DTU

DTU

Model-based Machine Learning

Introduction

Outline

Motivation

A practical example

Probabilistic programming

Structure

Team

- Teacher:

Filipe Rodrigues

Associate Professor, Department of Technology, Management and Economics
Intelligent Transportation Systems
Transportation Science
https://orcid.org/0000-0001-6979-6498

View Scopus Profile

Email
rodr@dtu.dk

Website
http://www.fprodrigues.com

Akademivej, 358, 177
2800 Kgs. Lyngby
Denmark

- TAs:
  - Danya Li
  - Emil Toft
  - Francisco Santos
  - Marianthi Kallidoni
  - Matteo Pozzar

Motivation - focus on the problem!

- Traditional machine learning
  - A very wide range of algorithms

Motivation - focus on the problem!

- Traditional machine learning
  - A very wide range of algorithms
  - When faced with a new problem, map it onto one of the existing machine learning methods

Motivation - focus on the problem!

- Traditional machine learning
  - A very wide range of algorithms
  - When faced with a new problem, map it onto one of the existing machine learning methods
  - Focus is on algorithms

Motivation - focus on the problem!

- Traditional machine learning
  - A very wide range of algorithms
  - When faced with a new problem, map it onto one of the existing machine learning methods
  - Focus is on algorithms
- Probabilistic Graphical Models
  - No need to learn about the huge range of traditional methods

Motivation - focus on the problem!

- Traditional machine learning
  - A very wide range of algorithms
  - When faced with a new problem, map it onto one of the existing machine learning methods
  - Focus is on algorithms
- Probabilistic Graphical Models
  - No need to learn about the huge range of traditional methods
  - When faced with a new problem, think about
    - the number and types of variables in the problem domain
    - which variables affect each other
    - what the effect of changing one variable is on another variable
    - encode that knowledge in the form of a model

Motivation - focus on the problem!

- Traditional machine learning
  - A very wide range of algorithms
  - When faced with a new problem, map it onto one of the existing machine learning methods
  - Focus is on algorithms
- Probabilistic Graphical Models
  - No need to learn about the huge range of traditional methods
  - When faced with a new problem, think about
    - the number and types of variables in the problem domain
    - which variables affect each other
    - what the effect of changing one variable is on another variable
    - encode that knowledge in the form of a model
  - Focus is on the problem
    - includes all the assumptions about the problem domain
    - highly tailored models for specific scenarios

MBML in a nutshell

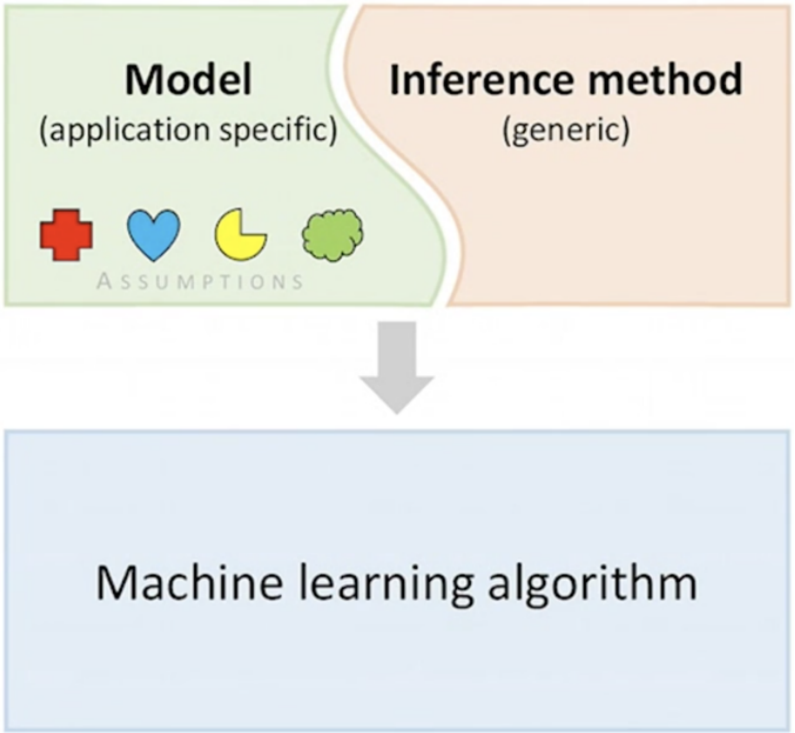


Figure: MBML big picture[1]

[1] In Christopher Bishop's Keynote, "Model-Based Machine Learning"
https://www.microsoft.com/en-us/research/video/keynote-model-based-machine-learning

Motivation - Why probabilistic models?

- **Uncertainty** in the real world
  - partial knowledge of state of the world
  - noisy observations
  - inherent stochasticity

Motivation - Why probabilistic models?

- **Uncertainty** in the real world
  - partial knowledge of state of the world
  - noisy observations
  - inherent stochasticity
- It is essential to account for uncertainty when building models of reality!
- **Probability theory** provides us with a consistent framework for quantifying and manipulating uncertainty

Motivation - A recent example...

starting to cross

crossing

stopping

bending-in

Motivation - A recent example...

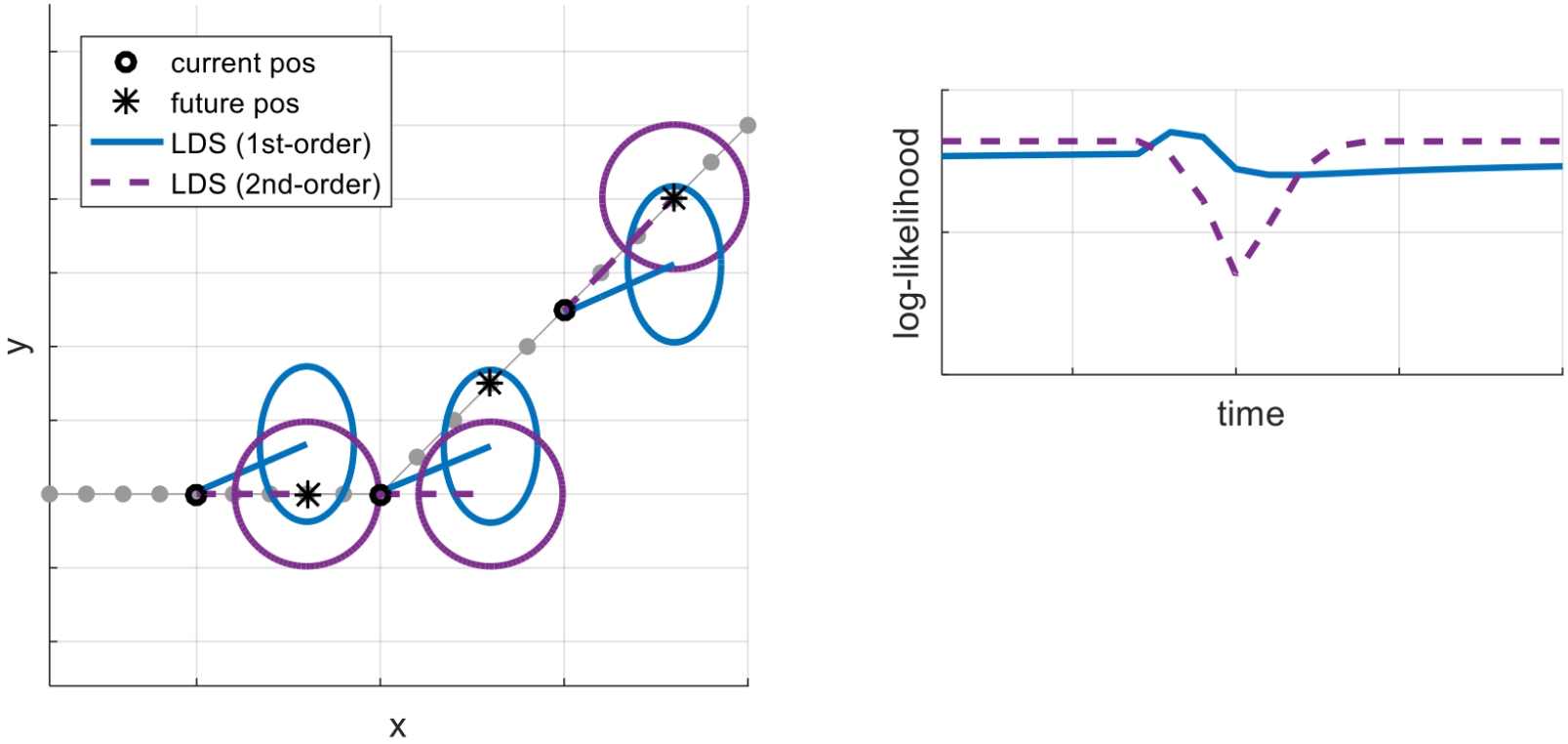
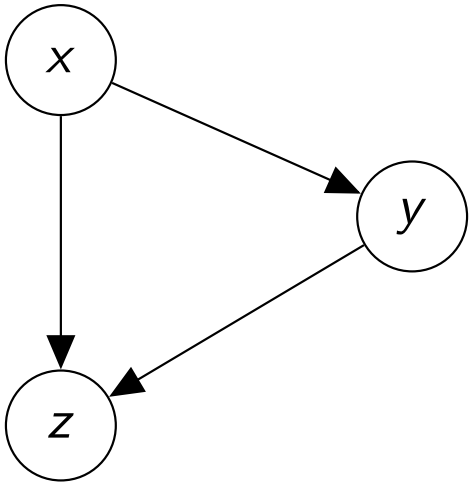


Figure: From Julian Kooij, TUDelft

Motivation - A recent example...

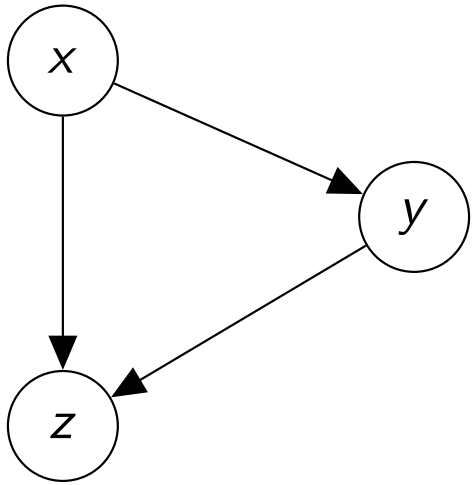
- Video

Motivation - Why probabilistic graphical models (PGMs)?



- **Nodes** - Represent random variables
- **Arrows** - Represent causal relationships

Motivation - Why probabilistic graphical models (PGMs)?



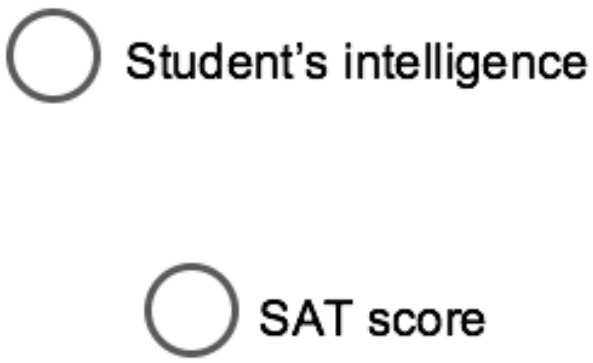
- Provide an **intuitive** and **compact** way of representing the structure of a probabilistic model
  - gives us insights about the properties of the model (e.g. relationships between variables, conditional independencies)
  - helps us communicate and design new models

A practical example

- A recruiter looking for the most intelligent students

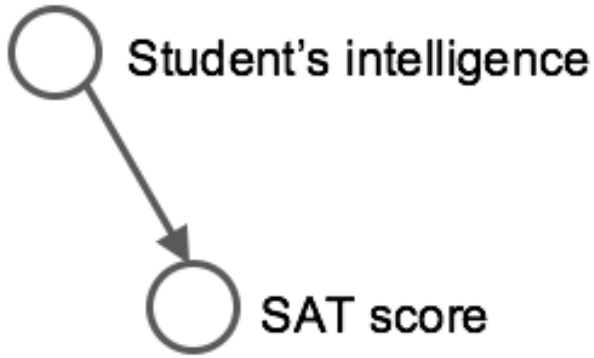
A practical example

- A recruiter looking for the most intelligent students



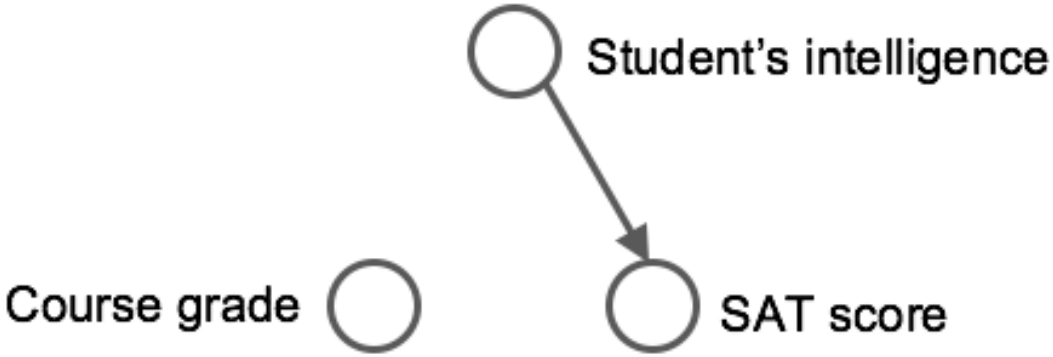
A practical example

- A recruiter looking for the most intelligent students



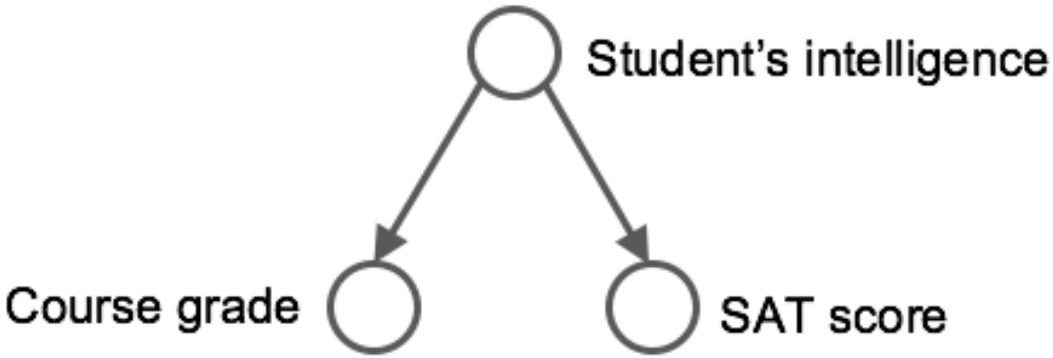
A practical example

- A recruiter looking for the most intelligent students



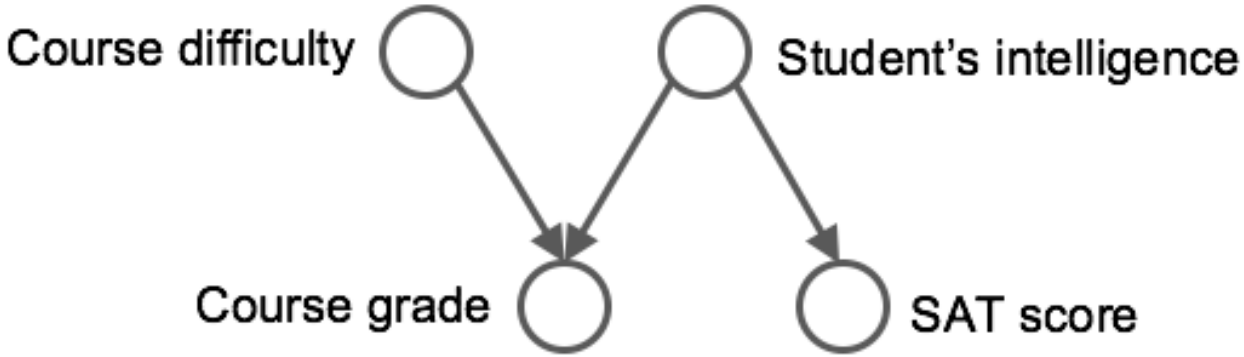
A practical example

- A recruiter looking for the most intelligent students



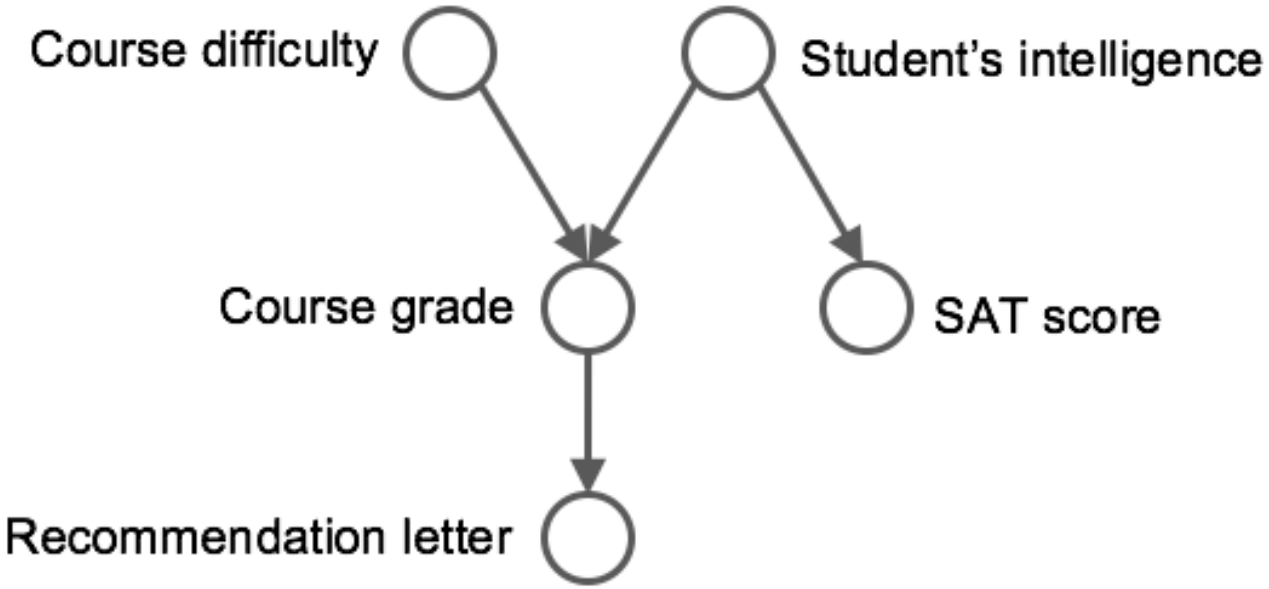
A practical example

- A recruiter looking for the most intelligent students



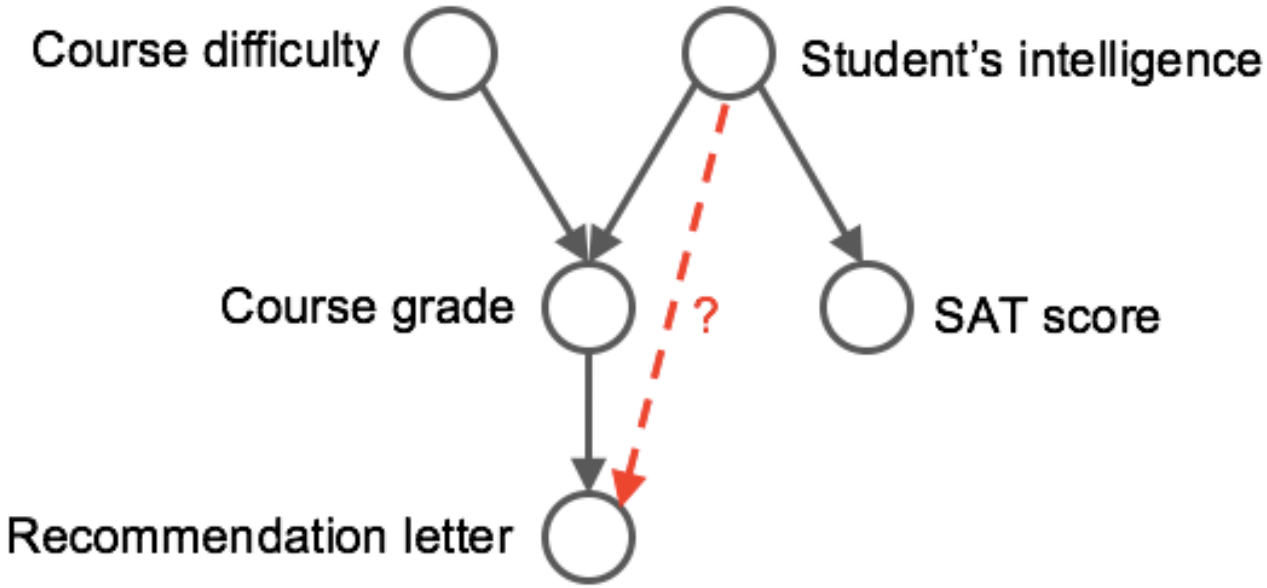
A practical example

- A recruiter looking for the most intelligent students



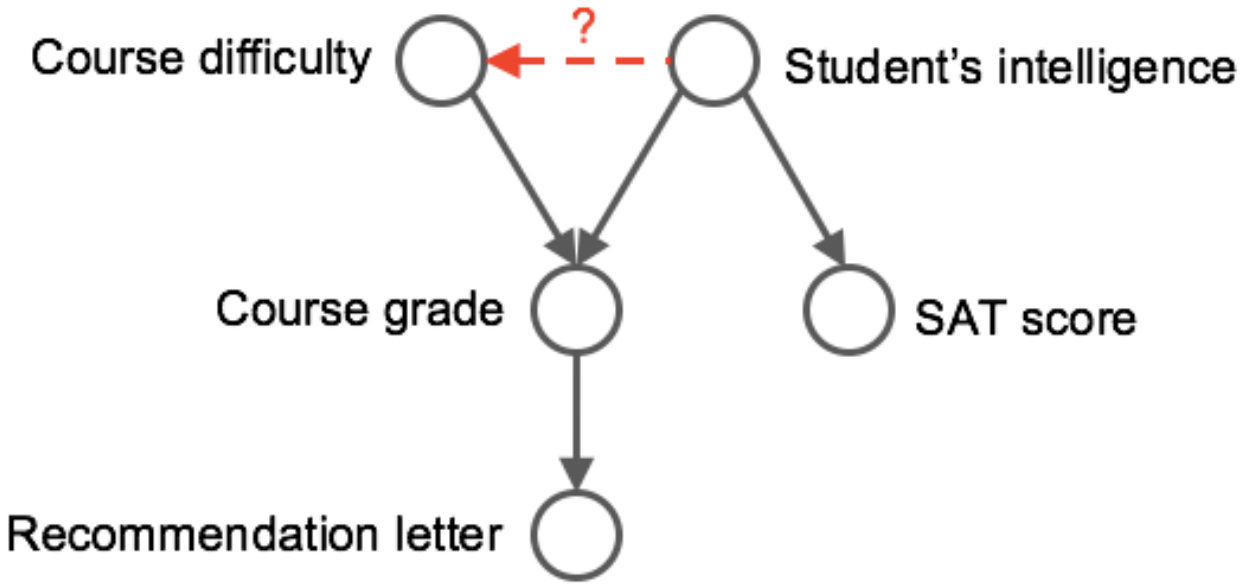
A practical example

- A recruiter looking for the most intelligent students



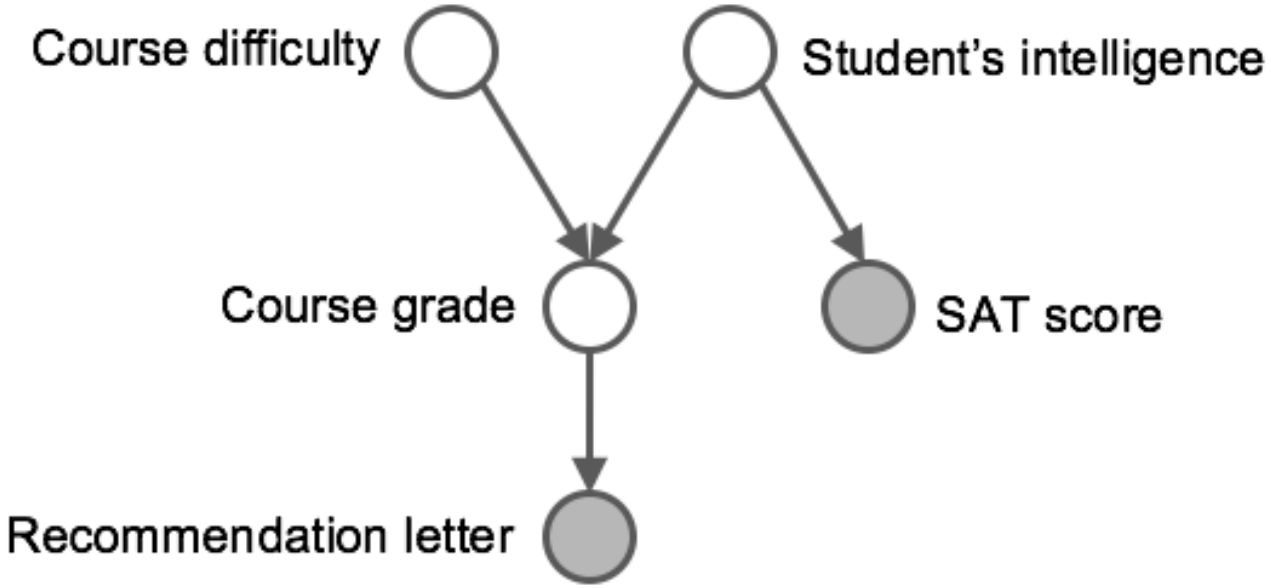
A practical example

- A recruiter looking for the most intelligent students



A practical example

- A recruiter looking for the most intelligent students



A (familiar) example - Linear regression

- y - target (a.k.a. response or dependent variable)
- $\mathbf{x}$ - vector of explanatory variables (a.k.a. predictor or independent variables)
- Target y is assumed to be a function of $\mathbf{x}$

$$y = f(\mathbf{x})$$

A (familiar) example - Linear regression

- y - target (a.k.a. response or dependent variable)
- $\mathbf{x}$ - vector of explanatory variables (a.k.a. predictor or independent variables)
- Target y is assumed to be a function of $\mathbf{x}$

$$y = f(\mathbf{x})$$

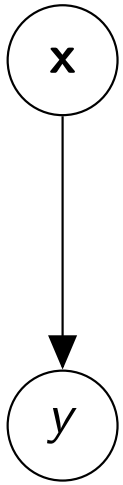
- Linear regression assumes f to be a **linear function** of $\mathbf{x}$

$$y = \beta^T \mathbf{x} + \epsilon$$

- β is a vector of coefficients
- $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is an error term (accounts for observation noise)

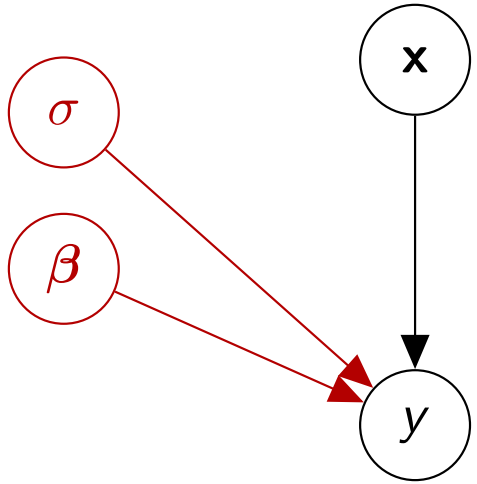
Linear regression as a graphical model

- We have a set of N observations of the targets y_n which depend on their corresponding explanatory variables $\mathbf{x}_n$



Linear regression as a graphical model

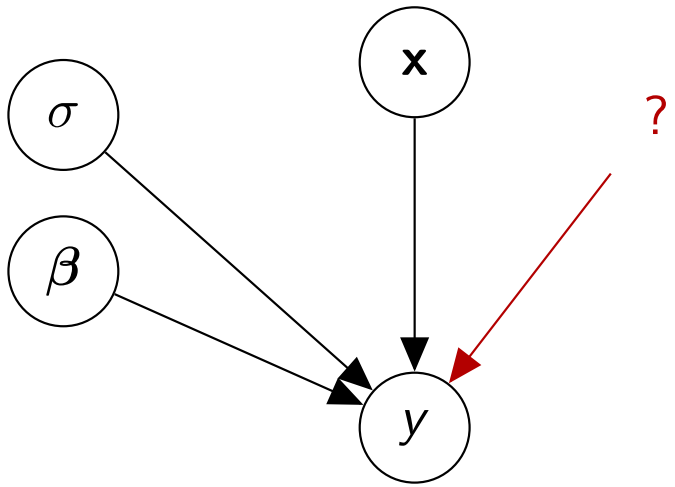
- We have a set of N observations of the targets y_n which depend on their corresponding explanatory variables $\mathbf{x}_n$



- We need to include parameters...

Linear regression as a graphical model

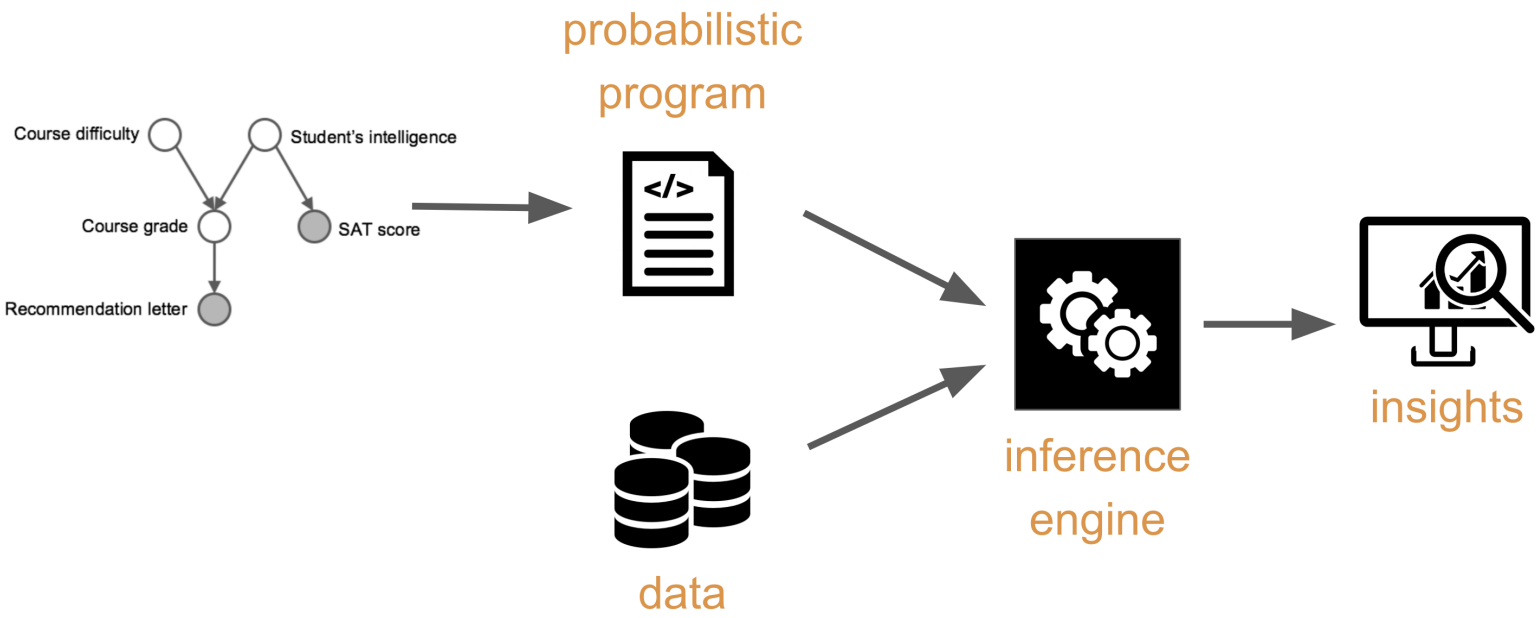
- We have a set of N observations of the targets y_n which depend on their corresponding explanatory variables $\mathbf{x}_n$



- We need to include parameters...
- In fact, we can even include other models (within our model)!

$$y = \boldsymbol{\beta_{lr}}^T \mathbf{x} + \boldsymbol{\beta_k} kernel_model + \boldsymbol{\beta_{dl}} deep_learning_model...$$

Probabilistic programming



About Pyro

- Pyro is a probabilistic programming language written in Python and supported by PyTorch on the backend
  - Pyro enables flexible and expressive deep probabilistic modeling
  - Unifies the best of modern deep learning and Bayesian modeling
  - Check `https://pyro.ai`

About Pyro

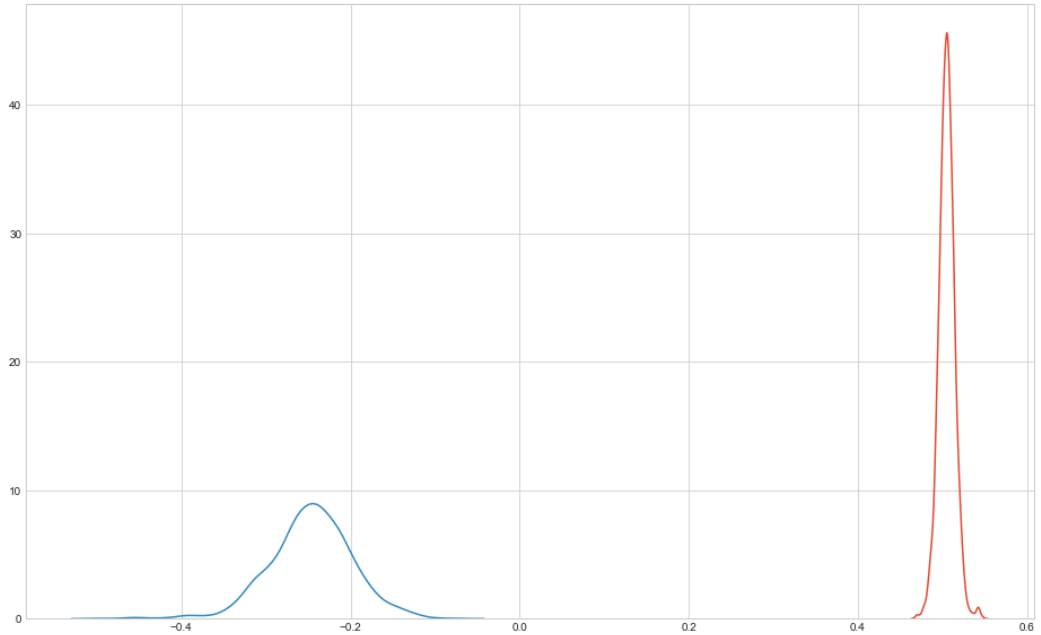
- Pyro is a probabilistic programming language written in Python and supported by PyTorch on the backend
  - Pyro enables flexible and expressive deep probabilistic modeling
  - Unifies the best of modern deep learning and Bayesian modeling
  - Check `https://pyro.ai`

- Example: Linear regression in Pyro

```
def model(x,y):
  beta0 = pyro.sample("beta0", pyro.distributions.Normal(0, 1))
  beta1 = pyro.sample("beta1", pyro.distributions.Normal(0, 1))
  sigma = pyro.sample("sigma", pyro.distributions.HalfCauchy(1))
  with pyro.plate("data", len(y)):
    pyro.sample("y", pyro.distributions.Normal(beta0+beta1*x, sigma), obs=y)
```

Example with Linear Regression

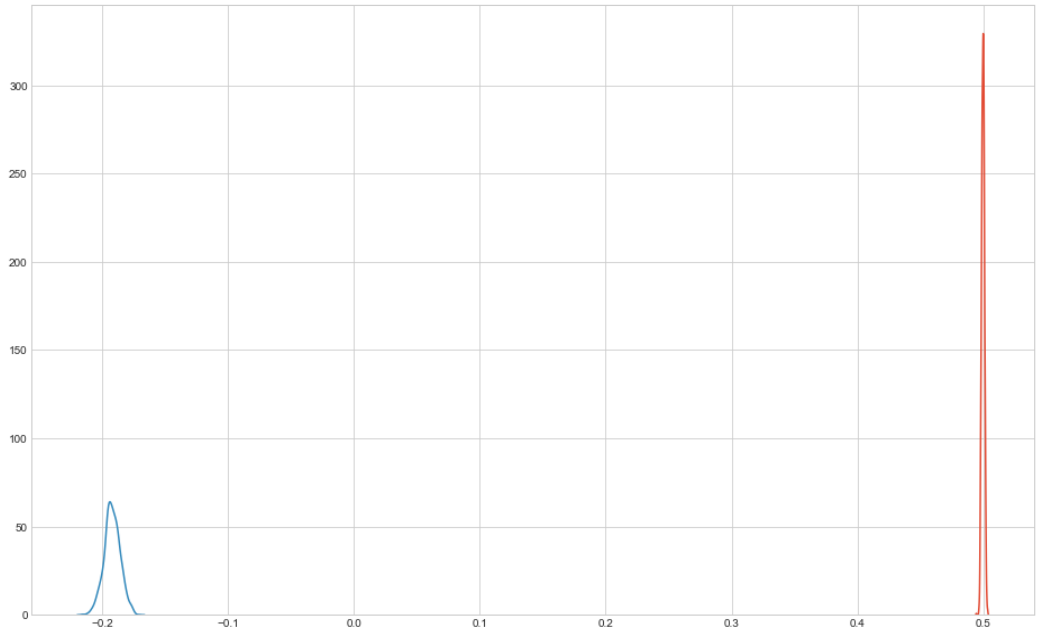
- Running inference... (N=10)



	mean
beta1	0.51
beta0	-0.25
sigma	0.1

Example with Linear Regression

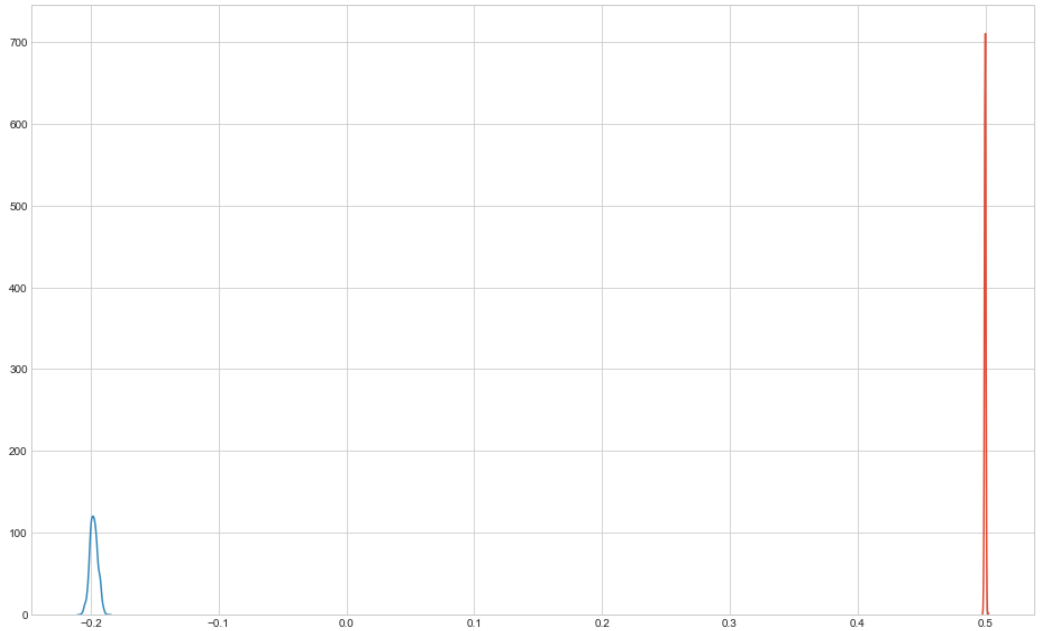
- Running inference... (N=200)



	mean
beta1	0.5
beta0	-0.19
sigma	0.1

Example with Linear Regression

- Running inference... (N=1000)

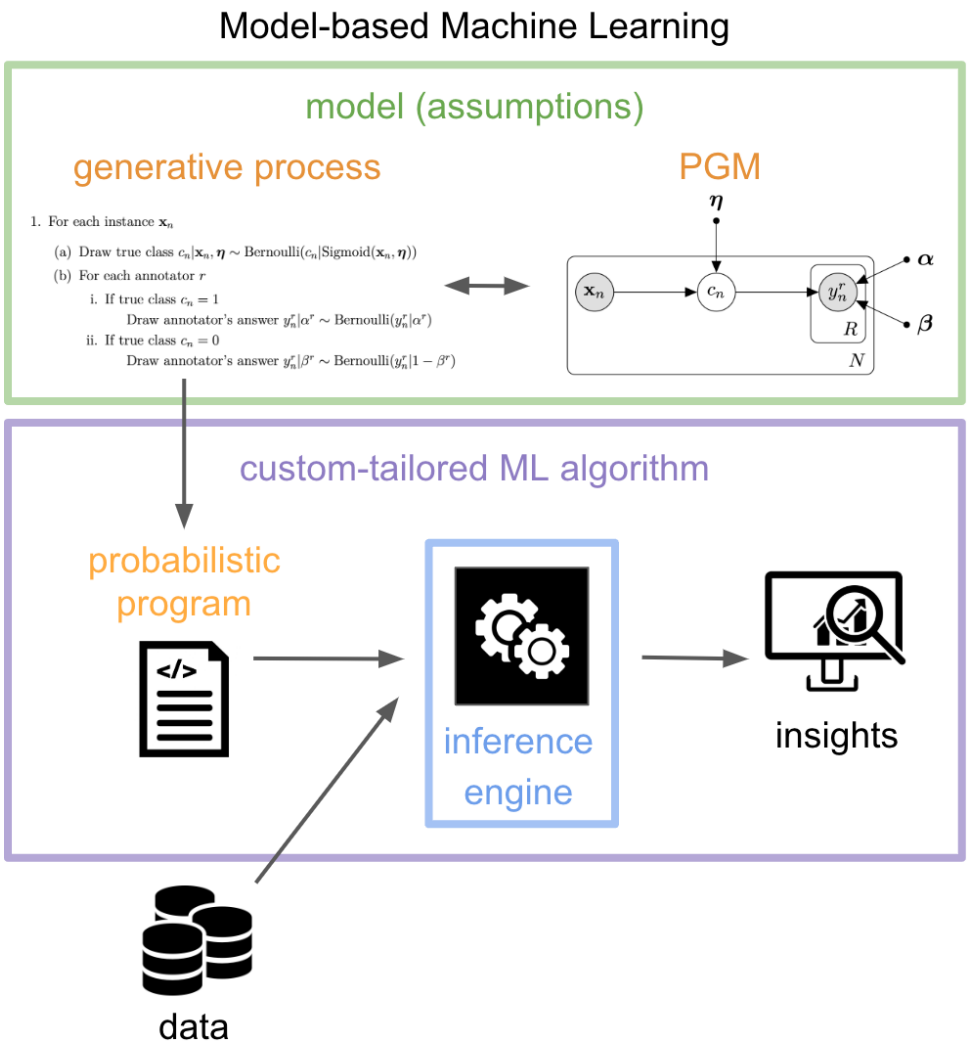


	mean
beta1	0.5
beta0	-0.2
sigma	0.1

Road to MBML

Foundations of PGMs and probabilistic programming
Probabilistic modelling in various contexts - your modelling toolbox
Bayesian inference (exact and approximate)

Week	Topic
1	Intro to the course + Prob. review
2	PGM foundations
3	PGM foundations II
4	Freq. vs Bayesian + Prob. Prog. + Mixture models
5	Regression models
6	Classification and Hierarchical models
7	Temporal models
8	Topic Models
9	Markov-chain Monte Carlo (MCMC)
10	Variational inference
11	Generative models
12	Gaussian processes
13	Project support



Method

- Interleaved blocks of theory and practice
- Python and Pyro(+PyTorch)
- Jupyter notebooks (in class)

Requirements

- Programming is important (preferably Python)
- Confident knowledge about basic statistics and probability
- Creativity, domain knowledge

Evaluation

- 2 in-class mini-tests, 25% each (dates: March 25, May 6)
  - Dates are final!
  - Multiple choice questions
  - Closed book - you can have 1 sheet of paper with you with notes
- Group project with individualized report, 50% (deadline: May 15)
  - Free topic
  - Groups of exactly 4 students (you will be randomly assigned if not)

Practical details

- Turn on notifications in DTU Learn
  (we are not responsible for missed announcements!)

Practical details

- Turn on notifications in DTU Learn
  (we are not responsible for missed announcements!)
- Some advice:
  - Don't skip class (keep the routine for you Wednesday afternoon)
  - Try to do all notebooks yourself before checking solutions
  - Don't be shy to ask for clarifications and help!