Course Plane

Dear all

This course plan is meant to serve as your guide throughout the course and as an outline of my expectations to you. However, it might be subject to slight change.

About the course

The Course is centered around machine learning in conflict prediction and not machine learning in general. Thus, we will only look at some very specific algorithms. There are tons more to look at, but we simply do not have time in this course to cover everything. Indeed, this course should be seen as an “advanced introductory” course. It is advanced because you need a lot of prerequisite knowledge and/or skills to do machine learning in conflict prediction, but it is also introductory in the sense that we only have time to go through the fundamentals. Hopefully, your future studies will allow you to dive deeper into the subject.

Naturally, if you want to go into more advanced stuff (e.g. Gaussian Processes and Neural Networks) please let me know and i’ll provide you with some resources.

How to approach the curriculum

You all come from different backgrounds; some know a lot about coding, some know a lot about conflict, some know a bit of both and some are quite new to all of it. As such, you should prioritize you weakness when going through the curriculum. If you have not programmed in python before focus extra on VanderPlas and Friedman. If you are a used to python then pay less attention to VanderPlas and if you know your machine learning then pay less attention to Friedman. If you know your python and machine learning focus on the articles. The core of the course is the practical exercises. As such, you will likely not have time to read the whole curriculum throughly (assuming you have other courses and meaningful lives). As such the curriculum should be seen as a collection of potential useful references for this course as well as your future endeavors.

For people not familiar with Python I highly recommend going through VanderPlas (2016) chapter 1-5 as soon as possible.

Exercises

All-in-all there will be 6 exercises, each divided into two parts. E.g. 1A, 1B, 2A, 2B etc. You will get a new part every week (except in the fall break). The deadline for submitting the exercises to feedback is every second week (Monday). Here you are expected to hand in both part A and B of a given exercise. E.g 1A and 1B (see plan below).

There will be no (official) group work. Everybody will be coding. You will evaluate each others work (two-way blind) through Peergrade. This will ensure you’ll get a lot of direct feedback but also that you get exposed to other peoples solutions and code. Thus, if you did not get it right, you

will be able to revisit your code and correct it. If you do this throughout the course, the exam will be a breeze.

The last couple of weeks are kept clear (no exercises here) so you can focus on incorporating feedback (if you have not done that yet) and better solutions into your existing code. If your code is perfect feel free to dive deeper into the curriculum. Thus, you should have plenty of time to get your code and skills ready before the exam.

There will naturally still be lectures these last weeks where I will talk about ‘what comes next’ and discord will be open for general help, guidance and questions.

Exam

It is a Portfolio Exam. It will mirror the exercises pretty closely with both coding tasks along with questions regarding understanding, interpretation, reflection, pros and cons etc. The main difference between the exam and the weekly exercises is that I will supply no code for the exam. However, fear not. If you have simply finished your exercises and incorporated (and reflected upon) the relevant feedback, you will have everything you need to get a passing - likely also good - grade.

Note on math and Friedman et al. 2001

There are a lot of math in Friedman et al. 2001. For some of you this will be fine. For some of you this will be inconvenient, and for some of you it’ll bring back substantial traumas. However, I implore you all to read through (the most important part of) it despite whatever disposition you might have towards math.

The book is perfectly readable without understanding all (or even most) of the math. And at some point the math might even start to make sense. Getting comfortable reading technical text will benefit you hugely down the line - even if the math still doesn’t make a 100% sense.

If you want to brush up you math skills (now or in the future) here are three book titles to get you started:

- **Easy**: Moore, Will H., and David A. Siegel. A mathematics course for political and social research. Princeton University Press, 2013.
- **Less easy**: Gill, Jeff. Essential mathematics for political and social research. Cambridge: Cambridge University Press, 2006
- **Still manageable**: Deisenroth, Marc Peter, A. Aldo Faisal, and Cheng Soon Ong. Mathematics for machine learning. Cambridge University Press, 2020.

---

1) Introduction: getting the data

Week 36

Agenda:

- How the course is structured

- Getting the data

Exercise

1A. Deadline Monday 14.09.2020
Peer Feedback

Must reads:

Everything on the curriculum for this first lecture is pretty central. Thats why it's on the curriculum for the first lecture. That being said, if you have not done so yet I highly recommend that you read: **Hegre, Håvard, et al. "ViEWS: a political violence early-warning system." Journal of peace research 56.2 (2019)**

Readings:

```
Hegre, Håvard, et al. "Introduction: Forecasting in peace research." (2017): 113-124.

Sundberg, R. and Melander, E. (2013). Introducing the ucdp georeferenced event dataset. Journal of
 Peace Research, 50(4):523 - 532.

Croicu, M. and Sundberg, R. (2017). Ucdp ged codebook version 18.1. Department of Peace and Confli
ct Research, Uppsala University

VanderPlas, Jake. Python data science handbook: Essential tools for working with data. " O'Reilly
Media, Inc.", 2016. Capter 1-4
```

---

2) Filling in the gaps: Interpolation of missing data

Week 37

Agenda:

- Getting PRIO dynamic data merged with the rest of the data
- Seeing some examples of predictions to get a grasp of the field and for inspirations (see curriculum).

Exercise

1B. Deadline Monday 14.09.2020

Must reads:

The curriculum for this lecture does not have a direct link to the excercies (1B). Instead it is meant to give you and idea of regarding state-of-the-art (SOTA) in the field. As such, if you only read one thing for this lecture it should be **Hegre, Håvard, et al. "ViEWS: a political violence early-warning system." Journal of peace research 56.2 (2019): 155-174.**

Readings:

```
Weidmann, N. B. and Ward, M. D. (2010). Predicting conflict in space and time. Journal of Conflict
 Resolution, 54(6):883 - 901.

Hegre, H., Nyg_ard, H. M., Karlsen, J., Strand, H., and Urdal, H. (2013). Predicting Armed Conflic
t, 2010{2050. International Studies Quarterly, 57(2):250 - 270.

Perry, C. (2013). Machine learning and conflict prediction: a use case. Stability: International J
ournal of Security & Development, 56(2(3)). (18 pages)

Hegre, Håvard, et al. "ViEWS: a political violence early-warning system." Journal of peace researc
h 56.2 (2019): 155-174.
```

---

3) Probing prediction: out-of sample predictions and evaluation metrics

Week 38

Agenda:

- Introducing the concept of out of sample prediction
- Create train and test
- Generate some baseline models to test

Exercise

- 2A Deadline Monday 28.09.2020

Must reads:

The curriculum for this lecture goes a bit more into the reason we use predictions to evaluate our models. If you only read one, let it be **Friedman, J., Hastie, T., and Tibshirani, R. (2001). The elements of statistical learning. Springer series in statistics New York, NY, USA. Chapter 7** . Don’t worry if you do not get everything, just try to understand the intuition. Alternatively see **Ward, M. D., Greenhill, B. D., and Bakke, K. M. (2010). The perils of policy by p-value: Predicting civil conflicts. Journal of Peace Research, 47(4):363 - 375** for a more ‘soft’ introduction.

Readings:

Colaresi, Michael, **and** Zuhaib Mahmood. "Do the robot: Lessons from machine learning to improve conflict forecasting." Journal **of** Peace Research 54.2 (2017): 193-214.

Ward, M. D., Greenhill, B. D., **and** Bakke, K. M. (2010). The perils **of** policy **by** p-value: Predicting civil conflicts. Journal **of** Peace Research, 47(4):363 - 375.

Schrodt, P. A. (2014). Seven deadly sins **of** contemporary quantitative political analysis. Journal **of** Peace Research, 51(2):287-300.

Friedman, J., Hastie, T., **and** Tibshirani, R. (2001). The elements **of** statistical learning. Springer series **in** statistics New York, NY, USA. Chapter 1, 2 **and** 7

---

4) Few and far between: Imbalanced data and evaluation metrics

Week 39

Agenda:

- Evaluating (imbalanced) data

Exercise)

- 2B Deadline Monday 28.09.2020

Must reads:

- If you only read one paper for this lecture it should be **He, H. and Garcia, E. A. (2008). Learning from imbalanced data. IEEE Transactions on Knowledge & Data Engineering, 9:1263 - 1284.**

Readings:

He, H. **and** Garcia, E. A. (2008). Learning from imbalanced data. IEEE Transactions on Knowledge & Data Engineering, 9:1263 - 1284.

King, G. **and** Zeng, L. (2001a). Explaining rare events **in** international relations. International Organization, 55(3):693 - 715

King, G. **and** Zeng, L. (2001b). Improving forecasts **of** state failure. World Politics, 53(4):623 - 658.

Goldstone, J. A., Bates, R. H., Epstein, D. L., Gurr, T. R., Lustik, M. B., Marshall, M. G.,Ulfelder, J., **and** Woodward, M. (2010). A global model **for** forecasting political instability. American Journal **of** Political Science, 54(1):190 - 208.

5 Drawing the line: Machine Learning Algorithms 1

Week 40

Agenda:

- Linear Regression (regression)
- Logistic Regression (Classification)
- Decision Trees (regression and Classification)
- (Optional Naive Bayes)
- (Optional SVM)

Exercise

3A Deadline Monday 12.10.2020

Must reads:

- If you only read one chapter for this lecture it should be 9 (since I assume you are all familiar with Linear and Logistic regression). Understanding Decision Trees will be central for understanding Random Forest and XGboost later on.

Readings:

```
Friedman, J., Hastie, T., and Tibshirani, R. (2001). The elements of statistical learning. Springe
r series in statistics New York, NY, USA. Chapter 3, 4, 9 (  not on curriculum but  6.6.3 for Naiv
e Bayes and 12 for SVM)
```

---

6 Stronger together: Ensemble model

Week 41

Agenda:

- Introducing the concept of ensembles
- Tying together with imbalanced data and undersampling
- Bayesian correction
- Quantifying (kinds of) Uncertainty

Exercise

3B

Must reads:

- To be honest you should read everything here. I think both articles are very worth-while, but the concept of ensembles is better presented in the book. Maybe skim the book if your are in a hurry, but do try to read the articles.

Readings:

```
Hegre, Håvard, Håvard Mokleiv Nygård, and Ranveig Flaten Ræder. "Evaluating the scope and intensit
y of the conflict trap: A dynamic simulation approach." Journal of Peace Research 54.2 (2017): 243-
26

Ward, Michael D., and Andreas Beger. "Lessons from near real-time forecasting of irregular leaders
hip changes." Journal of Peace Research 54.2 (2017): 141-156.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). The elements of statistical learning. Springe
r series in statistics New York, NY, USA. Chapter 16
```

FALL BREAK, week 42

---

7 Thinking about theory: Feature engineering

Week 43

Agenda:

- Connection to theory
- Feature engineering

Exercise

4A

Must reads:

Well, the only 'real' text on the curriculum for this lecture is simply an example of a proxy feature. I will upload **a more general note on actual feature engineering**. Read that.

Readings:

```
Weidmann, Nils B., and Sebastian Schutte. "Using night light emissions for the prediction of local
 wealth." Journal of Peace Research 54.2 (2017): 125-140.

a more general note on feature engineering
```

---

8) Thinning the herd: Feature importance and selection

Week 44

Agenda:

- Feature importance
- Feature selection

Exercise

4B

Must reads:

That text that you know you should have read, but you haven’t come around to yet - yes, that text. You know the text i’m talking about… You should read that text.

Readings:

```
That text we both know you haven't read yet..
```

---

9) A forest full of tress: Machine Learning algorithms 2, week 45

Agenda:

- Random Forest

Exercie

5A

Must reads:

Well.. That will be **Friedman, J., Hastie, T., and Tibshirani, R. (2001). The elements of statistical learning. Springer series in statistics New York, NY, USA. Chapter 15**. Understanding Random Forest will be central for understanding XGboost later on - and it is a canonical and seminal algorithm.

Readings:

```
Friedman, J., Hastie, T., and Tibshirani, R. (2001). The elements of statistical learning. Springe
r series in statistics New York, NY, USA. Chapter 15
```

---

10) Boosting

Week 46

Agenda:

- Presenting boosting
- Adaboost

Exercise

- 5B

Must reads:

Again, you really should read the one chapter presented here **Friedman, J., Hastie, T., and Tibshirani, R. (2001). The elements of statistical learning. Springer series in statistics New York, NY, USA. Chapter 10**
. It will be central for understanding XGboost later on.

Readings:

```
Friedman, J., Hastie, T., and Tibshirani, R. (2001). The elements of statistical learning. Springe
r series in statistics New York, NY, USA. Chapter 10
```

11) Connecting the dots: : Machine Learning algorithms 3

Week 47

Agenda:

- introducing the XGboost algorithm

Exercise

6A

Must reads:

This test is a must read: **Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pages 785-794. ACM**. Perhaps one of the best and most robust machine learning algorithms for structured data and general problems out there at the moment. This should be one of your go-to models.

Readings:

```
Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the
22nd acm sigkdd international conference on knowledge discovery and data mining, pages 785-794. AC
M.
```

12 Tying the knots: more XGboost

week 48

Agenda:

- More XGboost

Exercise

- 6B

13 Where to go from here: better models and novel data

Week 49

Agenda:

- Other algorithms: Gaussian Processes and Neural Networks
- Other data: text and images as data

Must reads:

Read at least **Mueller, H. F. and Rauh, C. (2016). Reading between the lines: Prediction of political violence using newspaper text. American Political Science Review, 2(112):358-375**. Its is a nice text and the idea has huge potential. The methods, however, are already a bit outdated but that is the name-of-the game in the field of machine learning right now. Everything moves really fast.

Grimmer and Stewart is not actual on the curriculum and it is also a bit dated by now (since most people use Neaural Networks; RNNs, LSTMs etc. for text operations now). However it is still a very good - and seminal - read.

Exercise

None. Well, fix you old code! Implement feedback; steal that cool code you evaluated; do all the neat things you should have done from the beginning but didn't know about. You will thank yourself at the exam.

Readings:

```
Chadefaux, T. (2014). Early warning signals for war in the news. Journal of Peace Research, 51(1):5
 - 18.

Mueller, H. F. and Rauh, C. (2016). Reading between the lines: Prediction of political violence us
ing newspaper text. American Political Science Review, 2(112):358-375.

Grimmer, Justin, and Brandon M. Stewart. "Text as data: The promise and pitfalls of automatic cont
ent analysis methods for political texts." Political analysis 21.3 (2013): 267-297. (Not strictly
on the curriculum)
```

14 Predictions about predictions: summing up the course

Week 50

Agenda:

- Potentials and pitfalls of predictions
- General questions

Exercise

None. Though, if you have not finished VanderPlas chapter 1-5 go do that. Really.

Must reads:

Two very short texts. You can manage to read both - especially since there are no official exercise in progress.

Readings:

```
Cederman, L.-E. and Weidmann, N. B. (2017). Predicting armed conflict: Time to adjust our expectat
ions?

Chadefaux, T. (2017). Conflict forecasting and its limits. Data Science, 1(1-2):7 - 17.
```