    8  186    8
##      6   84   39   95    3   30   31  595    0   81    0
##      7   13   40    0   15   45   36    5  757   39   78
##      8   15   18    5   53   22   37    3    8  809    4
##      9   23   31    4   19  288   14    0  187   79  364
```

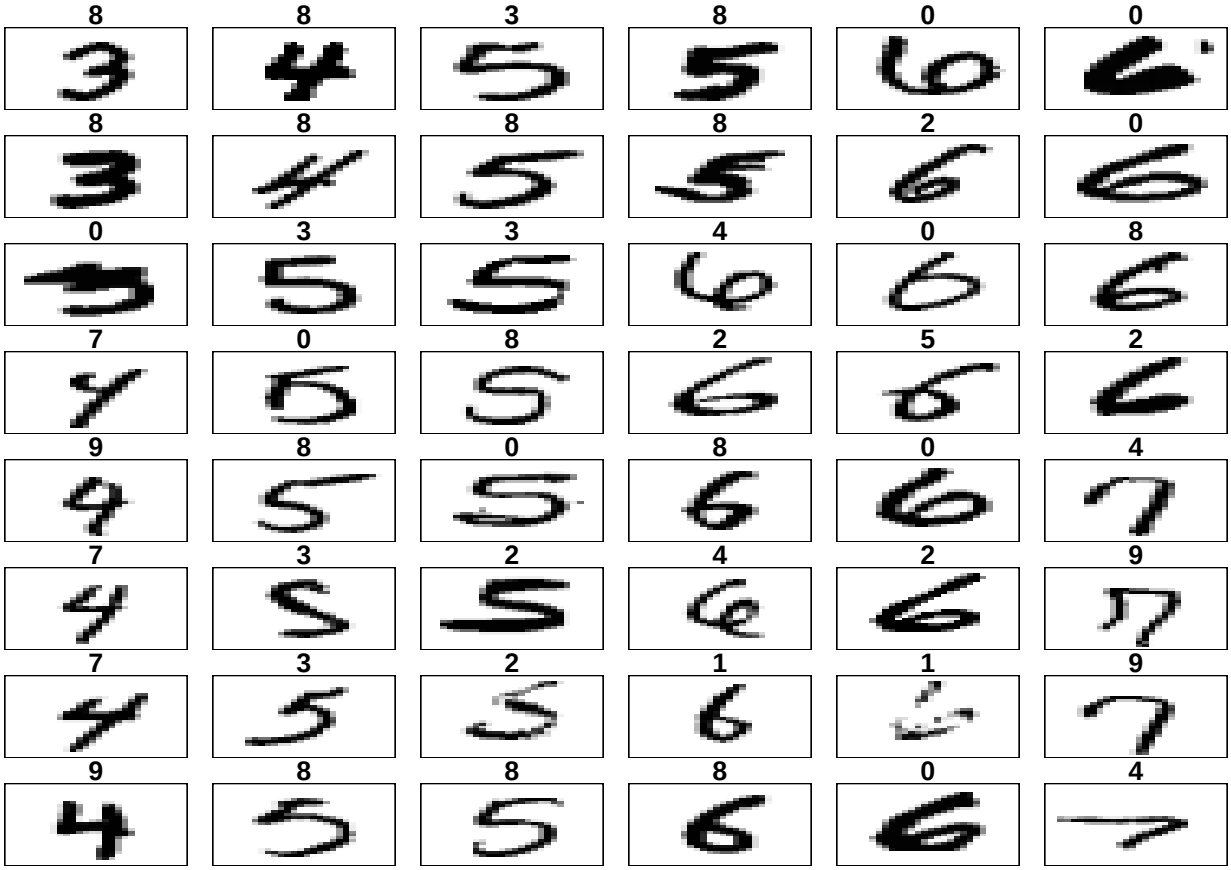
```
## TEST misclassification rate
mean(!(y_test == preds_test))
```

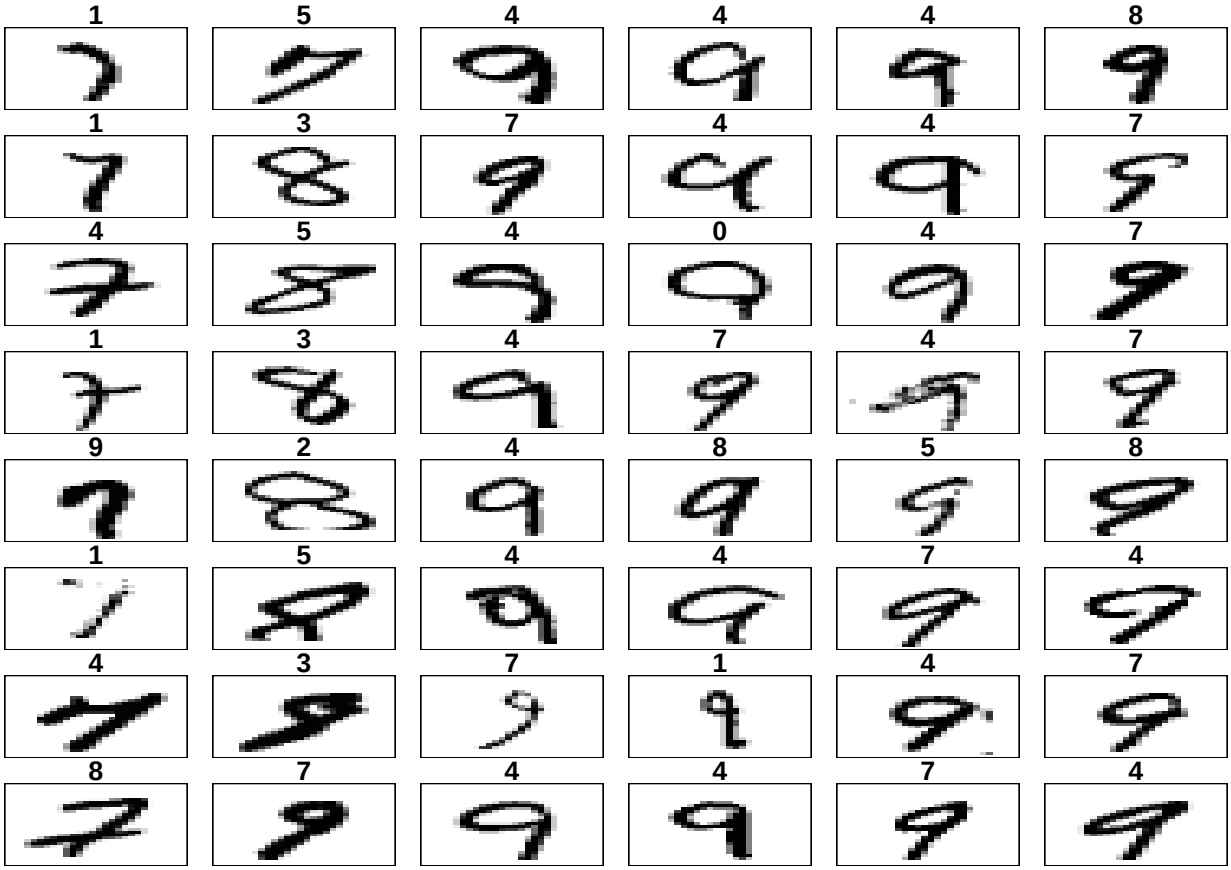
```
## [1] 0.3145
```

```
## sort vectors so outputs are grouped
x_test_sort <- x_test[order(y_test), , ]
y_test_sort <- y_test[order(y_test)]
preds_test_sort <- preds_test[order(y_test)]

## get misclassified obs
wrong <- which(!(y_test_sort == preds_test_sort))

## plot a sample of misclassified obs
plot_wrong <- wrong[sample(x = 1:length(wrong), size = 3*8*6)] %>%
```

```
  sort()

## plot params
par(mfcol = c(8, 6))
par(mar = c(0, 0.5, 1, 0.5))

for (i in plot_wrong) {
  plot_mnist_array(plt = x_test_sort[i, , ],
                   main_label = preds_test_sort[i])
}
```



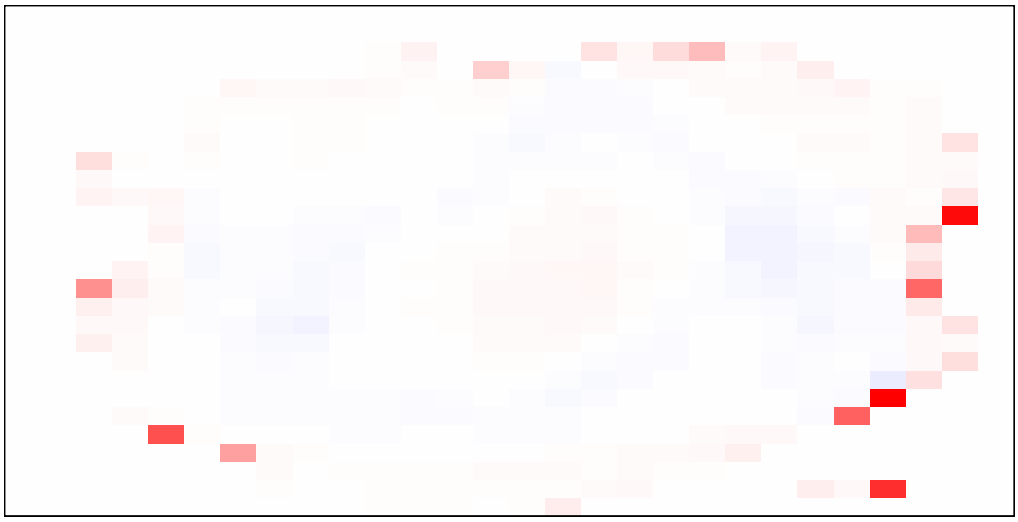

8 8 3 8 0 0
8 8 8 8 2 0
0 3 3 4 0 8
7 0 8 2 5 2
9 8 0 8 0 4
7 3 2 4 2 9
7 3 2 1 1 9
9 8 8 8 0 4



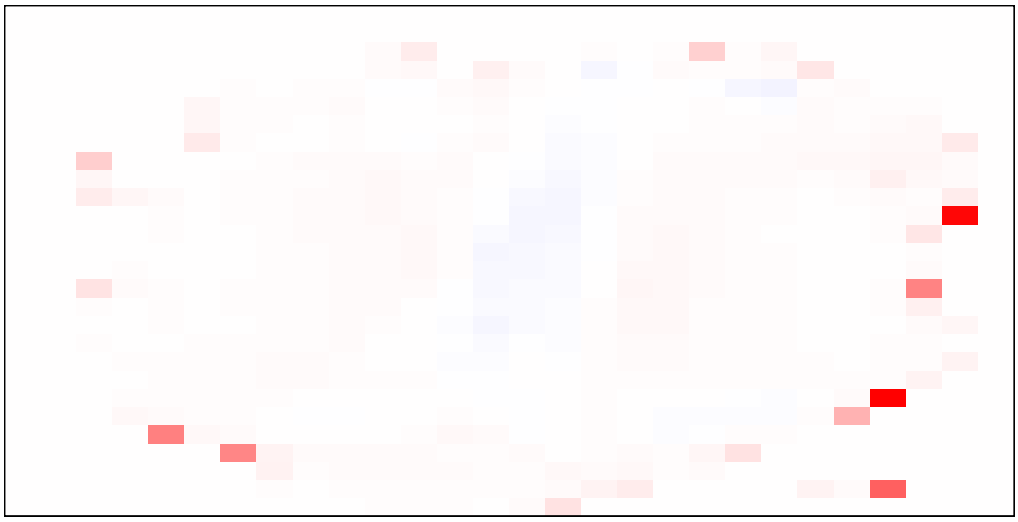

```
par(mfcol = c(1, 1))
```

model heatmaps

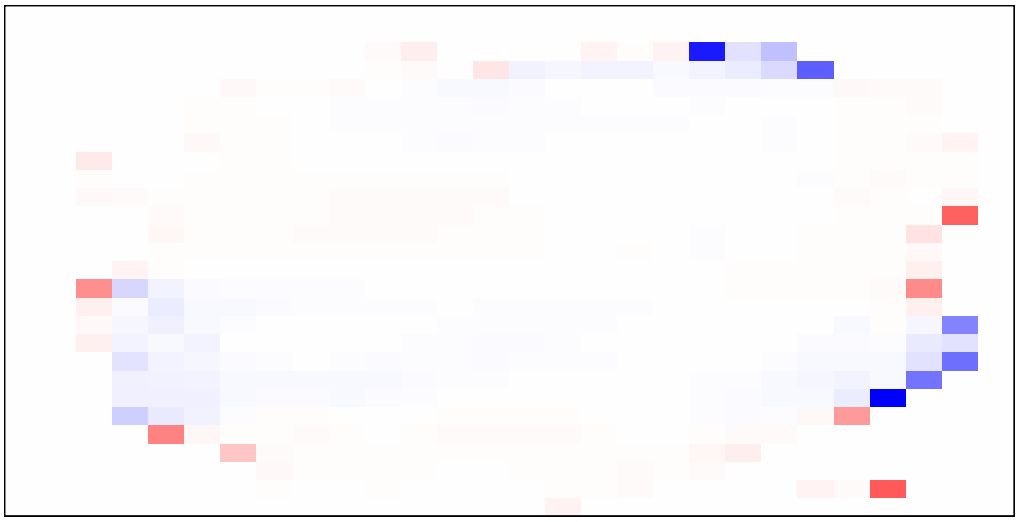
```
## get coefficients into matrices
model_coef <- coef(init_model, s = init_model$lambda.min) %>%
  lapply(as.matrix) %>%
  lapply(function(x) matrix(x[-1, ], nrow = 28, ncol = 28)) %>%
  ## take sigmoid activation just to help viz
  lapply(function(x) 1 / (1 + exp(-x)) - 0.5)

## plot
# par(mfcol = c(4, 3))
# par(mar = c(0, 0.5, 1, 0.5))

mapply(FUN = plot_mnist_array,
       plt = model_coef,
       main_label = names(model_coef),
       color = TRUE)
```


0

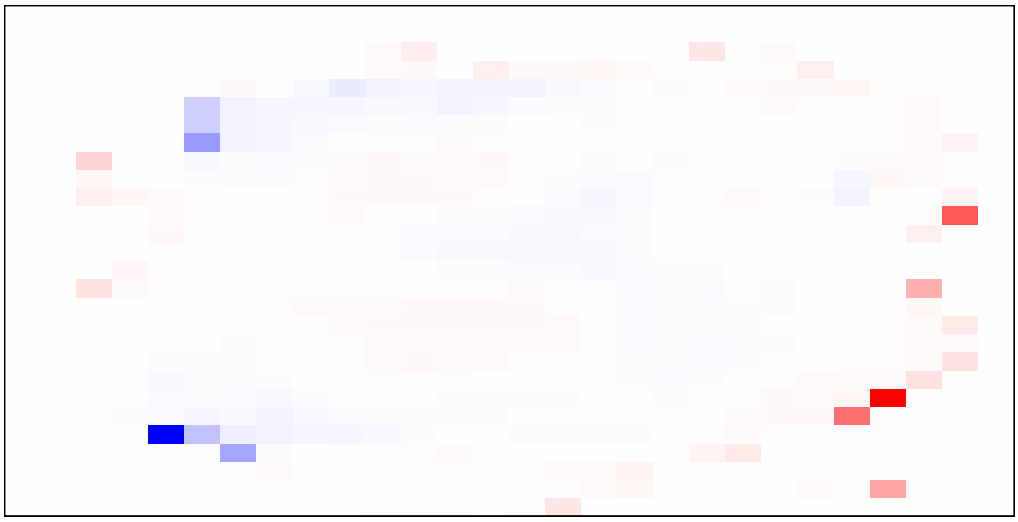
1



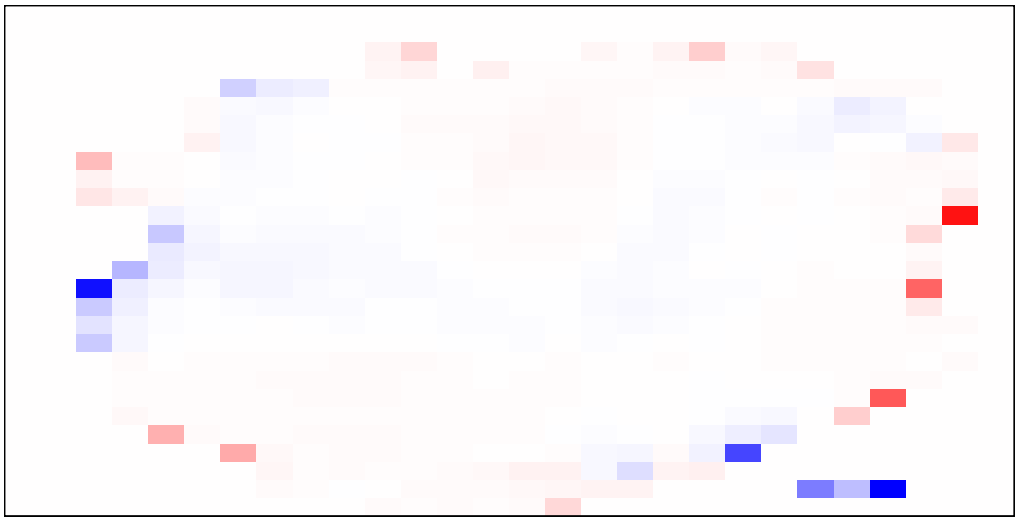
2



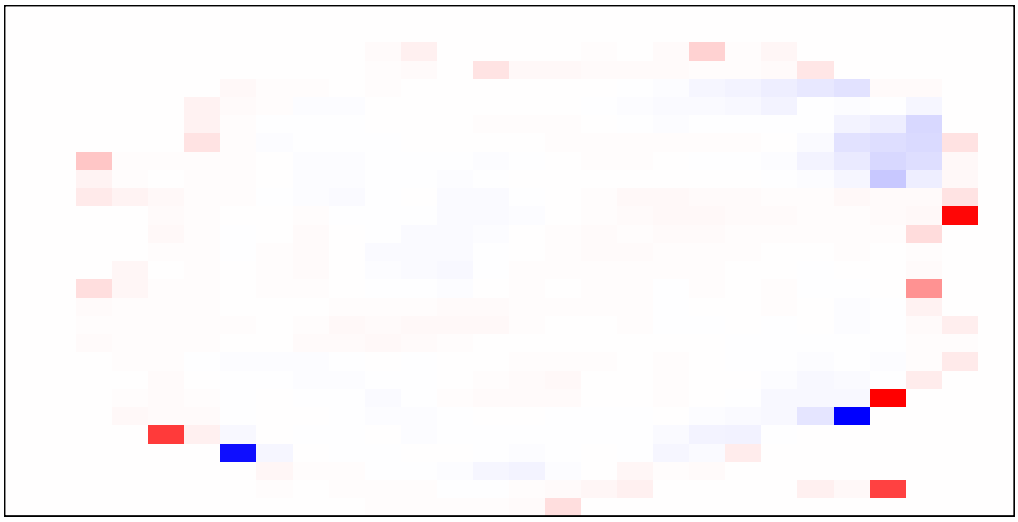
3



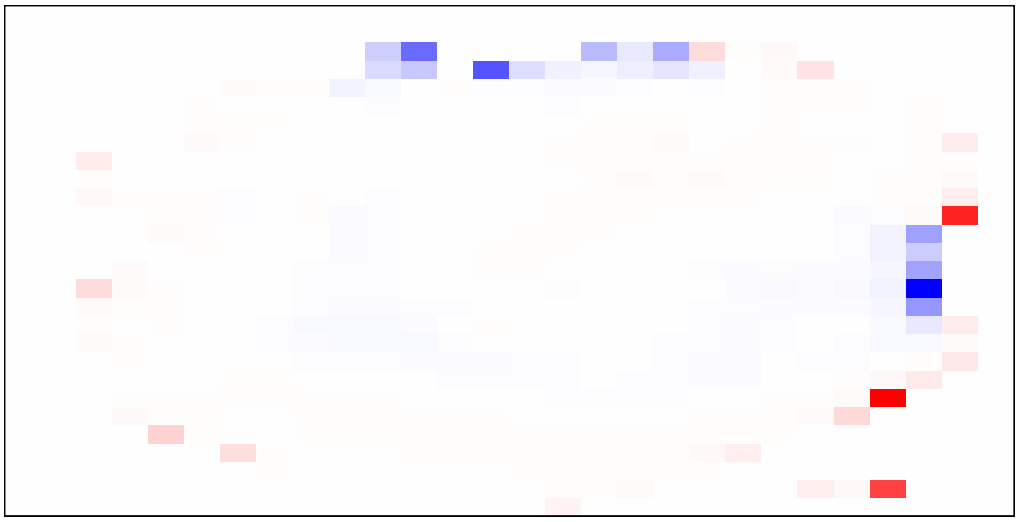
4



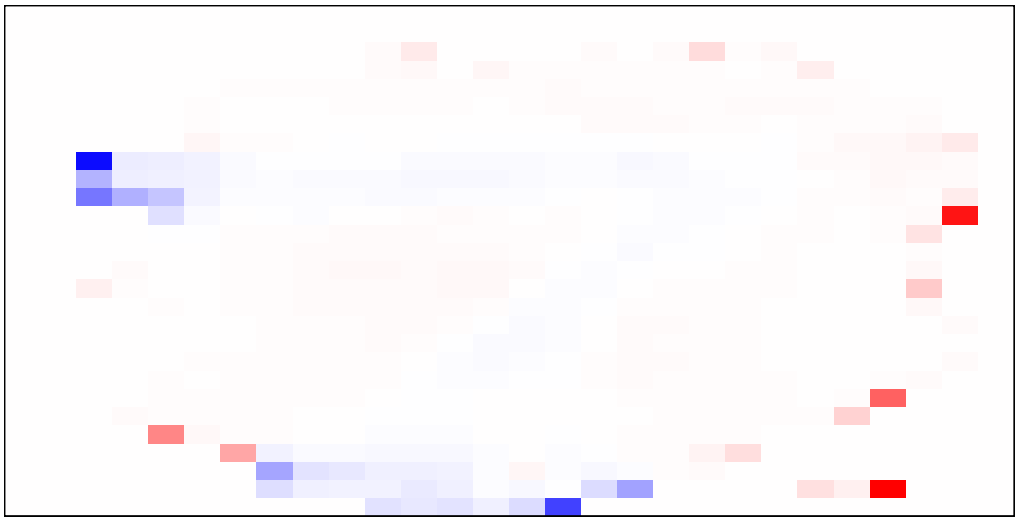
5



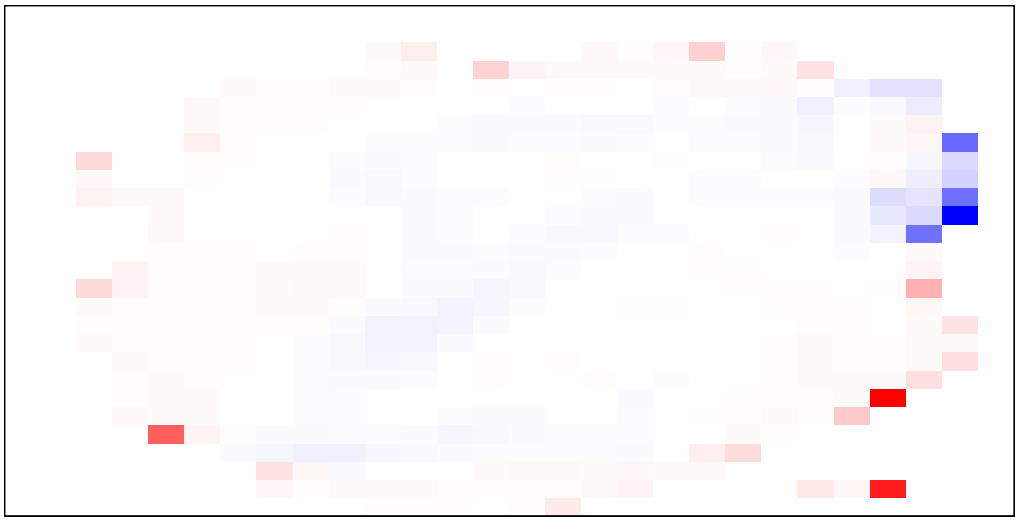
6

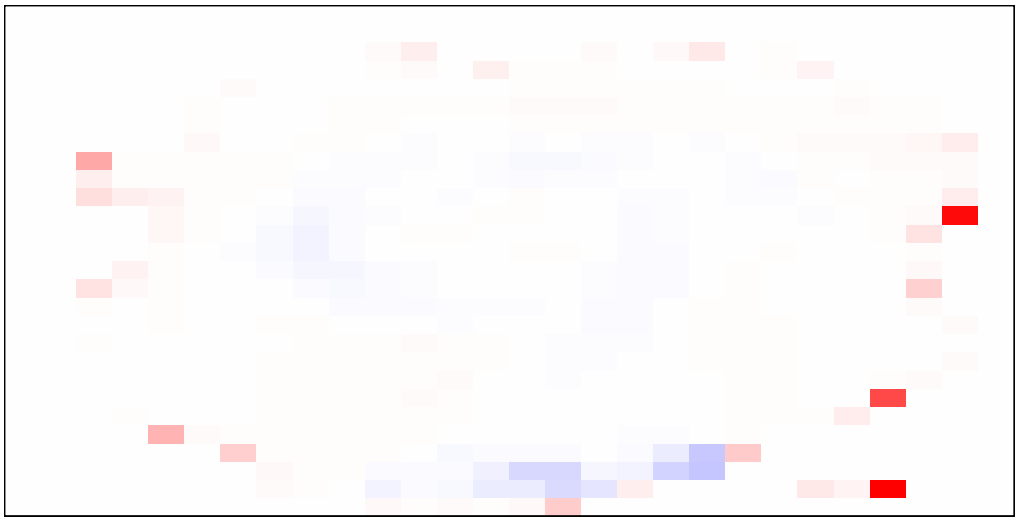


7



8






```
## $`0`
## NULL
##
## $`1`
## NULL
##
## $`2`
## NULL
##
## $`3`
## NULL
##
## $`4`
## NULL
##
## $`5`
## NULL
##
## $`6`
## NULL
##
## $`7`
## NULL
##
## $`8`
## NULL
```

```
##
## $`9`
## NULL
```

no outside cells model

earlier runs of the above sections revealed that for a regularization method that does not perform variable selection, odd importance is given to outermost cell for prediction. Thus, those will be removed:

```r
## set training data size
# n <- nrow(x_train_2)
n <- 100

indices <- sample(x = 1:nrow(x_train_2),
                  size = n)

## init data
x_multi <- x_train_2[indices, ]
y_multi <- y_train[indices]

## drop outer cells
x_multi <- x_multi[, rep(seq(146, 622, 28), each = 18) + rep(0:17, times = 18)]
```

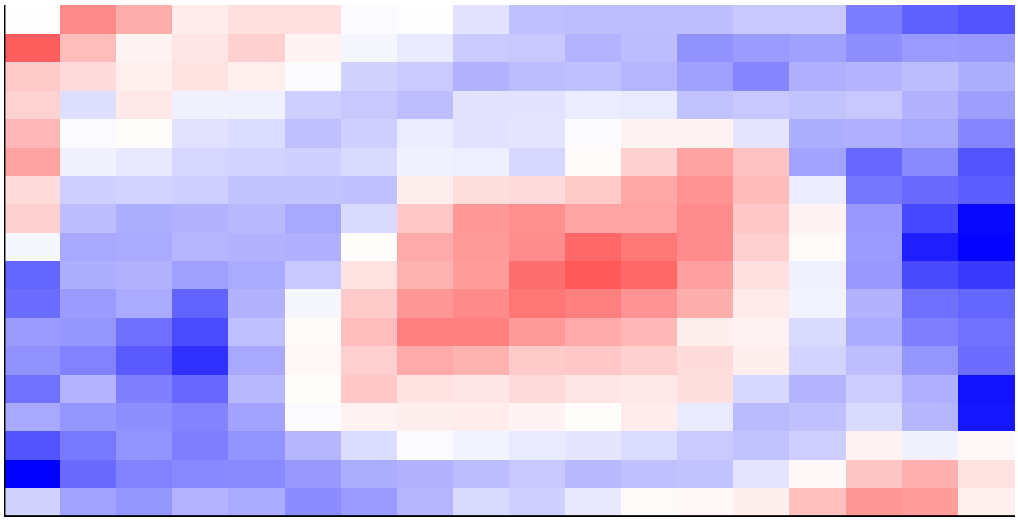
```r
## for the sake of the coefficients viz, setting alpha = 0
init_model <- cv.glmnet(x = x_multi %>% as.matrix,
                        y = y_multi %>% factor,
                        family = "multinomial",
                        alpha = 0)
```

```
## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

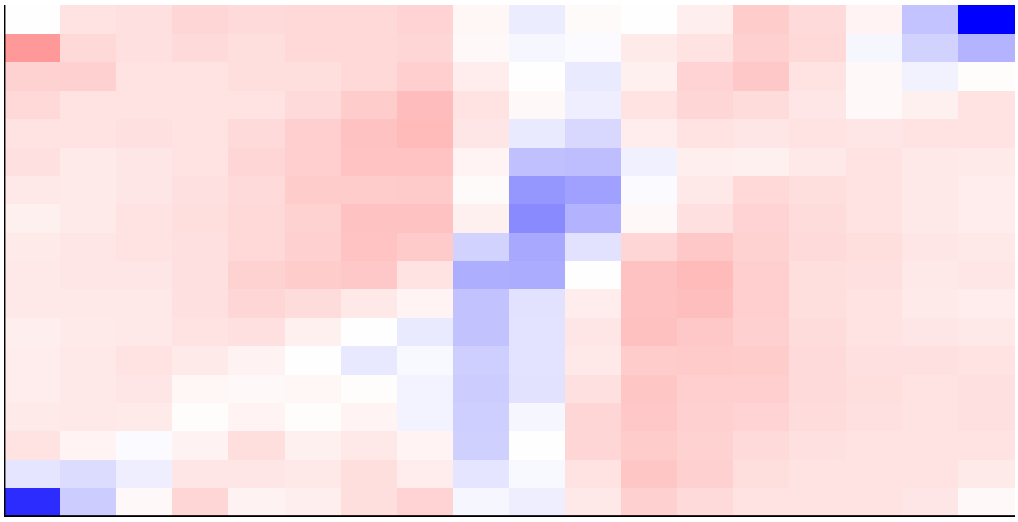
```
## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

```
multi_model <- predict(init_model,
                       x_multi %>% as.matrix,
                       s = init_model$lambda.min,
                       type = "response")
```

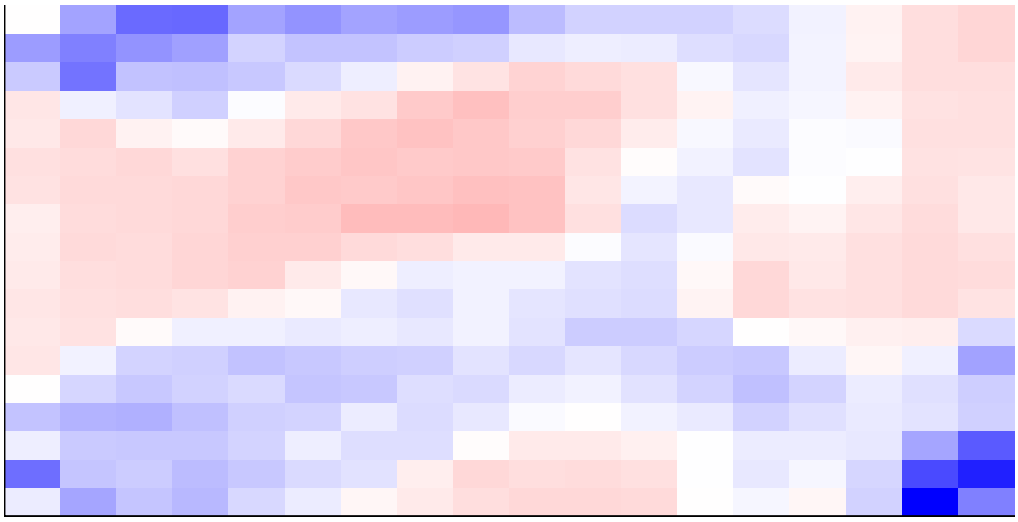
```
## get coefficients into matrices
model_coef <- coef(init_model, s = init_model$lambda.min) %>%
  lapply(as.matrix) %>%
  lapply(function(x) matrix(x[-1, ], nrow = 18, ncol = 18)) %>%
  ## take sigmoid activation just to help viz
  lapply(function(x) 1 / (1 + exp(-x)) - 0.5)

mapply(FUN = plot_mnist_array,
       plt = model_coef,
       main_label = names(model_coef),
       color = TRUE,
       dim_n = 18)
```


0

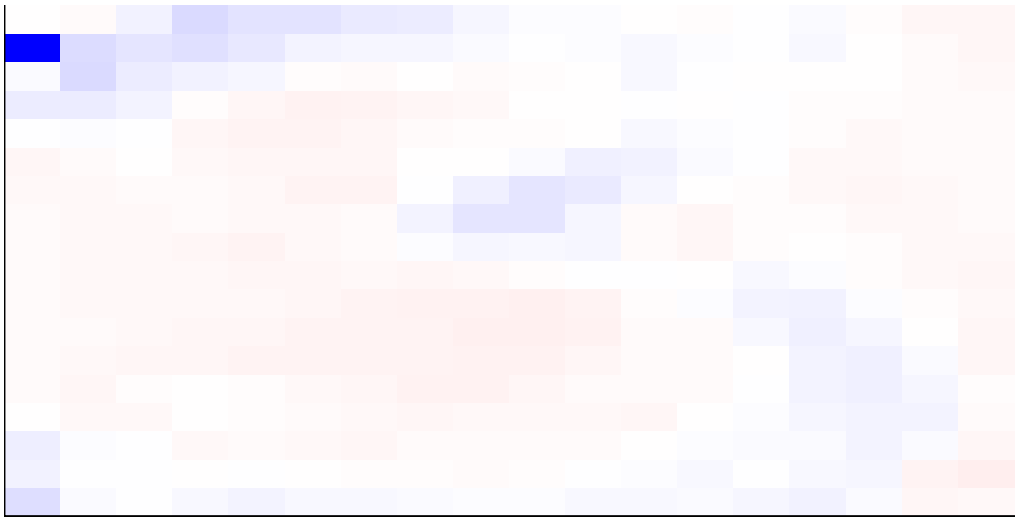
1



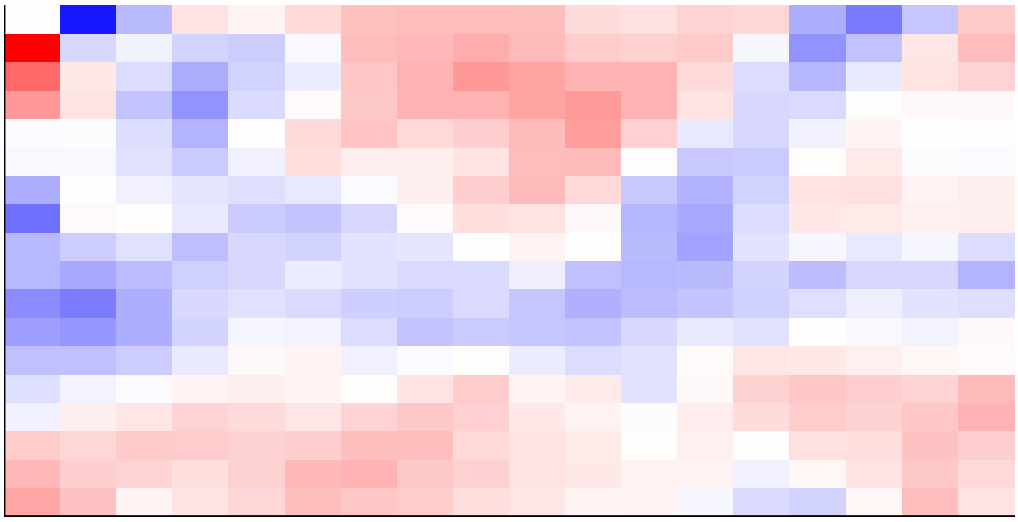
2



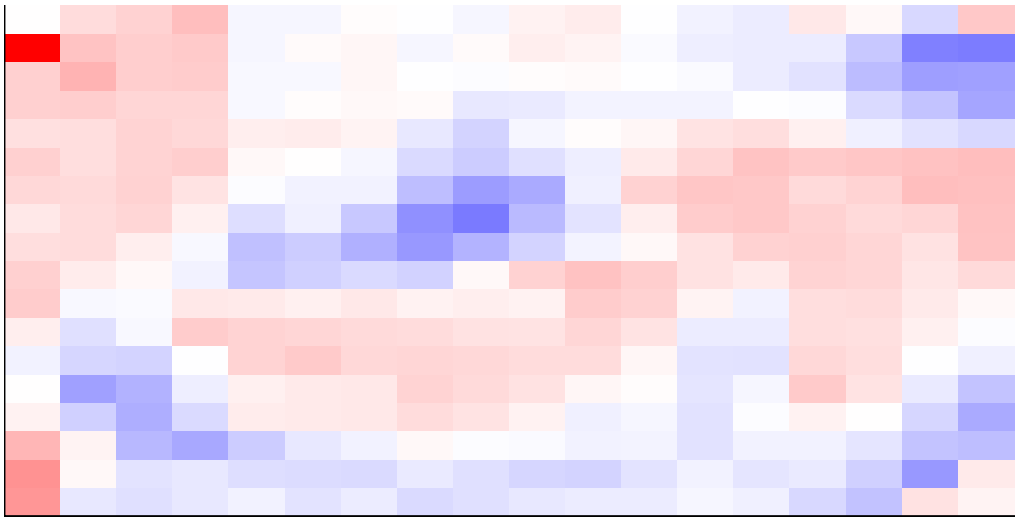
3



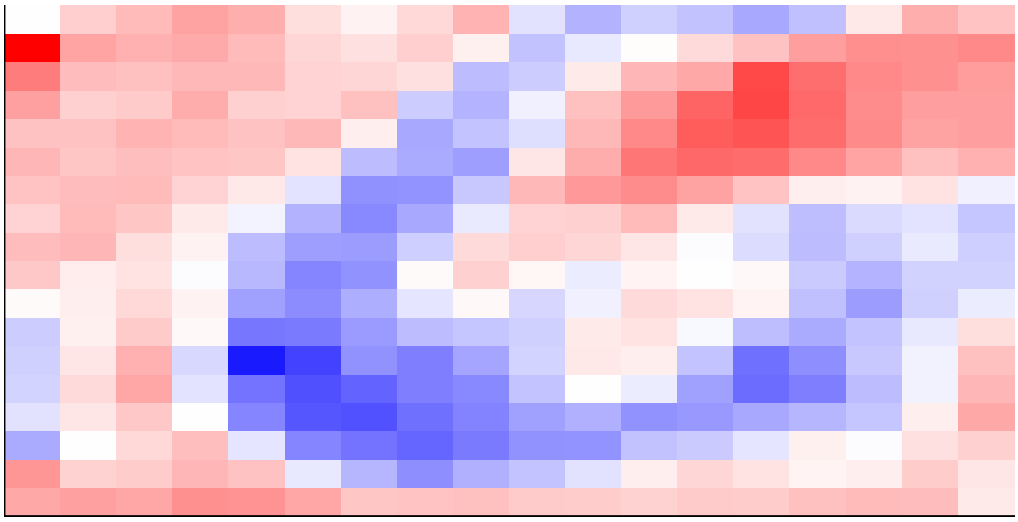
4



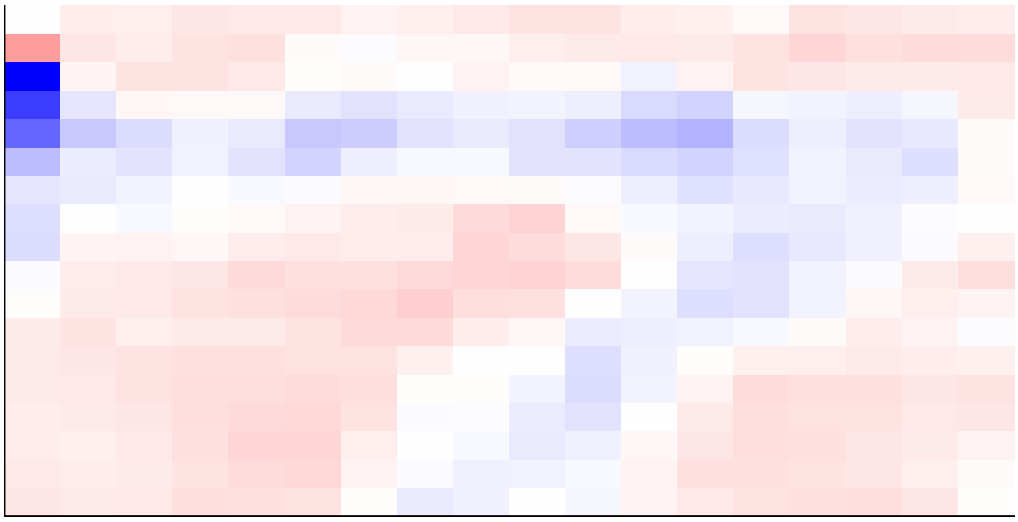
5



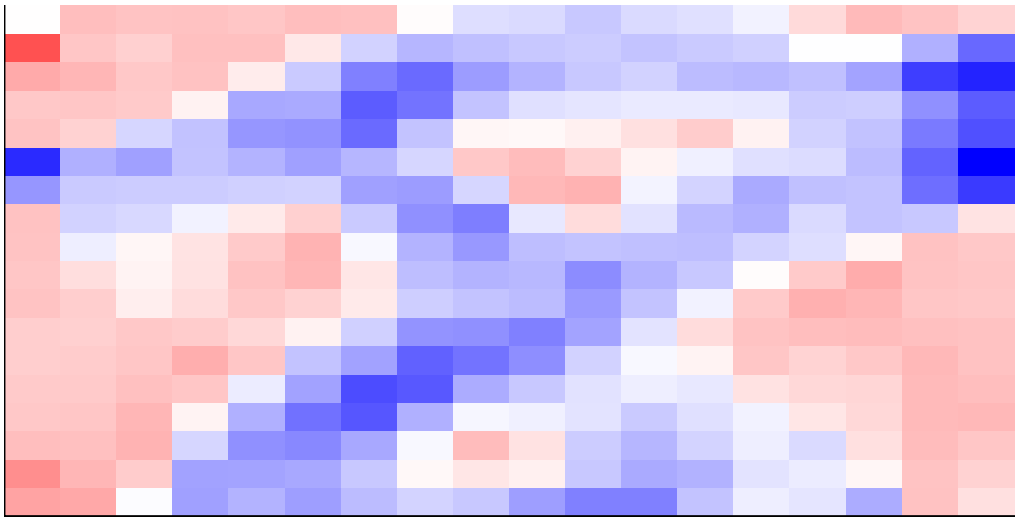
6

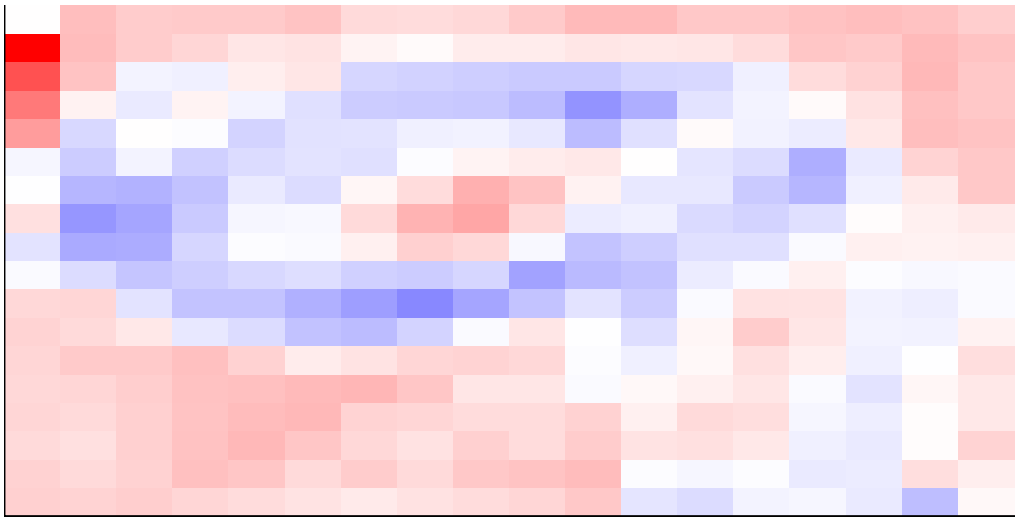


7



8






```
## $`0`
## NULL
##
## $`1`
## NULL
##
## $`2`
## NULL
##
## $`3`
## NULL
##
## $`4`
## NULL
##
## $`5`
## NULL
##
## $`6`
## NULL
##
## $`7`
## NULL
##
## $`8`
## NULL
```

```
##
## $`9`
## NULL
```