Machine Learning for Science & Society

Contents

Syllabus

Notes

- Spring 2022
- time: MW 3:00-4:15pm
- Professor: Sarah Brown
- course number: CSC 592: Topics in Computer Science
- Credits: 4

In this class, we will address the challenges in applying machine learning to scientific research and in high stakes social contexts. On the science side, we will examine the role of ML in research, in particular how it works within knowledge production and how to evaluate ML in line with domain norms. On the social side, we will consider how to ensure ML-based algorithmic decision making systems uphold social values, with a focus on fairness. While these two applications are distinct, many of the challenges translate into common technical problems. Some of the common challenges include:

- missing data
- noisy or missing labels
- multiple objectives

We will look at a range of strategies for identifying and mitigating these problems including:

- robust evaluation
- model inspection
- explanations
- interpretable models

Format

This will be a synchronous course offered in person.

The course will involve:

- reading and evaluating ML research papers
- facilitating and participating in class discussions of the papers
- producing a replication, demo, or illustration of one concept covered for a broader audience
- completing a project using ML in a scientific or social domain
- writing a CS conference style (short & concise) final paper on their project

graduate students are encouraged to do a project related to their research

Prerequisites

To be successful in this class students should have:

- past experience with machine
- basic programming skills
- familiarity with concepts in probability, linear algebra, and calculus that appear in ML

varying skill in these topics is ok, but a general understanding of the basic ideas is important.

Complete this Google form to request a permission number from Professor Brown to enroll in this course. Note that you must be enrolled at URI to take this course and be logged into your URI google account to view that form.

Basic Facts

Meetings

This class will meet on Monday and Wednesday 3-4:15pm. via Zoom (link provided to registered students via BrightSpace)

> **Tip**
>
> You can subscribe to the Brightspace Calendar to access it in your favorite calendaring tool.

Instructor

Professor Sarah M Brown is an Assistant Professor in Computer Science. Her current research aims to answer the question, “How can machine learning produce AI systems that make fair decisions?”

Office Hours

By appointment, link on Brightspace.

Schedule

We will meet synchronously via Zoom: Tu 5:30-8:15

This course will proceed in three main parts: overview, deep dives, and wrap up.

Structure

Overview

In the first part of the course we will review ML basics, set norms for interaction and complete a survey of the topics that we will cover for the rest of the semester.

In this part of the class, Professor Brown will lead synchronous sessions. Students will be responsible for reading overviews, refreshing background material, and choosing an area for their course project. Students will start with an introductory demo or replication as a mini project.

Deep Dives

During the middle of the course we will spend one week on each topic. There will be 1-3 papers to read each week.

Students will be responsible for presenting papers in class on a rotating basis.

During this time students will have milestones where they need to complete interim steps for their course project. The first milestone will be a proposal that includes the specific products for the remainder of the milestones based on a template.

Conclusion

In the end of the course, we will focus on integrating ideas across multiple topics.

We will also workshop students’ projects, giving substantive feedback prior to the final submissions.

Final projects will be evaluated through a presentation and paper

Weekly topics

Class	Topic	Reading	Activities
2021-01-29	Introduction	None	introductions, expectation setting
2021-02-01	Probability Review	Model Based ML, chapter 1	reading discussion, setting
2021-02-03	ML Process & Mutual information preview	Scikit learn getting started,	live coding
2021-02-08	Missing Data: Intro strategies	Handling Missing Values when Applying Classification Models & Missing data imputation using statistical and machine learning methods in a real breast cancer problem	Paper discussion led by Daniel
2021-02-10	Missing data with graphical models and causal reasoning	Graphical Models for Inference with Missing Data & Missing Data as a Causal and Probabilistic Problem	Paper discussion led by Julian
2021-02-15	Current Challenges in Missing data	Handling Missing Data in Decision Trees: A Probabilistic Approach & How to miss data? Reinforcement learning for environments with high observation cost	Paper discussions by Xavier and Zhen
2021-02-17	Current Challenges in Missing data	How to deal with missing data in supervised deep learning	Paper discussion by Madhukara, Replication & testing discussion,
2021-02-22	Fairness	fairml classification chapter and friedler empricial comparison paper	Empirical setup
2021-02-24	Fairness	Reading	preview of lasso and admm constraint to multiobjecitve reformulation
2021-03-01	Multi-objective & constrained opt	Elastic Net	Paper presentation by Daniel, try out elastic net & LASSO in scikit learn
2021-03-03	Multi-objective & constrained opt	A critical review of multi-objective optimization in data mining: a position paper	Paper presentation and discussion by Zhen
2021-03-08	Latent Variable Models	Gaussian Mixture Models and Topic Models	Paper presentaiton by Xavier
2021-03-10	Latent Variable Models	Indian Buffet Process and Auto-Encoding Variational Bayes	Paper presentation by Madhukara
2021-03-15	Missing or Noisy labels	Learning with Noisy Labels and Semi Supervised Learning	Julian and Daniel
2021-03-17	Noisy Labels as a model for Bias	Recovering from biased data: Can fairness constraints improve accuracy and Fair classification with group dependent label noise	Zhen
2021-03-22	Interpretable & Explanation Intro	A Survey of Methods for Explaining Black Box Models	Xavier
2021-03-24	A Case for Interpretability over Explanation	Why are we explaining black box models and Learning Certifiably optimal rule lists for categorical data	Madhukara
2021-03-29	Models for Explanation	Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV) and A unified approach to interpreting model predictions	Zhen
2021-03-31	Choosing Explanations and using explantions	How can I choose an explainer? An Application-grounded Evaluation of Post-hoc Explanations Actionable Recourse in Linear Classification	Daniel
2021-04-05	What are the risks of explanations	Model Reconstruction from Model Explanations	Xavier
2021-04-07	What does Interpretable mean	Towards A Rigorous Science of Interpretable Machine Learning and Towards falsifiable interpretability research	Madhukara
2021-04-12	Meta issues	The Scientific Method in the Science of Machine Learning and Value-laden Disciplinary Shifts in Machine Learning	Sarah

Class	Topic	Reading	Activities
2021-04-13	Meta issues	Roles for computing in social change	Sarah
2021-04-19	Project Presentations	projects	presentations with peer feedback
2021-04-21	Project Presentations	projects	peer feedback
2021-04-26	Review and Project Reflections	Paper feedback	presentations with revision plans

Table 1 Schedule

Learning Outcomes and Evaluation

This course has goals with respect to the knowledge and research skills.

Evaluation will be with respect to each of the outcomes and based on a level of mastery: general awareness, competency, or mastery.

By the end of the course students will be able to:

- Critique common ways that social or scientific applications of ML require violating ML algorithm assumptions and ways to mitigate or adapt.
- Evaluate ML Research papers for their applicability to scientific and social applications of ML.
- Communicate about ML and its limitations work to varied audiences
- Apply ML to scientific and social data responsibly

Activities

- reading and evaluating ML research papers
- facilitating and participating in class discussions of the papers
- producing a replication, demo, or illustration of one concept covered for a broader audience
- completing a project using ML in a scientific or social domain
- writing a CS conference style (short & concise) final paper on the project

Evaluation

The grading scheme is rooted in achieving the learning outcomes.

Presentations (20%)

Presenting papers and participating in class will contribute to demonstrating a basic awareness at each of learning objective.

Each class session will be evaluated on if you contribute to discussion or not. This includes both asking questions and answering questions.

Each time you present will be evaluated on specification, your presentation should:

- summarize the key takeaways for the reading(s) in your own words
- summarize key details for understanding to facilitate the discussion
- discussion of strengths and weaknesses of the paper & method
- describe how this paper relates to bigger ideas in the course or your own work

You'll present 5 times, but must meet specification for at least 3.

When you present you don't have to have all the answers, you can have open questions.
The goal is that you guide the discussion by doing the above and opening the floor up for questions.

Questions that will help organize your preparation, but may apply variably to different readings:

- What is the key question that drove the research?
- What is the main finding?
- What is the model assumed in the paper?

- Did they include experimental results? If so:
  - do the experiments support the claims?
  - what additional experiments would help make the result make more sense?
  - how broad are the experiments, are the context-specific or general?
- Is there an analytical result? if so:
  - do the conditions for the proof make sense?
  - are they realistic?
  - what questions do you have about the proof?

Discussions and Exercises (20%)

For each topic we cover in class, you should engage fully in the class discussion and practice exercises that are provided if applicable.

To demonstrate engagement you must:

- provide a good faith attempt at any exercises provided
- contribute to the discussion (comments and questions both count)

Translation mini project (10%)

For this assignment you can choose any topic other than the one your project is for and produce a short demo, illustration, or replication that makes some aspect of the the topic accessible for a broader audience.

For this, you must submit a one paragraph proposal that describes your demo Once that’s approved that it will count, you have two weeks to build your demo or replication. The latest your demo may be submitted is at the same time as your final project.

The proposal (2%) will be graded on specification and may be resubmitted until successful. Your demo proposal must:

- state the topic from class your demo relates to
- state the format/medium your demo will take:
  - illustration, replication, interactive visualization, etc
- describe the target audience (a particular type of scientists, impacted people, software engineers, layperson, etc)
- describes what your demo will do by answering the relevant questions from the list below:
  - what will a person learn by reading/ using your demo?
  - if it’s interactive what will vary? what will be the inputs?
  - what specific result will you replicate?
- describe a demo that is an appropriate scope (not too large or too small)

The demo will be graded on specification (6%) and can be revised and resubmitted one time. Your demo must:

- describe a topic accurately
- be accessible to the specified topic model
- meet the description in the proposal

With your demo or after, submit a one paragraph reflection (2%) describing what you learned doing this exercise. The reflection will be graded on completion.

Project (50%)

The final project is a chance to dive deeply into one of the course topics. It has the following timeline. Percentages below are of the total grade.

Date	Milestone	Submission format	grading
2021-02-19	Area Selection	Consultation meeting and general questions	feedback only
2021-03-01	Topic Selection	Objectives and scope of work	completion (5%)
2021-03-15	Proposal	Problem statement, lit review, method	specification, with revisions (10%)
2021-04-02	Checkin	Consultation meeting and prelim result	specification, no revision (5%)
2021-04-13	Rough draft	Draft ready for peers to read	feedback only, per paper specs
2021-04-19,21	Presentation	talk in class	specification (10%)
2021-04-26	revision plan	plan for final revision, minor extensions	feedback only, per paper specs
2021-05-07	final paper	final paper submitted for grading	specification (15%)
2021-05-07	final reflection	final paper submitted for grading	completion (5%)

Proposal Specifications

Submit a 1.5- 2page proposal in the ACM Proceedings format.

Your proposal should include a concise problem statement, a preliminary literature review that situates your project, a description of method(s) you will use to answer your questions in your project, and the expected outcomes of your project.

The proposal will be graded on if it meets the specification or not, but you will be able to revise and resubmit if the first submission does not. To meet specification it must:

- be the right length
- be the right format
- include all sections
- be written clearly
- describe the problem, clearly identifying what the specific goals of your project are
- describe a tractable project
- summarize relevant literature for the problem context
- summarize relevant course-related literature for your project
- describe what you will do in your project
- describe what the end outcome of your project.

Checkin Specifications

- scheduled on time
- at least one dimension of progress from proposal

Presentation Specifications

Your presentation should:

- include an agenda for the talk
- describe the problem
- summarize relevant backround
- clearly identify what you did
- describe findings
- include concluding remarks on reflection/possible extensions

Paper Specifications

Your final paper should include a concise problem statement, a complete literature review that situates your project, a description of method(s) used your project, findings, and a discussion or future work section.

For it to meet specification it must:

- be the right length
- be the right format
- include clearly marked sections indicating the required content
- be written clearly
- describe the problem, clearly identifying the specific goals of your project
- summarize relevant literature for the problem context
- summarize relevant course-related literature for your project
- include clear description of what was accomplished
- include a clear summary of results (may include null results/ failed findings)

Spring 2021

Class 1: Introductions

Introductions & Goals

Course Admin

- Brightspace
- Zoom
- Google docs or markdown in the future?
- Website

Learning outcomes

knowledge research

- identify common problems and solutions in scientific application of ML
- identify common challenges and solutions for social applications: fairness,
- implement and extend research papers

Activities

- reading and evaluating ML research papers
- facilitating and participating in class discussions of the papers
- producing a replication, demo, or illustration of one concept covered for a broader audience
- completing a project using ML in a scientific or social domain
- reflect on methodologies used in this type of research
- writing a CS conference style (short & concise) final paper on their project

Model Based ML and this course

https://www.mbmlbook.com/toc.html

- missing data
- noisy or missing labels
- multiple objectives

We will look at a range of strategies for identifying and mitigating these problems including:

- robust evaluation
- model inspection
- explanations

- interpretable models

ML and Probability Review

admin

- collaborative notes
- brightspace will be updated later this week
- grading details by Wed
- environment for coding demos

More formalism

- model
- prediction algo
- cost function
- objective

Probability

- sample distros

Practical Application of ML & Pipelines

Class 3: ML Pipelines

Goals when using ML

1. Understand about the data (data science/ actual science) probability more statistics, maybe fit another examine model parameters, inspect them
2. understanding about Naive bayes fit different data varies
3. claims about the learning algorithm run multiple algorithms on the same data possibly multiple data

Basic setup

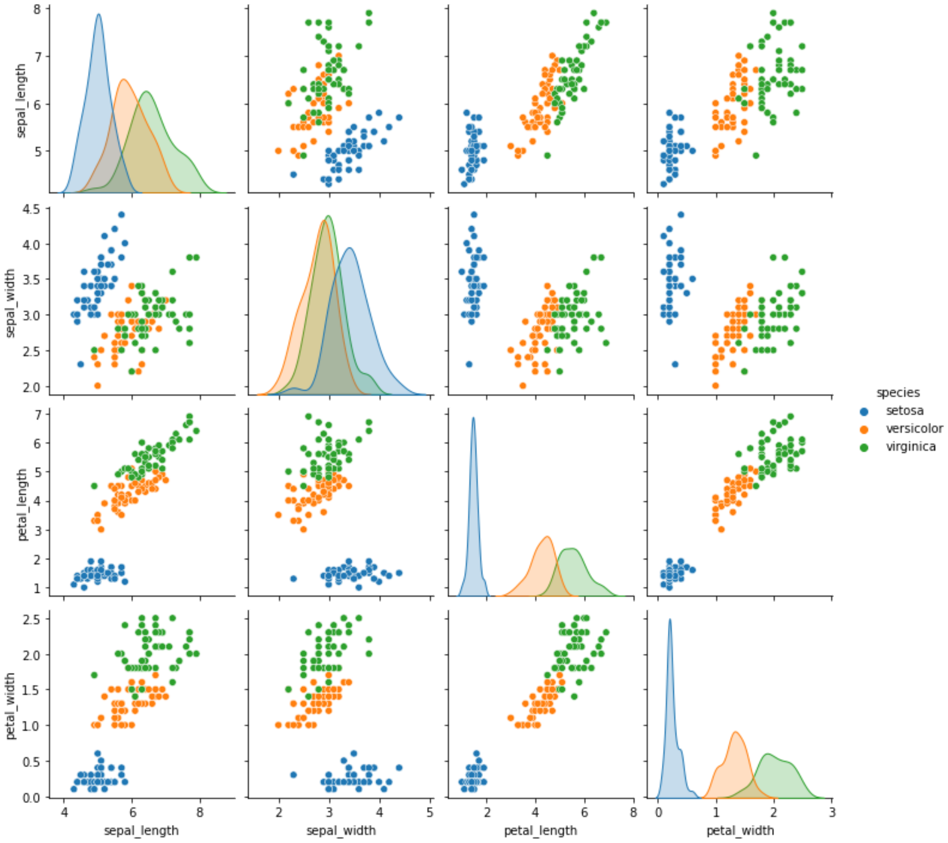
1. test train
2. training parameters
3. estimator objects
4. fit model parameters
5. metrics
6. cross validation

```
import pandas as pd
import seaborn as sns
import numpy as np
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import confusion_matrix, classification_report
from sklearn import datasets
```

```
iris_df = sns.load_dataset('iris')
```

```
sns.pairplot(iris_df, hue='species')
```

```
<seaborn.axisgrid.PairGrid at 0x7fdff8ac3110>
```



```
X,y = datasets.load_iris(return_X_y=True)
```

```
X.shape
```

```
(150, 4)
```

```
y.shape
```

```
(150,)
```

```
X_train, X_test, y_train, y_test = train_test_split(X,y,)
```

```
gnb = GaussianNB()
```

```
gnb.__dict__
```

```
{'priors': None, 'var_smoothing': 1e-09}
```

```
gnb.fit(X_train,y_train)
```

```
GaussianNB()
```

```
gnb.__dict__
```

```
{'priors': None,
 'var_smoothing': 1e-09,
 'classes_': array([0, 1, 2]),
 'n_features_in_': 4,
 'epsilon_': 3.1439253826530615e-09,
 'theta_': array([[5.01428571, 3.4452381 , 1.45714286, 0.24285714],
        [5.82285714, 2.70571429, 4.18857143, 1.3       ],
        [6.50571429, 2.93428571, 5.50857143, 1.99428571]]),
 'var_': array([[0.12360545, 0.15438209, 0.02482994, 0.01054422],
        [0.2303347 , 0.09711021, 0.22729796, 0.04      ],
        [0.37939592, 0.07996735, 0.30592653, 0.08625306]]),
 'class_count_': array([42., 35., 35.]),
 'class_prior_': array([0.375 , 0.3125, 0.3125])}
```

```
X_test[0]
```

```
array([6.7, 3. , 5. , 1.7])
```

```
y_pred = gnb.predict(X_test)
```

```
y_pred[:5]
```

```
array([2, 2, 1, 0, 1])
```

```
y_test[:5]
```

```
array([1, 2, 1, 0, 1])
```

```
confusion_matrix(y_test, y_pred)
```

```
array([[ 8,  0,  0],
       [ 0, 10,  5],
       [ 0,  0, 15]])
```

```
gnb.score(X_test,y_test)
```

```
0.868421052631579
```

```
gnb2 = GaussianNB(priors=[.5,.25,.25])

gnb2_cv_scores = cross_val_score(gnb2,X_train,y_train)
```

```
np.mean(gnb2_cv_scores)
```

```
0.9553359683794467
```

```
gnb_cv_scores = cross_val_score(gnb,X_train,y_train)
```

```
np.mean(gnb_cv_scores)
```

```
0.9553359683794467
```

```
print(classification_report(y_test,y_pred))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00         8
           1       1.00      0.67      0.80        15
           2       0.75      1.00      0.86        15

    accuracy                           0.87        38
   macro avg       0.92      0.89      0.89        38
weighted avg       0.90      0.87      0.86        38
```

```
gnb.predict_proba(X_test)
```

```
array([[6.91259697e-158, 2.02779380e-002, 9.79722062e-001],
       [5.57443611e-220, 2.13139294e-004, 9.99786861e-001],
       [6.39966685e-097, 9.92924692e-001, 7.07530770e-003],
       [1.00000000e+000, 4.67010851e-017, 6.12586347e-024],
       [1.04945137e-113, 9.73531592e-001, 2.64684083e-002],
       [1.42284693e-063, 9.99921877e-001, 7.81227042e-005],
       [2.80075077e-263, 2.77856412e-007, 9.99999722e-001],
       [3.45127426e-107, 9.06239381e-001, 9.37606187e-002],
       [1.00000000e+000, 1.59329672e-014, 8.84248378e-021],
       [5.47437479e-090, 9.99496846e-001, 5.03154037e-004],
       [3.44208825e-141, 1.35983611e-001, 8.64016389e-001],
       [1.22964296e-124, 4.00880662e-001, 5.99119338e-001],
       [7.51410601e-287, 1.49470734e-011, 1.00000000e+000],
       [9.59642930e-213, 1.51285370e-006, 9.99998487e-001],
       [3.41855866e-189, 1.14087619e-003, 9.98859124e-001],
       [2.40303924e-085, 9.99400959e-001, 5.99041022e-004],
       [6.86353628e-288, 8.55590829e-011, 1.00000000e+000],
       [7.75830380e-219, 1.96208822e-007, 9.99999804e-001],
       [5.11316281e-107, 9.42646742e-001, 5.73532579e-002],
       [1.00000000e+000, 2.30661854e-010, 2.12374555e-017],
       [3.07524384e-224, 1.38357344e-007, 9.99999862e-001],
       [5.82816017e-071, 9.99987066e-001, 1.29340942e-005],
       [3.39598438e-096, 9.95793005e-001, 4.20699459e-003],
       [1.00000000e+000, 6.12033613e-017, 1.46941701e-023],
       [1.20019371e-208, 1.41634369e-007, 9.99999858e-001],
       [1.00000000e+000, 4.11944045e-017, 8.19744263e-024],
       [6.40726523e-174, 9.18327432e-003, 9.90816726e-001],
       [1.00000000e+000, 2.06995850e-013, 1.12904573e-019],
       [6.43924944e-250, 3.87576323e-009, 9.99999996e-001],
       [2.40419465e-131, 3.85543557e-001, 6.14456443e-001],
       [1.00000000e+000, 3.39619724e-017, 5.60193149e-024],
       [0.00000000e+000, 6.93028669e-013, 1.00000000e+000],
       [2.55787153e-232, 6.77790313e-007, 9.99999322e-001],
       [2.60103834e-085, 9.99497127e-001, 5.02872761e-004],
       [1.00000000e+000, 4.63813456e-016, 7.79701526e-023],
       [2.39203803e-247, 9.38063429e-008, 9.99999906e-001],
       [4.79950886e-155, 4.31983078e-002, 9.56801692e-001],
       [2.01501470e-155, 3.81967304e-001, 6.18032696e-001]])
```

Class 4: Missing Data: Basic techniques

Evaluation of missing data at training

- mulitple imputation
- ML based was better than imputation which is better than dropping samples
- example datasets: 45% of patients have at least 1 missing value

Imputation

- Mean imputation:
  - insert the mean based onthe other values
- Hot deck
  - mean-like with similarity
- Multiple inputation
  - 3 diff ways

Imputation ML

- MLP
  - fully connected
- Self organization
  - competitive learning
  - NN on modle of nodes in 2d grid,
- KNN
  - select closest complete case to impute values from
  - expensive for large datasets due to need to search everywhere for each missing value

Testing

- Train NN based on data imputed with each technie

Conclusions:

- in general, any imputation was better than deletion
- ML based performed better

Discussion & Questions

- interesting that even simple methods provide improvement
- SOM is sort of unclear how does that work?
- Review of MLP and sigmoid

Handling missing values At application time

- reduced models vs imputation.
- broad approach
- 15 common datasets

Techniques:

- Discard
- Acquire missing values
- Imputation
  - predictive value imputation
  - distribution based
  - unique values
- Reduced Feature Models
  - retrain for different feature models

Feature imputability impacts the distribution or predictive type of imputation

More complex model

- decision tree with bagging
- again, reduced model is the best strategy

Hybrid Models for efficient prediction

- reduced models
- a hybrid is a complete model with stored subset for most common missing features
- Reduced feature enseble
  - N models for N features
  - each one is missing one feature
  - average these together for final prediction
  - substantial reduction in when there is a single feature is missing

    - combine with imputation for multiple features
    - relative accuracy is better than imputation

General takeaways

- reduced models vs imputatation is a large improvement
- this is sort of an imputation

Weaknesses

- Didn’t check unique value imputation
- MCAR
- focused on

Overall Discussion

- How might the two problems interact?
    - if missing data at both train and prediction…
    - train using missing data without imputation for training the separate models
- Questions on these ideas
- What additional things might you need to consider when choosing one?
    - feature imputability at training
- what to do with time series data
- How to check if missing CAR?
    - look at collection technique
    -
- what do to with varying data per person
    - LSTM for time series data
    - hierarchichal modeling other wise
    - example of hierarchical with time series also

For Wednesday

1. Graphical Models for Inference with Missing Data
2. Missing Data as a Causal and Probabilistic Problem

Missing Data 3

Handling Missing Data in Decision: A probabilistic approach

key ideas

- A decision tree’s structure and notation
- Review of imputation
    - Predictive value imputation
        - mean, median or mode
        - make assumption that features are indpendent
        - surrogate splits, partition data using another feature to
- XG Boost

Expected Predictions:

- impute all possible completions as once to avoid strong dist assumptions
- consistent for MCAR and MAR
- expensive, but density can help reduce

- tractably compute the exact expected predictions
- loss minimization

Experiments

- for a single dataset, outperforms in general

Discussion

- generally easier
- given single dataset, of results, how much do we trust this?
- what does this provide as an advantage
- NP hard

How to miss data?: Reinformcent leanring for environgments iwith high obseration cost

Key points

Reinforcement learning

- cost associated with making accurate observations
- goal directed
- RL agent tries to

Problem setting:

- \(o \sim p_0(o_t |s_t; \beta)\)
- beta is accuracy og obs
- r is old reward

Scenario A:

- observed cangle vs

Big picture: manipulating how the data collection

Discussion

- survivorship bias?
- right left imbalance for figure 3
- simple pendulum example helped overcome the background lacking
- figures

General

Try writing out a missingness graph for a problem of choice, some scenario where you imagine there would be missing data, or an example dataset that you can find.

MIssing data

supervised

Background

- Hadamard
-

Readings for next week:

http://sorelle.friedler.net/papers/fairness_comparison_fat19.pdf https://fairmlbook.org/

- introduction and classification chpaters (1 and 2)

Intro to Fairness

Handwritten notes on shared drive

Comparison of Fairness Interventions

Paper discussion

Spring 2022

Notes will be added after the semester starts.

By Sarah M Brown