MACHINE LEARNING UNIVERSITY

ML for Leaders: Solving Business Problems

Prof. Thomas Gilbert
Amazon Scholar
June 2021

Course Schedule

Day One	Day Two	Day Three
ML LINGO, PROCESS, AND STAKEHOLDERS	ML BUSINESS GOAL AND PERFORMANCE METRICS	ML, CAUSALITY, AND EXPERIMENTS
Case #1: Churn	Case #2: Janus	Case #3: Catalove
Statistical learning and machine learning	Business and ML decision questions	Business heuristics versus science
Regressions and forecasting	Decision trees and random forests	Case #4: Emails and Sales
Process and stakeholders	Performance metrics	Causality and experiments

MACHINE LEARNING UNIVERSITY

Module 2: ML Business Goal and Performance Metrics

Case #2: New Customer Troubles?

Module 2: ML Business Goal and Performance Metrics

Case #2: Bona Fide and Janus

- Please turn your cameras on and participate
  - Raise your hand and/or unmute and/or write in chat
  - I will cold call!

- What is the business problem in early 2017?

- What is Jasmine’s business problem?

- What is a hypothesis behind Jasmine’s business problem? What data supports it? Who can help with historical data analysis?

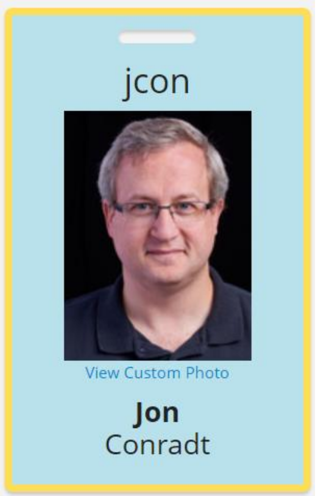
- Why does it matter if an account is fake or real if only real accounts will continue to make purchases?

- What is the ML version of the business problem that the Janus model is addressing?

- Why is the Bona Fide model not enough, i.e., why do we need the Janus model?

- How accurate are the Janus model's target labels, i.e., do we know for sure whether an account is held by a normal person or a suspected bad actor?

- Why does the random forest model use 128 decision trees and why does each tree have a depth of 8?



- Which parameters of the Janus model need to be set in agreement between science and business? Why?

- Can you interpret the significance of all 500 features used? Does high feature significance imply a positive relationship with the target variable? Why or why not?

Feature Name (column)	**Sig. Score**
NEW_CUSTOMER	0.29722496
NEW_CUSTOMER_MOBILE	0.12239406
PT_AUTHEN	0.04677547
PRE_HIT_COUNT	0.04133801
POST_PT_CHECKOUT	0.03202277
PT_CHECKOUT	0.02888195
POST_HIT_COUNT	0.02543502
PRE_PT_DETAIL	0.01949924
NTA_HOUR	0.01735604
HIT_COUNT	0.017224
PT_CHECKOUTADDRESSAW	0.01249372
PT_CHECKOUTPAYMENTAW	0.01213777
PRE_PT_AUTHEN	0.01177205
PRE_PT_CHECKOUTPREFETCH	0.01143146
PT_CHECKOUTSHIPOPTIONAW	0.01133428
POST_PT_GATEWAYMSHOP	0.01073951
DAY_1	0.0102
PT_MARKETPLACEREDIRECTAJAX	0.00968874
ACTIVITY_LENGTH	0.00874284
DAY_NEG1	0.00866397
POST_PT_DETAIL	0.00847364
POST_PT_CHECKOUTPAYMENTAW	0.00846187
POST_PT_CHECKOUTPREFETCH	0.00797943
PT_CHECKOUTPAYMENT	0.007549
POST_PT_DETAILWEBVIEW	0.00747085

- Do we care more about the Janus model's false positives or false negatives? Why? Should Amazon aggressively shut down suspected bad actors? Why or why not?

- Why do we now report the model's predicted percentage of new customers that are suspected actors in weekly business reviews?

	Supected Bad Actors (SBA) Predictions* - See footnote (i)	Weekly Metrics (MM)					SBA (% of total NTAs by Week)				
33											
34	Domestic B2C Shoppers	0.12	0.11	0.09	0.06	0.08	20.4%	21.1%	19.7%	12.3%	16.1%
35	+ Prime Domestic B2C	0.03	0.03	0.02	0.01	0.02	15.3%	16.6%	15.0%	9.1%	12.7%
36	+ Non-Prime Domestic B2C	0.08	0.08	0.07	0.04	0.06	23.5%	23.4%	22.1%	13.8%	17.7%

- Like all models, the Janus model will have to be reviewed. Changes to the model could lead to significant changes in the results presented in weekly business reviews. How should you approach these changes to avoid unnecessary deep dives that can waste downstream teams’ time and resources?

Case #2: Bona Fide and Janus

- Is there a problem with our NTA flow?
  - Do we need to redesign the website?
  - Are we running out of new customers?

“You don’t learn a lot about ducks by studying the decoys”

Business and ML Decision Questions

Module 2: ML Business Goal and Performance Metrics

Three Main Phases for the PM's ML Work

- Hypothesis development
  - » What is the business problem? What is the current and wanted customer behavior? What is our goal? What is the metric of success? Map entire business process like an engineer?
- Data analysis/analytics without ML
  - » What does the historical data tell us? What experiments can we run to test our hypotheses? Do we need to revise/narrow our objective?
- ML solution
  - » Simple model: Can we gain intuition and trust quickly?
  - » Full model: What are we predicting? How good are we at it?

From Business to ML Problem

Define your use case objective (business problem):

These are specific, measurable milestones that you reach on the way to achieving the organization's goal. While goals are broad, objectives are clear and quantifiable.

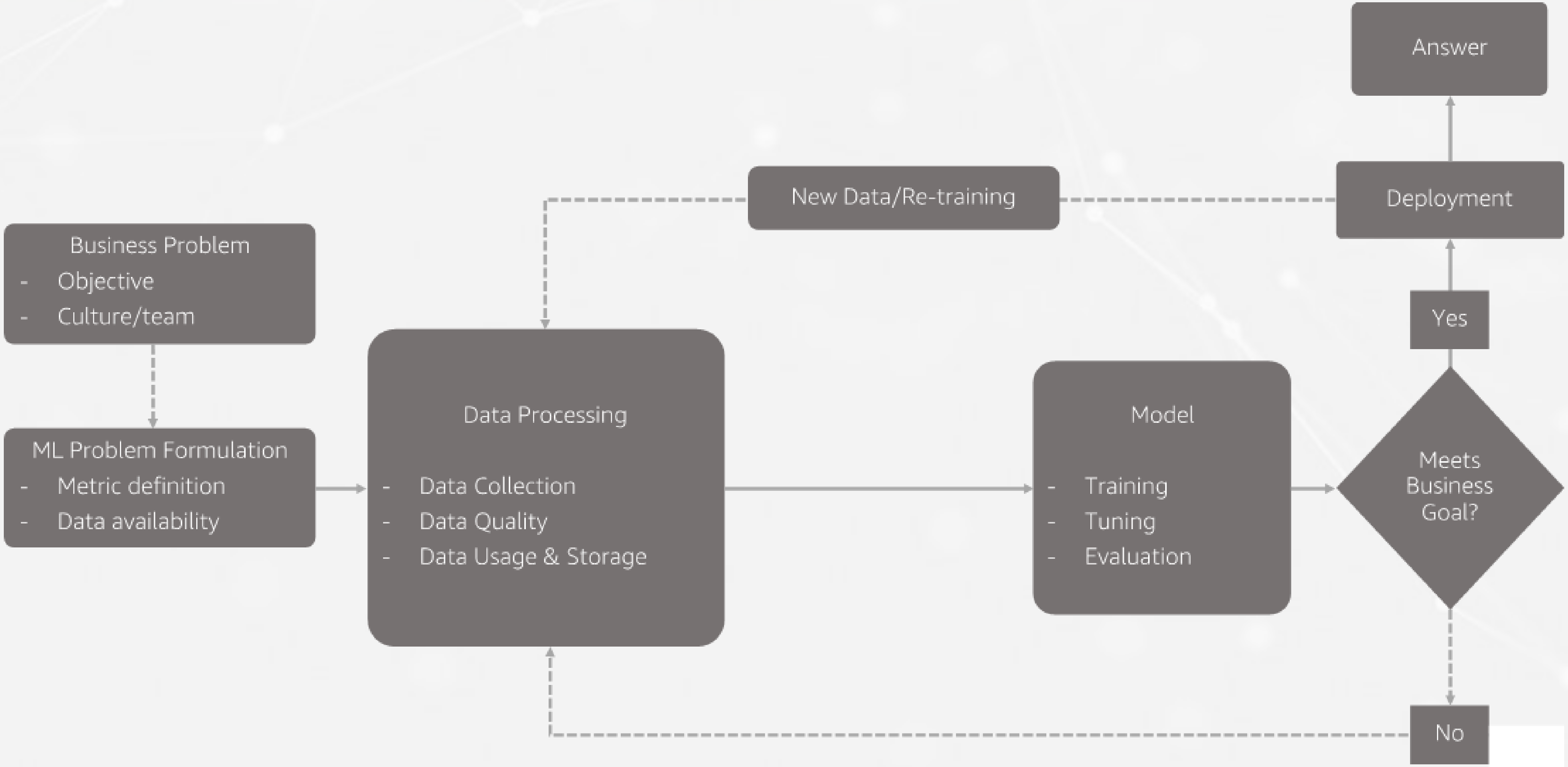
Translate the objective to an ML problem:

Express your objective using mathematical terms, e.g., predicting, calculating the likelihood, probability, propensity to, etc.

Outline your desired business outcome:

It should also be clear and quantifiable. If your ML model doesn't meet this outcome, you will need to go back and revise each part of your process as needed.

MLDQs along the Business Process



Decision Trees and Random Forests

Module 2: ML Business Goal and Performance Metrics

Algorithms

- Different business problems require different algorithms
- If you have labels: supervised learning
  - Numerical forecasting: regression
  - Binary classification: logistic regression
  - General classification: k-nearest neighbors or decision trees and random forest
- Quality of the labels (clean data) is upper bound on model performance
- If you have unlabeled data: unsupervised learning
  - Clustering: k-means

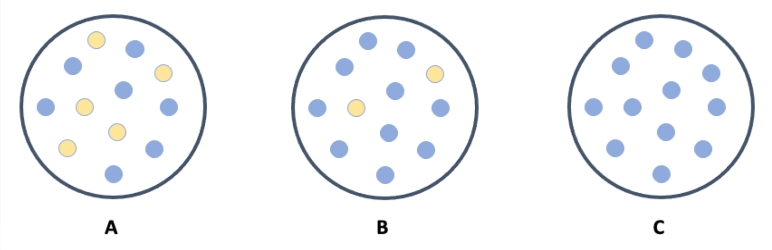
Training a Decision Tree

- Starting with a historical dataset containing features (input variables) and labels (output to be predicted), the goal is to create a model that predicts the latter by learning simple decision rules inferred from the former

Decision/label/output	Feature 1: cold	Feature 2: wind	Feature 3: windbreaker
Walk	58	Yes	Yes
Walk	75	Yes	Yes
Home	49	No	Yes
Home	47	Yes	No
Walk	69	No	No
Home	81	Yes	No
Walk	78	No	Yes

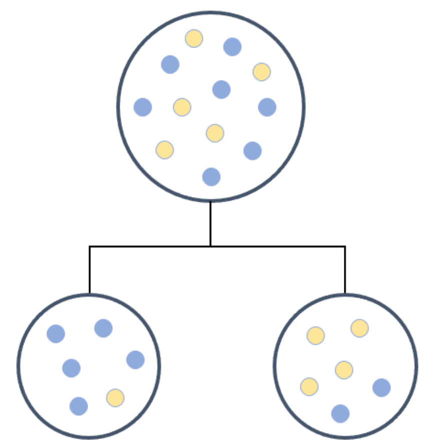
Node Purity

- At each branch, the algorithm picks the feature to create nodes of the highest possible purity (A = impure, C = pure)



- Such a split maximizes the information gain (measured using Gini index or entropy)

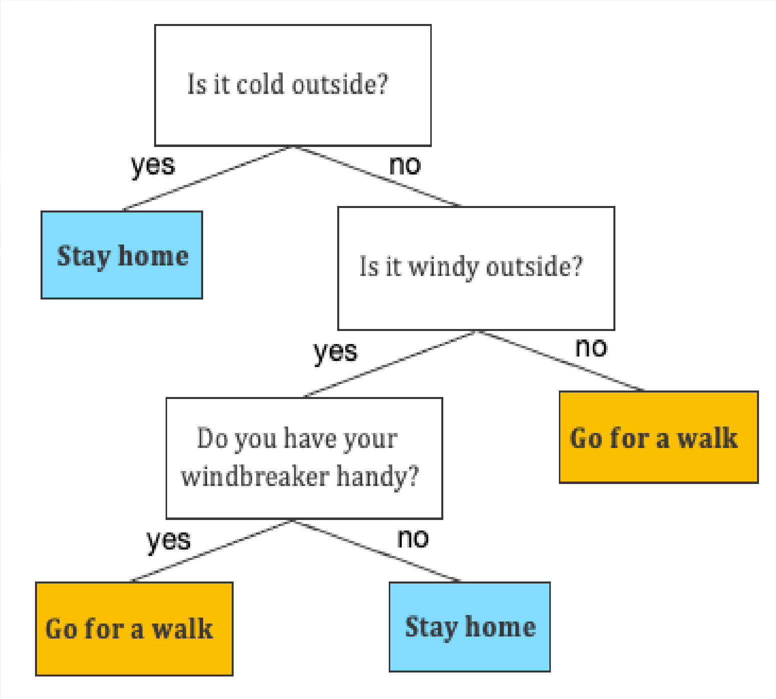
- Split generates a feature-based decision rule



Decision Tree

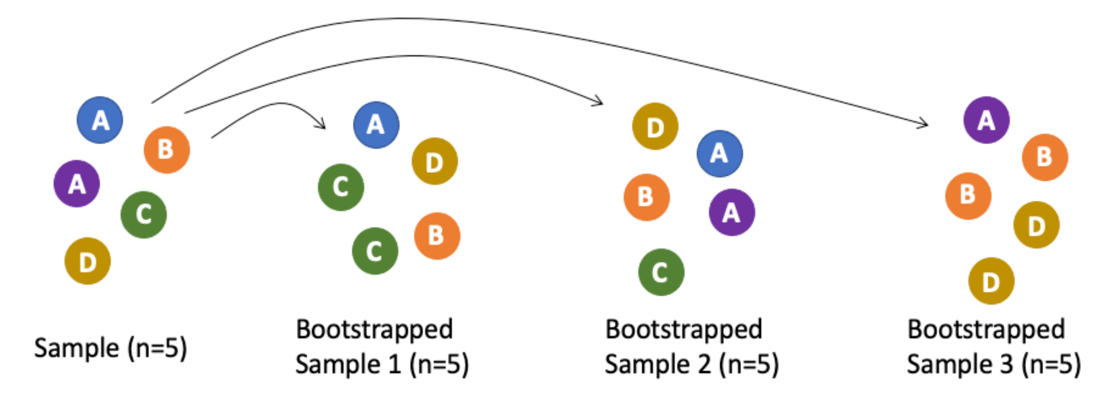
- Algorithm keeps splitting data until
  - » Maximum tree depth
  - » Minimum number of observations in each leaf

- How does tree make prediction given new data?



Random Forest

- Train an ensemble of decision trees from bootstraped samples of the dataset
  - Drawing random samples from our training set with replacement



- At every node of every tree, only a random small subset of features is allowed to be considered
  - Creates uncorrelated trees
- Prediction of the forest = aggregate vote of all the trees

Overfitting

- Decision trees are prone to overfitting, especially if the tree is too deep
  - » Too many splits: model is too complex and it simply memorized the noise

- Overfitting refers to the case when a model is so specific to the data on which it was trained that it is no longer applicable to different datasets

Model Bias and Variance

- Overly complex model is likely to overfit
  - » Captures too much noise from training data and hence will do poorly in testing data
- Overly simple model is likely to underfit
  - » Fails to capture true underlying relationship and hence will do poorly in testing data

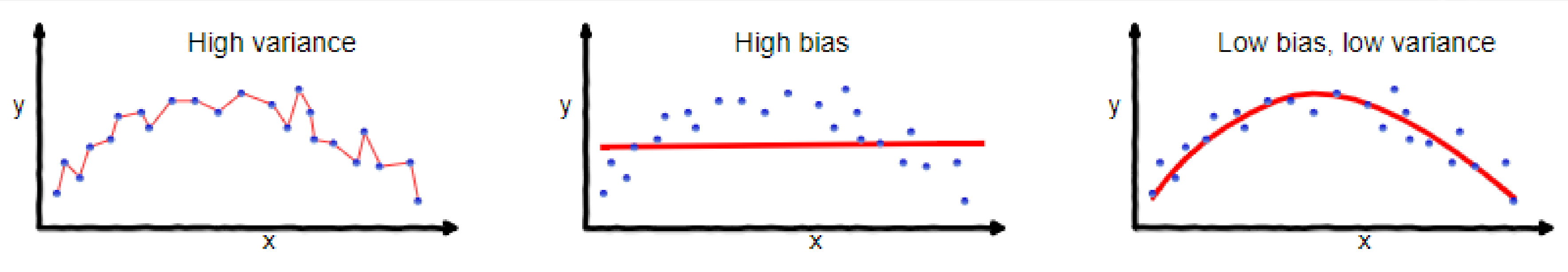
- Model error = bias^2 + variance + noise

- Bias = error from erroneous assumptions in model (underfitting)
- Variance = error from small fluctuations in training data (overfitting)

Bias-Variance Trade-Off

- Ideally, one wants to choose a model that both accurately captures the regularities in its training data, but also generalizes well to unseen data

- But typically, there is a trade off between variance and bias
  - » Simplify model to reduce variance but this increases bias
  - » Make model more complex to reduce bias but this increases variance



Performance Metrics

Module 2: ML Business Goal and Performance Metrics

Classification Model Performance

- The Janus model was trained using historical data
- It was then tested on historical data it had never seen before
  - Make a prediction using the features, but we know the actual outcome

	Truth: 1 purchase	Truth: 2+ purchases
Predict: 1 purchase	**True negative**	False negative
Predict: 2+ purchases	False positive	**True positive**

Classification Accuracy

		Prediction	
		Positive	Negative
True State	Positive	True Positive 18	False Negative 3
	Negative	False Positive 1	True Negative 15

Accuracy: The percent (ratio) of cases classified correctly

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

$$Accuracy = \frac{18 + 15}{18 + 1 + 3 + 15} = 0.89$$

$$(bad)\, 0 \leq Accuracy \leq 1\, (good)$$

Classification Accuracy with Imbalanced Data

True State \ Prediction	Positive	Negative
Positive	True Positive 2	False Negative 8
Negative	False Positive 2	True Negative 88

High Accuracy Paradox: Accuracy is misleading when dealing with imbalanced datasets - few True Positives, the ‘rare’ class, and many True Negatives, the ‘dominant’ class. High Accuracy even when few True Positives.

$$Accuracy = \frac{2 + 88}{2 + 2 + 8 + 88} = 0.90$$

Classification Precision

True State		Prediction: Positive	Prediction: Negative
True State	Positive	True Positive 2	False Negative 8
True State	Negative	False Positive 2	True Negative 88

Precision*: Accuracy of a predicted positive outcome

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{2}{2 + 2} = 0.50$$

$^{*}(bad)\ 0 \leq Precision \leq 1\ (good)$

Classification Recall

True State \ Prediction	Positive	Negative
Positive	True Positive 2	False Negative 8
Negative	False Positive 2	True Negative 88

Recall*: Measures model's ability to predict a positive outcome

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{2}{2+8} = 0.20$$

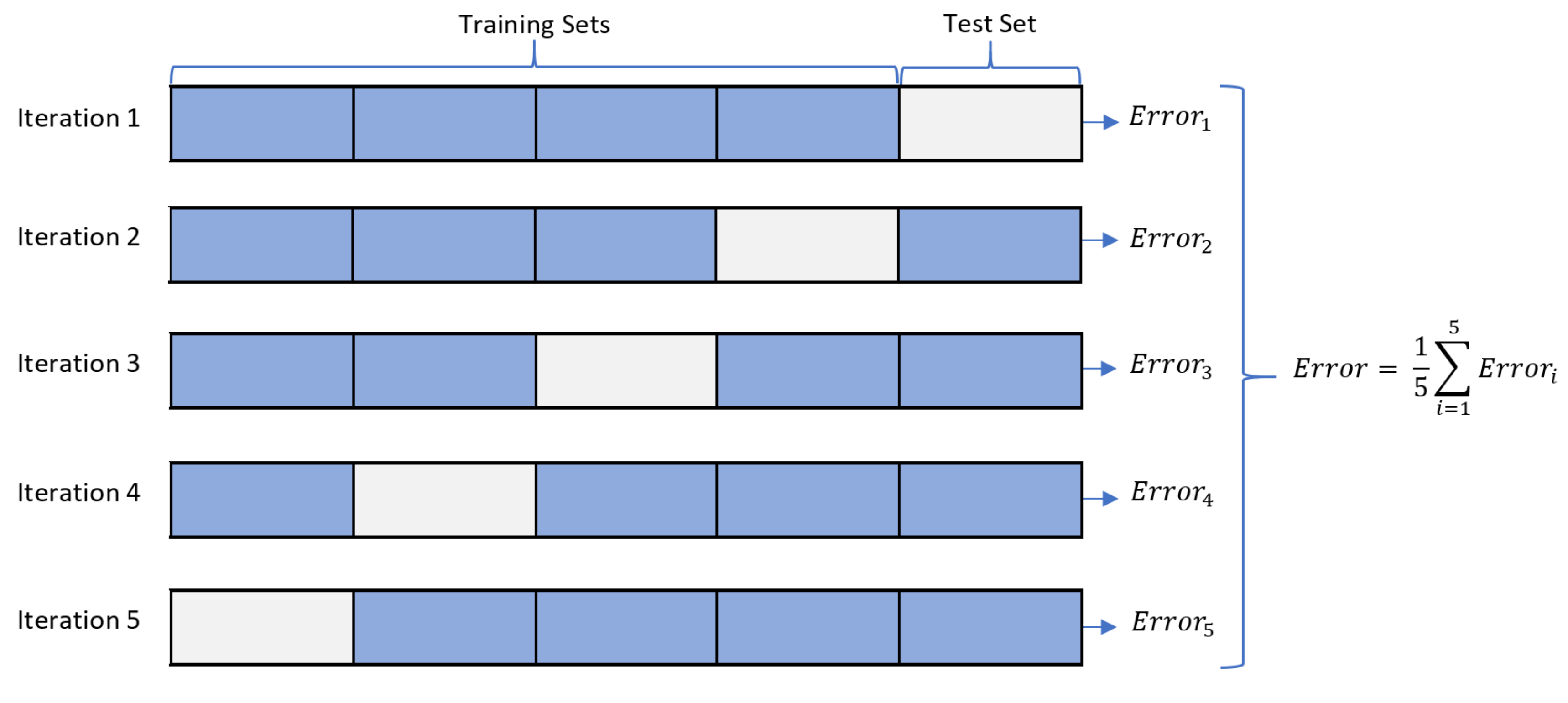
$^{*}(bad)\; 0 \leq Recall \leq 1\; (good)$

f1 Score for Classification

- f1 = 2*(precision*recall)/(precision+recall)
  - » Precision = TP/(TP+FP)
  - » Recall = TP/(TP+FN)
- Score between 0 (bad) and 1 (good)
- MLDQ: What mistakes are most costly for your business problem?
  - » Science and business must agree early on appropriate choice of performance metric given the business problem

Cross-Validation

- How do we evaluate accuracy especially if we do not have much data?
- Create n random data folds, train on k sets, test on remaining set, rotate k

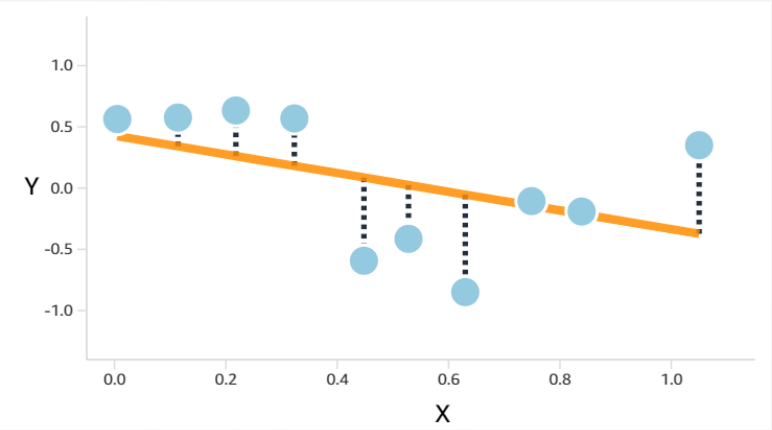


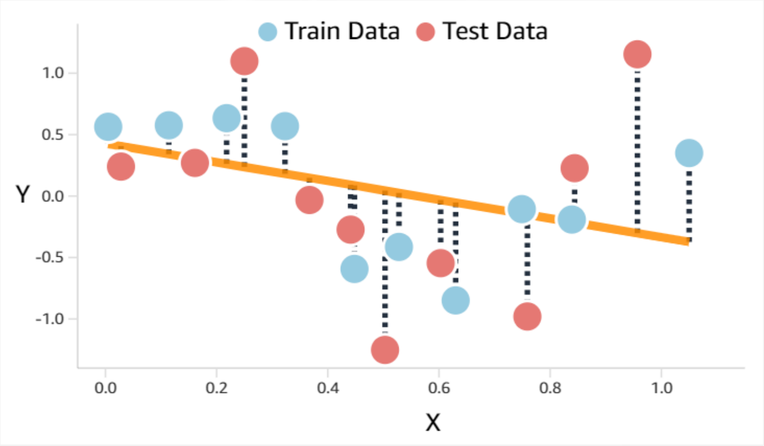
Regression Performance Metrics

- Algorithm minimizes the sum of squared errors
  - RMSE: square root of the mean squared error
  - MSE, MAE, and R^2 as well

- Test model performance on data model has never seen before

- Compare RMSE in testing data to RMSE in training data

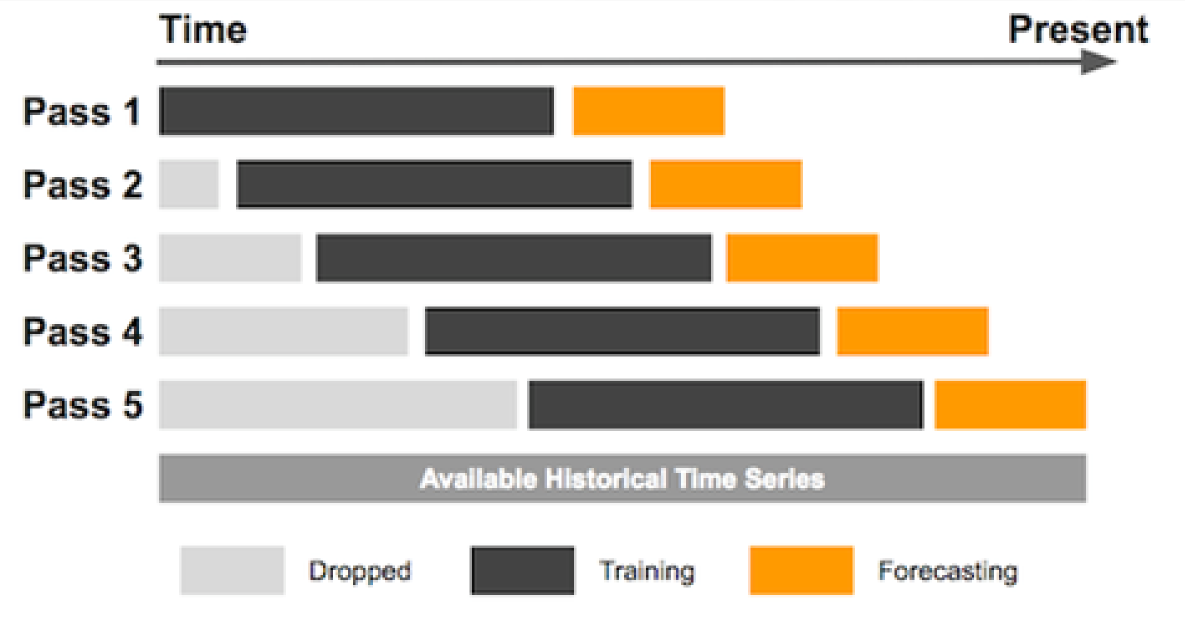




Backtesting

- With time-series data and time-series models, one cannot randomly draw historical data to create train/test sets
  - » Time-series features matter (seasonalities, past/present correlations…)

- We backtest



Additional Performance Questions for PM

- Ensure mapping between quality of model and goal of business problem

- % accuracy versus $ cost reduction
  - Benefit to company, contribution profits, DSI

- Debt of model setup
  - Monitoring, ongoing cost after production

- How confident are we in the model's predictions?
  - Uncertainty forecasting = confidence intervals
  - Sensitivity analysis

Conclusion

Module 2: ML Business Goal and Performance Metrics

Day 2 Takeaways

- The PM's job is not to do science but to manage the ML process
- Ask MLDQs
- Business success metric must line up with science model metric
- Decision trees and random forests are fast and powerful algorithms
  - f1 performance metric for classification

Course Schedule

Day One	Day Two	Day Three
ML LINGO, PROCESS, AND STAKEHOLDERS	ML BUSINESS GOAL AND PERFORMANCE METRICS	ML, CAUSALITY, AND EXPERIMENTS
Case #1: Churn	Case #2: Janus	Case #3: Catalove
Statistical learning and machine learning	Business and ML decision questions	Business heuristics versus science
Regressions and forecasting	Decision trees and random forests	Case #4: Emails and Sales
Process and stakeholders	Performance metrics	Causality and experiments

Day 3 Outlook

- Homework for today
  - You must read the case “Catalove: Targeting Amazon’s Christmas catalog”
    - You must answer the case’s online questions
  - You must read the case “More emails, more sales”
    - You must answer the case’s online questions

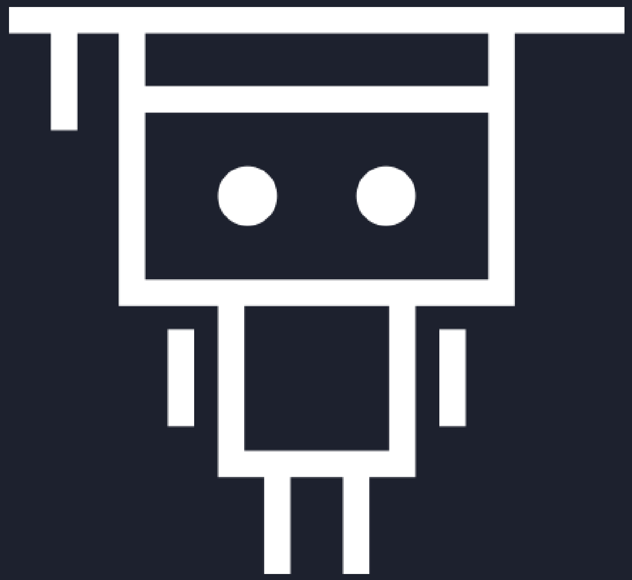
- We can now solve a business problem using ML
  - What is an experiment?
  - How do we reconcile predictions and causation?

Final Project

- To complete this course, you must
  - Submit a Word document with your proposed solution to the “Nudge Prime” case
    - Deadline is Thursday midnight PST
  - Review a peer’s submission, provide feedback, submit your feedback
    - Deadline is Friday midnight PST
  - After completion, student and manager receive confirmation email

- ML solution to business problem
  - What is the business problem?
  - What is the ML version of the business problem?
  - Data, labels, features, success and performance metrics, experiment, team?

MLU-Accelerated GRADUATE



Thank you!