

# ACTL3142

## Week 7: Machine Learning Ideas

---

Tetian Madfouni

[t.madfouni@unsw.edu.au](mailto:t.madfouni@unsw.edu.au)

# Announcements

## Assignment Part 2

- Due 26<sup>th</sup> July
- Do read the task carefully, some key things change, especially for the predictive task.

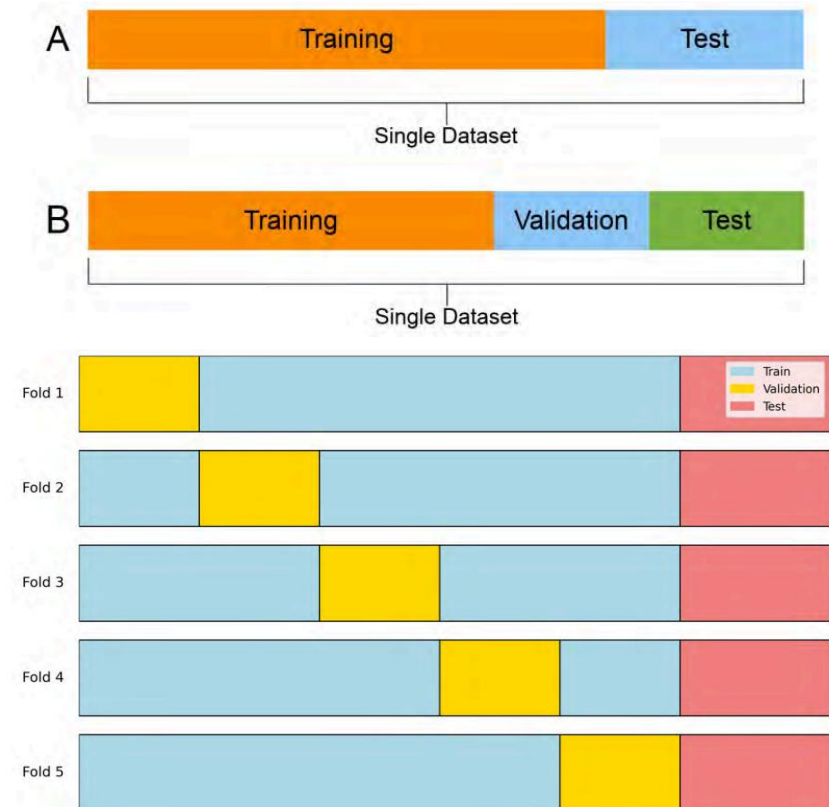
# Recap: Model Evaluation

What is the current way we've been splitting our data sets and why?

- Train/Test split – allows us to consider how our model will perform on unseen data and escape problems like overfitting. Can anyone think of any problems or shortcomings of this method? (Think of a model like KNN)

3 primary methods for overcoming those shortcomings were presented in the slides, what are they?

- Validation Set Approach – split the data set into Train/Validation/Test
- $k$ -fold Cross Validation – enhancement of above where we split Train into  $k$  Train/Validation groupings and average errors over the  $k$  folds
- Leave-One-Out Cross-Validation (LOOCV) – more specific version of  $k$ -fold CV



# Recap: Regularised Regressions

What are the two regularisation methods we learned this week?

- Ridge regression
- Lasso regression

What is the general goal of both? And how do they differ?

- To shrink coefficient values and simplify model without relying on computationally expensive methods like subset selection
- Ridge shrinks all values smaller, while Lasso tends to force the weakest predictors all the way to 0. This is changed by changing the form of the penalty, ridge uses the sum of squares of coefficients, while lasso uses their absolute values

