# Homework_1

Jeffrey LIANG

2/4/2021

```r
set.seed(123123)
sl_tr =
  read_csv(here::here("solubility_train.csv")) %>%
  janitor::clean_names()

sl_ts =
  read_csv(here::here("solubility_test.csv")) %>%
  janitor::clean_names()

x_tr = model.matrix(solubility ~ ., sl_tr)[, -1]

y_tr = sl_tr$solubility
```

## Q1

```r
sl_lm =
  train(solubility~.,
        data = sl_tr,
        method = "lm",
        trControl =
          trainControl(
            method = "cv",
            number = 10,
            repeats = 50
          ))

sl_lm
```

```
## Linear Regression
##
## 951 samples
## 228 predictors
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 856, 856, 856, 855, 856, 856, ...
## Resampling results:
##
```

```
##    RMSE    Rsquared  MAE
##    0.701   0.884     0.524
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
sl_lm_pre = predict.train(sl_lm,newdata = sl_ts)

RMSE(sl_lm_pre,sl_ts$solubility)
```

```
## [1] 0.746
```

# Q2