Images

Matthew Boyd

2/28/2022

```
library(tidyverse)
```

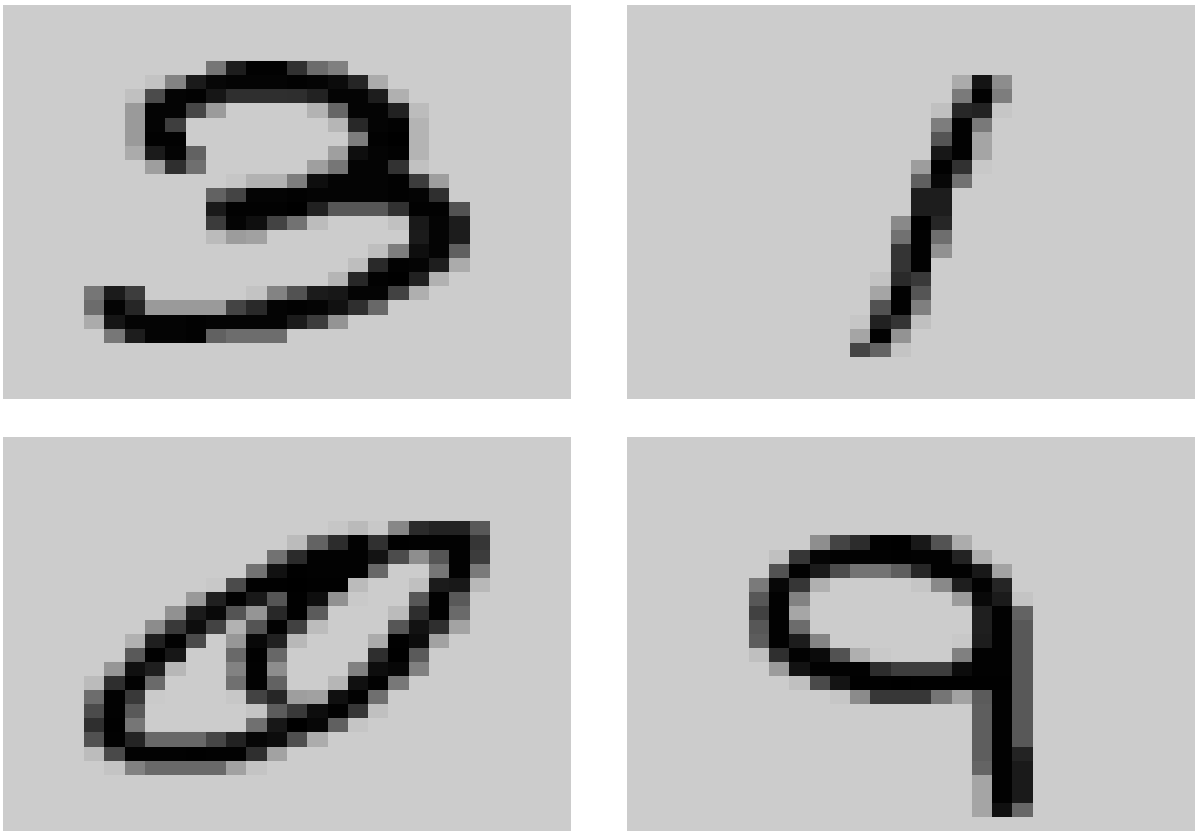
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(class)
library(caret)
```

```
## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(ggpubr)
load("~/DSCI 478/Kaggle Project/DSCI_478_Kaggle/knn_accuracy.Rdata")
df = read.csv("~/DSCI 478/Kaggle Project/digit-recognizer/train.csv")
test = read.csv("~/DSCI 478/Kaggle Project/digit-recognizer/test.csv")

# Visualize different numbers by row
image = function(row) {
  df %>%
    slice(row) %>%
    select(starts_with("pixel")) %>%
    pivot_longer(starts_with("pixel")) %>%
    mutate(x = (row_number() - 1) %% 28,
           y = -((row_number() - 1) %/% 28)) %>%
    ggplot(aes(x = x, y = y, fill = value)) +
```

```
    geom_raster() +
    theme_void() +
    scale_fill_gradient(high = "black",
                        low = "gray80") +
    guides(fill = "none")
}

one = image(15)

two = image(16)

three = image(150)

four = image(1600)

ggarrange(one, two, three, four, ncol = 2, nrow = 2)
```



```
knn_df %>%
  group_by(k) %>%
  summarize(accuracy = mean(accuracy)) %>%
  arrange(desc(accuracy)) %>%
  ggplot(aes(x = k, y = accuracy))+
  geom_line() +
  geom_point() +
  ylim(0.95, 0.97) +
```

```
labs(x = "K", y = "Cross Validation Accuracy",
     title = "KNN Cross Validation Accuracy") +
theme_minimal()
```