THE UNIVERSITY OF MELBOURNE

Workshop 10

COMP20008

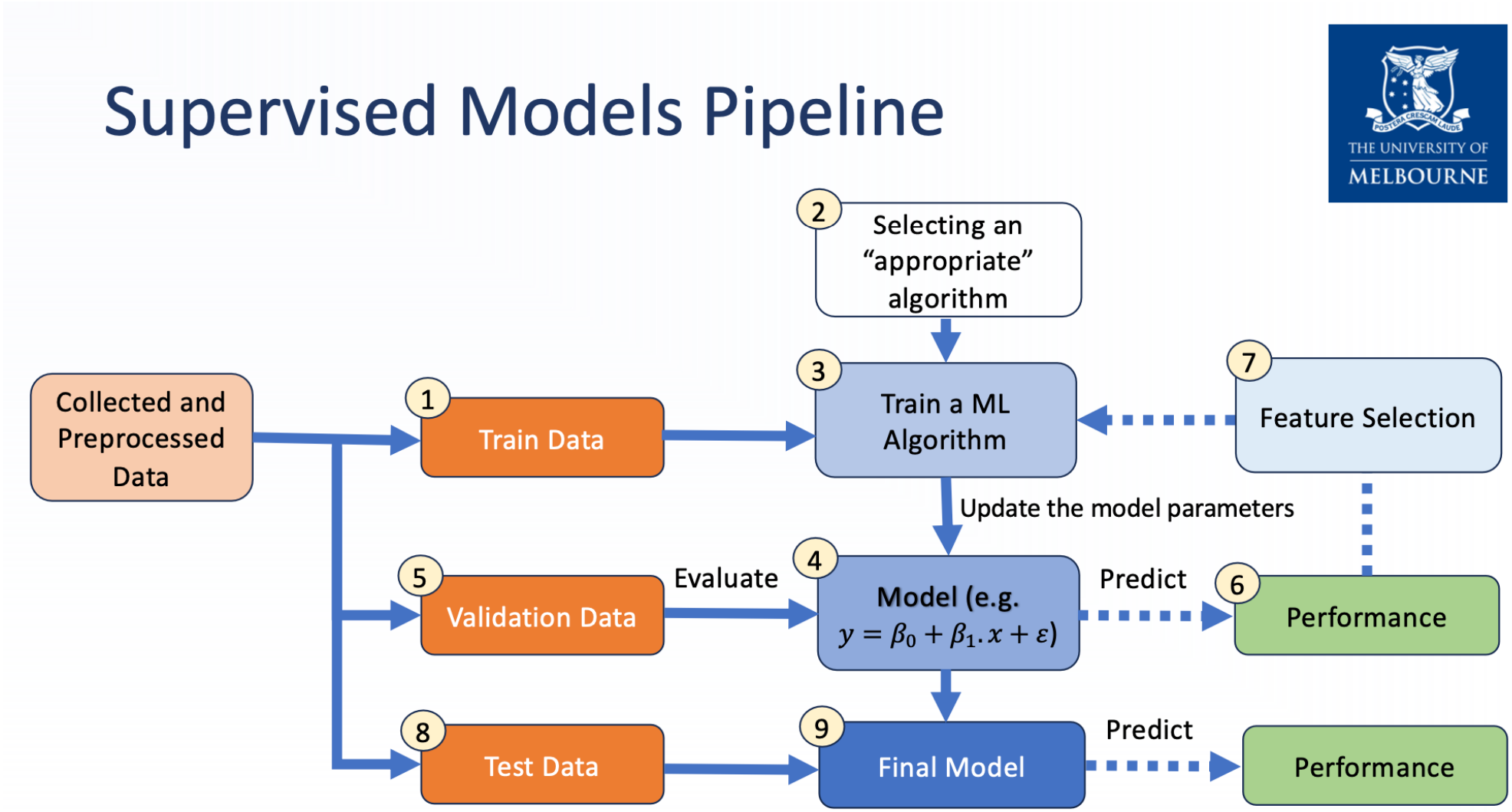
Elements of Data Processing

Zijie Xu

Agenda

- Supervised ML Evaluation
  - Why train/validation/test split
  - Holdout, Cross-validation & Stratification
  - Bootstrapping
- Performance metrics

Why train/validation/test split

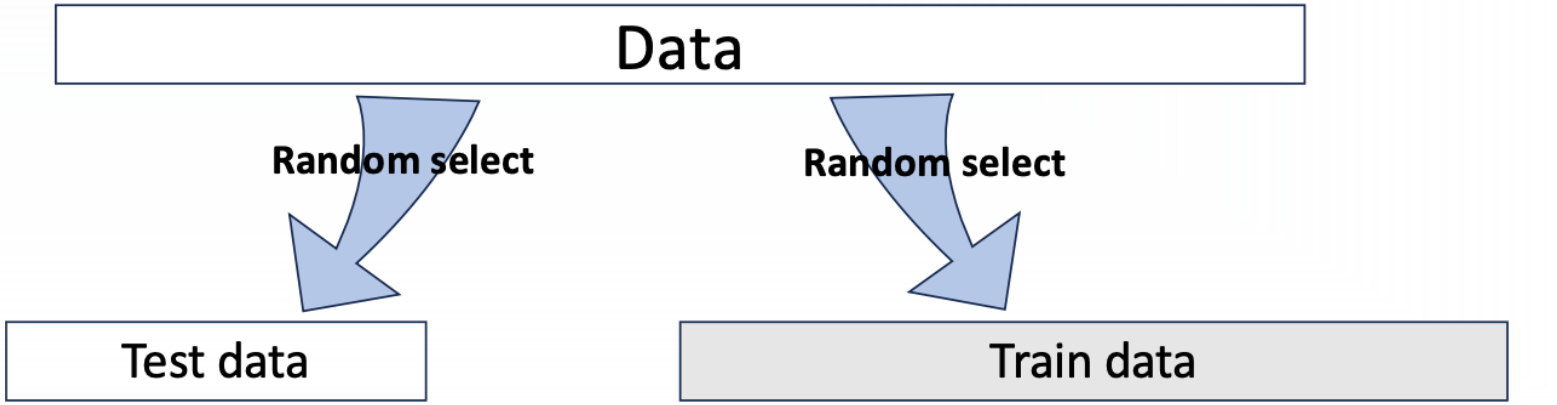


Evaluation methods

- We want to quantitatively determine how well our model performs on unseen data
- We split data into three parts to prevent data leakage
  - Model learns from training data
  - We tune hyperparameters on validation data
  - Fit tuned model to combined train & validation data, test its performance on testing data

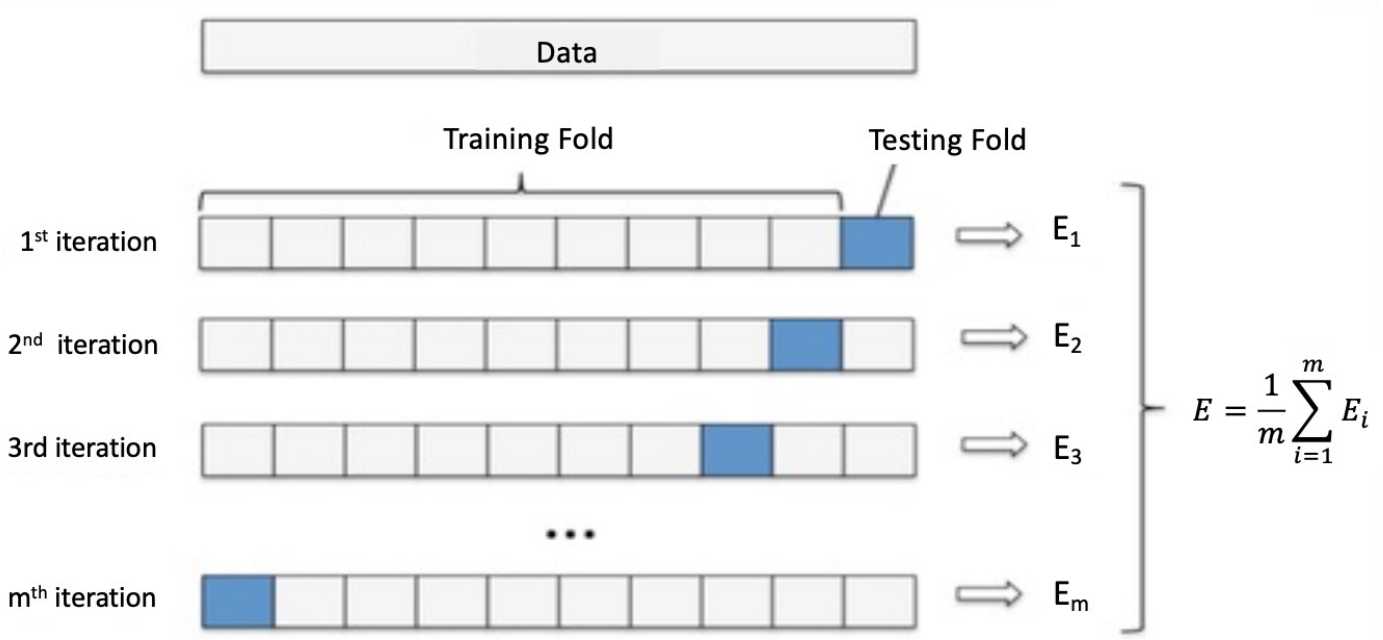
Evaluation methods

- Three methods
  - Holdout: Hold out a fixed proportion of data as testing set



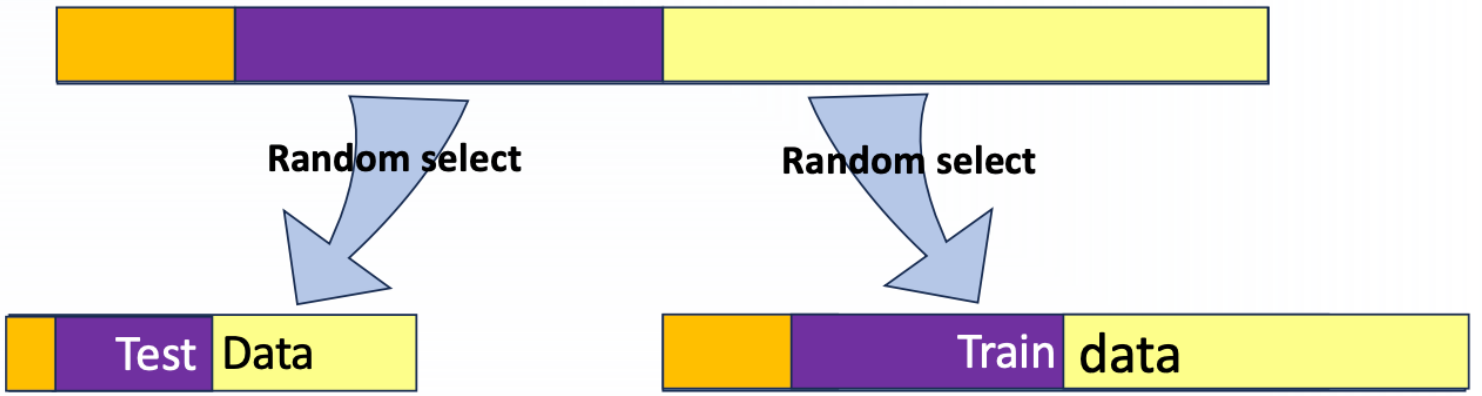
Evaluation methods

- Three methods
  - *m*-fold Cross-validation: Split the data into *m* folds, train on *m*-1, test on 1, repeat for all folds to and evaluate average performance



Evaluation methods

- Three methods
  - Stratification: preserves class distribution in training and testing set



Bootstrapping

- An additional method to evaluate model
  - Randomly draw sample from data set with replacement to obtained bootstrapped sample
  - Train model on bootstrapped sample, evaluate performance on corresponding test sample (data not seen in bootstrapped set)
  - Repeat many times to get average performance results

Performance metrics

- Some metrics for classification:
  - Accuracy $\frac{TP+TN}{P+N}$
  - Precision $\frac{TP}{TP+FP}$
  - Recall $\frac{TP}{TP+FN}$
  - F1-score $\frac{2*precision*recall}{precision+recall}$

Total = P+N

		Predicted ($\hat{Y}$)	
		$\hat{Y} = +1$	$\hat{Y} = -1$
Actual (Y)	Y = P	TP	FN
	Y = N	FP	TN

THE UNIVERSITY OF MELBOURNE

Thank you

More Resources: Canvas