

Evaluating performance of machine learning algorithms

APSC 8280: Machine learning applied to plant science

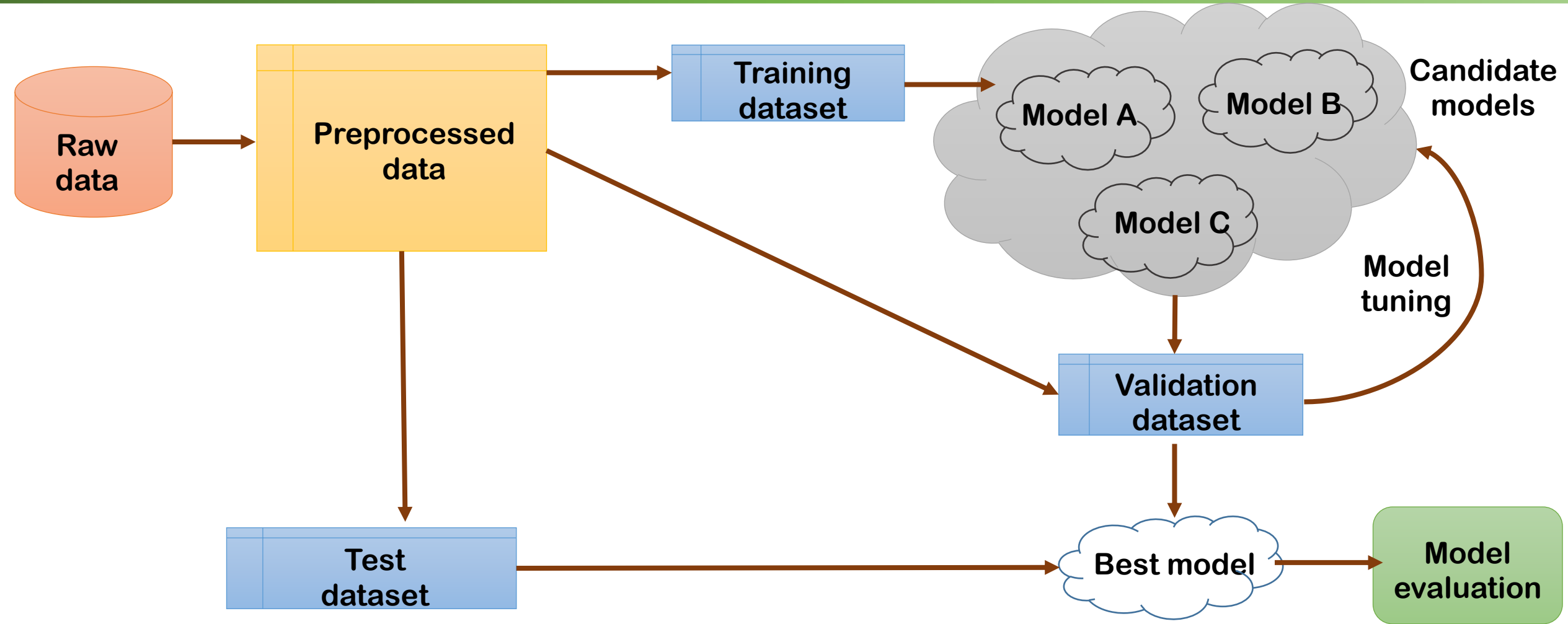
Outline

- **Meta performance metrics**
- **Estimating generalizability**
- **Regression**
- **Classification**

Meta performance metrics

- Time taken to train the model on the data
- Time taken for a fitted model to make predictions on new data instances
- Size of metadata that must be stored to use the model
- Suitability for online learning
- Interpretability

Estimating generalizability



Regression performance metrics

➤ Mean squared error (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

➤ Root mean squared error (RMSE)

$$RMSE = \sqrt{MSE}$$

➤ Mean absolute error

$$MAE = \frac{1}{n} \sum_{i=1}^n (|y_i - \hat{y}_i|)$$

➤ Root mean squared logarithmic error

$$RMSLE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(y_i + 1) - \log(\hat{y}_i + 1))^2}$$

➤ Coefficient of determination

$$R^2 = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}}$$

Classification performance metrics

- Accuracy
- True and false positive rate
- Sensitivity (true positive rate) and specificity
- False negative and false positive rate
- AUC
- Kappa statistic
- Cross-entropy (logloss)
- Gini index

Confusion matrix

		Predicted class	
		Spam	Not spam
Actual class	Spam	✓	✗
	Not spam	✗	✓

		Predicted class		
		Setosa	Virginica	Versicolor
Actual class	Setosa	✓	✗	✗
	Virginica	✗	✓	✗
	Versicolor	✗	✗	✓

We need most of our decisions to be correct

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Error rate} = 1 - \text{Accuracy}$$

We want users to see almost all of their legitimate email

$$\text{Specificity (true negative rate)} = \frac{TN}{TN + FP}$$

We want most of what we marked as spam to be actually spam

$$\text{Precision (positive predictive value)} = \frac{TP}{TP + FP}$$

We want to significantly reduce the amount of spam that users see

$$\text{Sensitivity (true positive rate, recall)} = \frac{TP}{TP + FN}$$

Balance between precision and recall

$$F\text{-measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{recall} + \text{precision}} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

		Predicted class	
		Spam	Not spam
Actual class	Spam	TP True positive	FN False negative
	Not spam	FP False positive	TN True negative

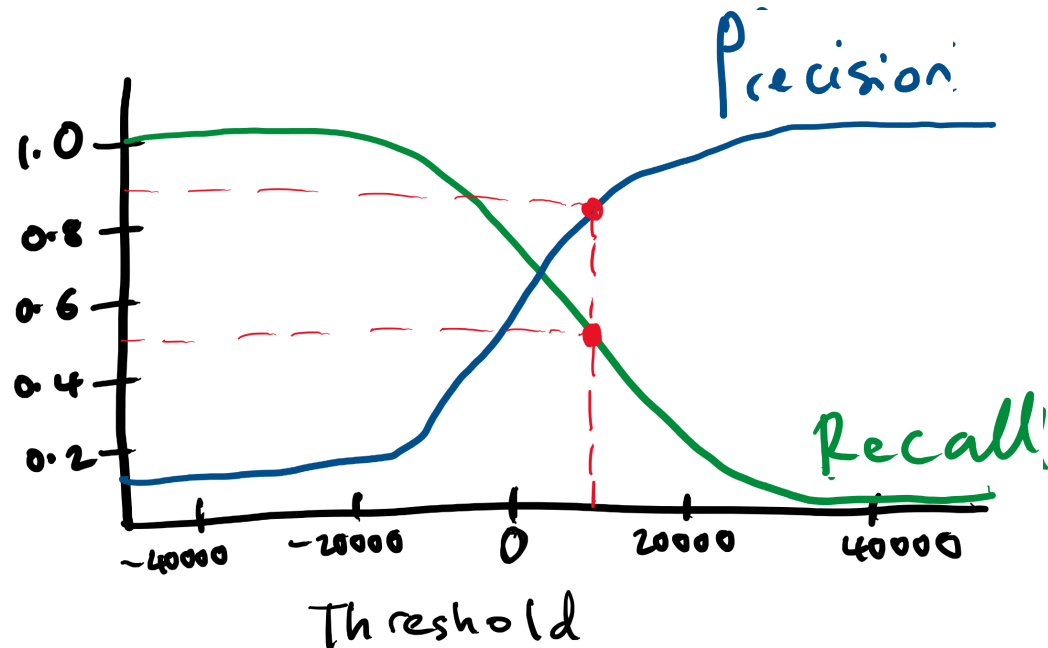
Achieving balanced performance

Class imbalance

Not spam	Spam
90%	10%

- Get more data
- Over sampling
- Under sampling
- Tree-based algorithms
- Use right metrics

Precision recall tradeoff



ROC curve



AUC

0.9 – 1.0 : Outstanding
0.8 – 0.9 : Excellent/Good
0.7 – 0.8 : Acceptable/Fair
0.6 – 0.7 : Poor
0.5 – 0.6 : No Discrimination