

Outline

- Rules
- Important dates
- Peer feedback and project evaluation
- Common mistakes, misconceptions and advice
- Great projects from last year
- Take-away ideas

Rules

- Topic is **free** (creativity highly encouraged!)
- No constraints on the dataset
- If in doubt, talk with us!
- We also provide some suggestions for projects

Important dates



Important dates

- 31 March - Milestone report (in groups, mandatory)
 - Project outline (1-2 pages):
 - Dataset and research question(s)
 - First draft of model(s) that you plan to try out (PGM + generative process)
 - Initial notebook (descriptive stats, data preparation)
 - Save notebook as PDF and append to project outline.
 - Submit one PDF to the peer feedback platform (we will provide more details soon).

Important dates

- 31 March - Milestone report (in groups, mandatory)
 - Project outline (1-2 pages):
 - Dataset and research question(s)
 - First draft of model(s) that you plan to try out (PGM + generative process)
 - Initial notebook (descriptive stats, data preparation)
 - Save notebook as PDF and append to project outline.
 - Submit one PDF to the peer feedback platform (we will provide more details soon).
- 13 April - Peer feedback (individually, mandatory)
- Everyone provides feedback on two milestone reports

Important dates

- 31 March - Milestone report (in groups, mandatory)
 - Project outline (1-2 pages):
 - Dataset and research question(s)
 - First draft of model(s) that you plan to try out (PGM + generative process)
 - Initial notebook (descriptive stats, data preparation)
 - Save notebook as PDF and append to project outline.
 - Submit one PDF to the peer feedback platform (we will provide more details soon).
- 13 April - Peer feedback (individually, mandatory)
- Everyone provides feedback on two milestone reports
- 22 May - Final delivery (in groups)
 - Fully self-explanatory notebook
 - 5-10 page report (incl. figures and tables)

Peer feedback and project evaluation

- Peer feedback
 - We will use peergrade.io to organise the peer feedback.
 - Each group makes one submission.
 - Everyone provides feedback on two milestone reports (individually)
 - Provide fair, respectful, caring and constructive feedback to your peers.
 - Consider the grading rubric for the final evaluation of the project when giving feedback.
 - Milestone submission and peer feedback make up 5% of the project grade.

Peer feedback and project evaluation

- Peer feedback
 - We will use peergrade.io to organise the peer feedback.
 - Each group makes one submission.
 - Everyone provides feedback on two milestone reports (individually)
 - Provide fair, respectful, caring and constructive feedback to your peers.
 - Consider the grading rubric for the final evaluation of the project when giving feedback.
 - Milestone submission and peer feedback make up 5% of the project grade.
- Project evaluation
 - We provide a detailed grading rubric on DTU Learn. Please read it carefully.

Common mistakes, misconceptions and advice

- No point putting priors on observed variables without any parents, unless doing imputation, otherwise these variables are always given

Common mistakes, misconceptions and advice

- No point putting priors on observed variables without any parents, unless doing imputation, otherwise these variables are always given
- Makes no sense to include observed variables that block the information path in the PGM
 - Consider for example: Intelligence \rightarrow Course grade \rightarrow Recommendation letter
 - If course grade is always observed, then it blocks the path between intelligence and recommendation letter. Therefore, “Intelligence \rightarrow Course grade” and “Course grade \rightarrow Recommendation letter” become two independent models.
 - There is no point estimating them jointly because there is no flow of information between the two (unless there is an alternative path, of course. . .)

Common mistakes, misconceptions and advice

- No point putting priors on observed variables without any parents, unless doing imputation, otherwise these variables are always given
- Makes no sense to include observed variables that block the information path in the PGM
 - Consider for example: Intelligence \rightarrow Course grade \rightarrow Recommendation letter
 - If course grade is always observed, then it blocks the path between intelligence and recommendation letter. Therefore, “Intelligence \rightarrow Course grade” and “Course grade \rightarrow Recommendation letter” become two independent models.
 - There is no point estimating them jointly because there is no flow of information between the two (unless there is an alternative path, of course. . .)
- Don't just think about what variables depend on other variables, and their distribution types. Think also of how you should model those dependencies (e.g. how to condition the parameters of a Beta distribution on another variable?)

Common mistakes, misconceptions and advice

- Be careful with the inclusion of latent variables that are discrete. Remember that they require special treatment in STAN/Pyro

Common mistakes, misconceptions and advice

- Be careful with the inclusion of latent variables that are discrete. Remember that they require special treatment in STAN/Pyro
- Choice of priors is often more of an art than a science. However, there are some recommendations and guidelines that you can try to follow, and people in statistics write papers about this a lot, but in the end it often boils down to a lot of trial and error. I recommend that you have a look at:
<https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>

Common mistakes, misconceptions and advice

- Be careful with the inclusion of latent variables that are discrete. Remember that they require special treatment in STAN/Pyro
- Choice of priors is often more of an art than a science. However, there are some recommendations and guidelines that you can try to follow, and people in statistics write papers about this a lot, but in the end it often boils down to a lot of trial and error. I recommend that you have a look at:
<https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>
- Try different types of inference algorithms: VI and MCMC, and make sure they are doing the right thing and you can trust the results

Common mistakes, misconceptions and advice

- Be careful with the inclusion of latent variables that are discrete. Remember that they require special treatment in STAN/Pyro
- Choice of priors is often more of an art than a science. However, there are some recommendations and guidelines that you can try to follow, and people in statistics write papers about this a lot, but in the end it often boils down to a lot of trial and error. I recommend that you have a look at:
<https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>
- Try different types of inference algorithms: VI and MCMC, and make sure they are doing the right thing and you can trust the results
- It is a generally good idea to start with a simple model and incrementally make it more complex towards the idealised/conceived PGM. Plus, try to have some baseline models for comparison whenever possible

Common mistakes, misconceptions and advice

- Be careful with the inclusion of latent variables that are discrete. Remember that they require special treatment in STAN/Pyro
- Choice of priors is often more of an art than a science. However, there are some recommendations and guidelines that you can try to follow, and people in statistics write papers about this a lot, but in the end it often boils down to a lot of trial and error. I recommend that you have a look at:
<https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>
- Try different types of inference algorithms: VI and MCMC, and make sure they are doing the right thing and you can trust the results
- It is a generally good idea to start with a simple model and incrementally make it more complex towards the idealised/conceived PGM. Plus, try to have some baseline models for comparison whenever possible
- Use ancestral sampling to generate artificial data, and run inference on the model using Pyro/STAN to see if it is able to recover the true values/parameters that were used to generate the data. This is a great way of guarantying that the model is correctly implemented, and that inference is working correctly

2019 proj.: Analysis group membership in Social Networks I



- Graph data of human interaction is called social networks. The interaction could for example be being friends on Facebook or talking to each other.
- The project inferred groups in social networks based on Game Of Thrones. Take a look at the data here <https://github.com/mathbeveridge/gameofthrones>
- Tried different models. Results below are from a Mixed membership stochastic block model - MMSBM.

- 1 For $n = 1..N$
 - 1 draw $\pi_n \sim \text{Dirichlet}(\alpha)$
- 2 For $k, k' = \{1, 2, \dots, K\} \times \{1, 2, \dots, K\}$
where $k \neq k'$
 - 1 draw $\lambda_{k,k'} \sim \text{Gamma}(a, b)$
- 3 For k in $1..K$
 - 1 draw $\lambda_{k,k} \sim \text{Gamma}(c, d)$
- 4 For all pairs $i, j \in \{1, 2, \dots, N\} \times \{1, 2, \dots, N\}$
where $i \neq j$
 - 1 $x_{i,j} \sim \text{Poisson}(\pi_i^T \lambda \pi_j)$;

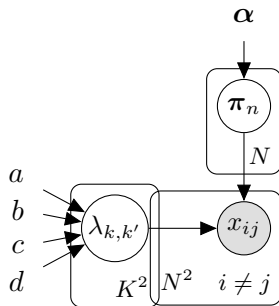


Figure: PGM for Mixed membership stochastic block model

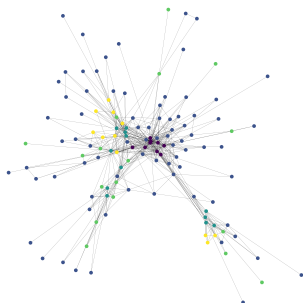


Figure: Each node is a GoT character from season 5. Colors indicate one of 5 inferred groups.

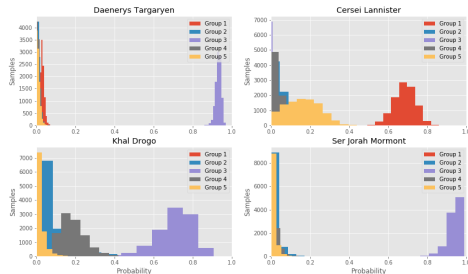


Figure: The histogram shows posterior probabilities of belonging to one of 5 groups for 4 main characters of GoT

2019 project: Bayesian semi-parametric soccer analysis

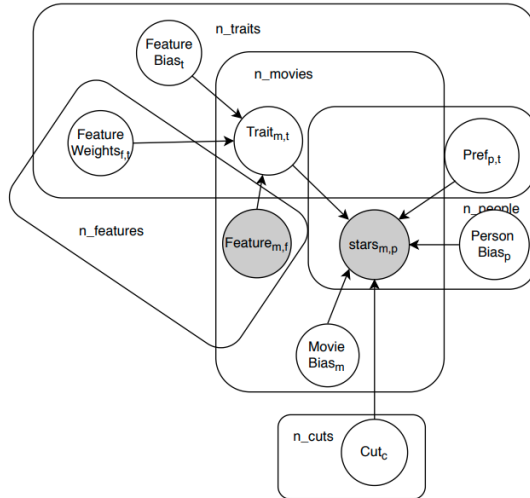
- Apply the time-to-event analysis methods to analyze the intensity with which goal scoring occurs in soccer matches
- Tried different time-to-event models from the class of Cox proportional hazards models and extend these to incorporate non-linear effects by use of the highly flexible Gaussian processes from the Bayesian non-parametrics toolbox
- The hazard in the i -th match at time interval j is

$$\lambda_{ij} = \lambda_j \exp(\beta_1 * skill_gap + \beta_2(t) * x_goal)$$

- Baseline of Cox model is a GP: $\log \lambda_0 \sim \mathcal{GP}(\mathbf{0}, K(t, t'))$
- Implemented everything in STAN and did a very good experimental evaluation

2019 project: Movie recommendations - MovieLens dataset

- Inspired by the example in Bishop's "Model-based Machine Learning" book, but with a good mix of creativity, lots of experimentation and insightful discussions



EV charging demand prediction incorporating demand diffusion

- Demand diffusion: Propagation of customers to different charging stations when the capacity of the desired station is met
- Bayesian autoregressive model with latent propagation process to predict EV charging demand
- Comparison with two other models: Without censoring or propagation variable

EV charging demand prediction incorporating demand diffusion

- 1 Draw propagation probability $p_{i,j} \sim \text{Dirichlet}(\alpha)$ with α being a proximity measure
- 2 For each charging station i :
 - 1 Draw transition coefficient $\beta_i \sim N(0, 1)$
 - 2 Draw standard deviation $\sigma_i \sim \text{HalfCauchy}(0.1)$
 - 3 Draw demand for initial time step $y_i^0 \sim N(\mu, 1)$
 - 4 Calculate censored demand $y_i^{*0} = \min(\text{Cap}_i, y_i^0)$
- 3 Draw coefficients $\tau \sim N(0, 1)$
- 4 For each charging station i :
 - 1 For each time step t :
 - 1 Draw non-censored demand:

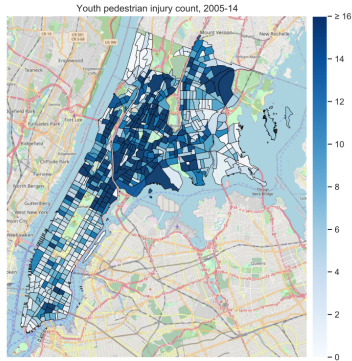
$$\mathbf{y}_i^t \sim N \left(\beta_i y_i^{t-1} + \tau X_i^t + \sum_{j \neq i} p_{i,j} (y_j^{t-1} - y_j^{*t-1}), \sigma_i^2 \right)$$
- 5 Calculate censored demand $y_i^{*t} = \min(\text{Cap}_i, y_i^t)$

Kaggle datasets



- Find a cool/crazy dataset on Kaggle - e.g. Avocado Prices (<https://www.kaggle.com/neuromusic/avocado-prices>)
- Formulate a research question(s) - e.g. how do the temporal dynamics of price change across regions
- Propose a PGM to model the data based on domain knowledge and assumptions of research question - e.g. hierarchical temporal model
- Fit model, extract results, revise model and data, propose model extensions, formulate new research questions... and iterate...

Bayesian Spatial Count Models



- For each area i , the observed count y_i follows a Negative Binomial distribution

$$y_i \sim \text{NB}(r, p_i), \quad (\text{likelihood})$$

$$p_i = \frac{\exp(\psi_i)}{1 + \psi_i}, \quad (\text{link function})$$

$$\psi = \beta^T \mathbf{X}_i + \phi_i, \quad (\text{linear function of inputs } \mathbf{X}_i)$$

where \mathbf{X}_i are the characteristics of the area and ϕ_i is a correlated error term

Bayesian Spatial Count Models

- For each area i , the observed count y_i follows a Negative Binomial distribution

$$y_i \sim \text{NB}(r, p_i), \quad (\text{likelihood})$$

$$p_i = \frac{\exp(\psi_i)}{1 + \exp(\psi_i)}, \quad (\text{link function})$$

$$\psi = \beta^T \mathbf{X}_i + \phi_i, \quad (\text{linear function of inputs } \mathbf{X}_i)$$

where \mathbf{X}_i are the characteristics of the area and ϕ_i is a correlated error term

- **Research topic 1:** How to have non-linear dependencies on inputs \mathbf{X}_i ?
 - Neural networks?
 - Hierarchical modelling?

¹Reference: <https://arxiv.org/abs/2007.03681>

Bayesian Spatial Count Models

- For each area i , the observed count y_i follows a Negative Binomial distribution

$$y_i \sim \text{NB}(r, p_i), \quad (\text{likelihood})$$

$$p_i = \frac{\exp(\psi_i)}{1 + \exp(\psi_i)}, \quad (\text{link function})$$

$$\psi = \beta^T \mathbf{X}_i + \phi_i, \quad (\text{linear function of inputs } \mathbf{X}_i)$$

where \mathbf{X}_i are the characteristics of the area and ϕ_i is a correlated error term

- **Research topic 1:** How to have non-linear dependencies on inputs \mathbf{X}_i ?
 - Neural networks?
 - Hierarchical modelling?
- **Research topic 2:** How to capture spatial correlations between areas?
 - Capture correlations through the noise term?¹

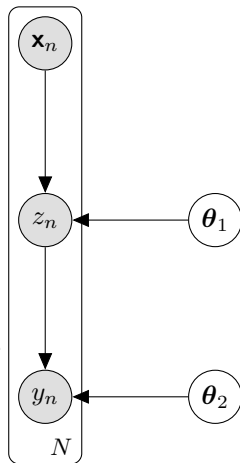
$$\exp(\tau \mathbf{W}) \phi \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$$

- Neural networks? E.g., convolutions or graph convolution neural nets?

¹Reference: <https://arxiv.org/abs/2007.03681>

Do Bayesian Neural Networks Need To Be Fully Stochastic?

- Bayesian neural network is a very active area of research
- Generative process
 - 1 Draw NN parameters $\theta \sim \mathcal{N}(\theta|\mathbf{0}, \tau\mathbf{I})$
 - 2 For the n^{th} observation
 - a Draw target $y_n \sim \mathcal{N}(y_n|f_{\text{nnet}}(\mathbf{u}_n), \sigma^2)$
- But do they need to be fully stochastic?
- **Research Ideas:**
 - ... can we only keep a fraction of θ_1 stochastic?
 - ... how do these work with time series?
 - ... analyse the Bayesian approach to neural networks.



²<https://arxiv.org/pdf/2211.06291.pdf>