University of California, Berkeley

Modelling correlation among multiple response variables

Fabian Wauthier, Liming Wang, Paul Waddell

27^{th} May, 2010

Motivation

- Need models that can represent complex correlations among choices.
- Example: Residence location, workplace, auto ownership, building type, tenure
- Choices may be correlated by unobserved (latent) lifestyle preferences
- Multinomial logit model cannot represent these correlations due to IIA property
- Want to avoid the IIA property, while maintaining computational efficiency.

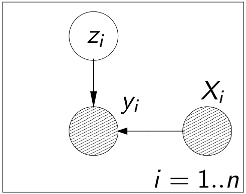
A natural modelling approach

- Allow choice dimensions to be represented with two (or more) level models
- Simplify modelling at lower levels, but allow for more complex interactions at higher level, including latent classes
- **Our interest**: modelling correlations among multiple responses

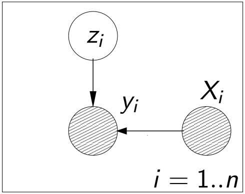
Graphical models

- Graphical models: a marriage of probability and graph theory
- A compact formalism to express complex probabilistic models

- **Notation**:
  - Nodes represent random variables
  - Un/shaded nodes were un/observed
  - Present arcs denote dependencies
  - Absent arcs condit. independencies
  - Rectangles represent replication



Kenneth Train's mixture model



- Single choice dimension y, external covariate x
- Avoids IIA property by discrete mixture over class z
- Efficient learning/prediction by EM algorithm
- Simple building block for more complex models

Inference/learning via EM algorithm

- Only some of the variables are observed.
  $\Rightarrow$ data likelihood integrates out unobserved variables.

$$\log p_\beta(y|x) = \log \int p_\beta(y|x,z)p(z)dz \tag{1}$$

- Optimising likelihood in terms of β is hard
- EM algorithm maximizes a lower bound on the likelihood

$$\log p_\beta(y|x) \geq \int q(z) \log \frac{p_\beta(y|x,z)p(z)}{q(z)} dz \tag{2}$$

  E-step: Set $q(z) = p_\beta(z|y,x)$
  M-step: $\hat{\beta} = \text{argmax}_\beta \int q(z) \log p_\beta(y|x,z)dz$

EM learning in Kenneth Train's model

- Mixing distribution $p(z)$ discrete
- Conditional likelihood

$$p(y|x,z) = \frac{\exp\{x^\top \beta_{y,z}\}}{\sum_{y'} \exp\{x^\top \beta_{y',z}\}} \tag{3}$$

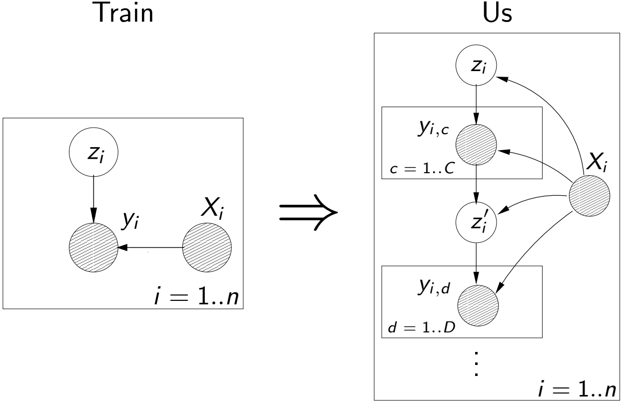
- E-step:

$$q(z) \propto p(z) p_\beta(y|x,z) \tag{4}$$

- M-step:

$$\hat{\beta} = \text{argmax}_\beta \sum_z q(z) \log p(y|x,z) \tag{5}$$

Our extension



- Multiple parallel responses of different type: Discrete, ordinal, interval, real-valued with various forms of noise.
- Conditionally mix over multiple choice dimensions at once.
- Responses can be conditioned on other responses

EM learning in extended model

- Conditional mixing distributions $p(z|x)$ still discrete
- Conditional likelihoods $p(y_c|x,z)$ given by response type
- Training: x, y are observed $\Rightarrow$ EM decouples across layers
- E-step:

$$q(z) \propto p(z|x) \prod_c^C p_\beta(y_c|x,z) \tag{6}$$

- M-step simplifies by conditional independence:

$$\hat{\beta} = \text{argmax}_\beta \sum_c^C \sum_z q(z) \log p(y_c|x,z) \tag{7}$$

$$\hat{\beta}_c = \text{argmax}_{\beta_c} \sum_z q(z) \log p(y_c|x,z) \tag{8}$$

Comparing to Train's algorithm

- E-step:

  Train: $q(z) \propto p(z) p_\beta(y|x, z)$
  Us: $q(z) \propto p(z|x) \prod_c^C p_\beta(y_c|x, z)$

- M-step:

  Train: $\hat{\beta} = \text{argmax}_\beta \sum_z q(z) \log p(y|x, z)$
  Us: $\hat{\beta}_c = \text{argmax}_{\beta_c} \sum_z q(z) \log p(y_c|x, z)$

- For $C = 1$ our model reduces to Train's.
- Otherwise: E-step not much more complicated
  M-step essentially unchanged

Results for 7 latent classes

	Predicted correlation	Actual correlation
`auto` vs `bldg`	$\begin{bmatrix} 0.01 & 0.05 \\ 0.23 & 0.20 \\ 0.31 & 0.05 \\ 0.13 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 0.01 & 0.05 \\ 0.23 & 0.20 \\ 0.32 & 0.05 \\ 0.13 & 0.01 \end{bmatrix}$
`auto` vs `tenure`	$\begin{bmatrix} 0.02 & 0.05 \\ 0.23 & 0.20 \\ 0.28 & 0.08 \\ 0.12 & 0.02 \end{bmatrix}$	$\begin{bmatrix} 0.01 & 0.06 \\ 0.22 & 0.21 \\ 0.29 & 0.07 \\ 0.13 & 0.02 \end{bmatrix}$
`bldg` vs `tenure`	$\begin{bmatrix} 0.55 & 0.14 \\ 0.10 & 0.21 \end{bmatrix}$	$\begin{bmatrix} 0.55 & 0.14 \\ 0.10 & 0.21 \end{bmatrix}$

Potential Applications

- Nested models of location choice, with sampling of alternatives
- Joint model of residence location, workplace, auto ownership, building type, tenure
- Modeling lifestyle clusters as a fuction of age, income, economic conditions (e.g. price of gas)
- Potential extensions to deal with dynamic choice situations
- Potential extension to deal with spatial correlation among alternatives
- Potential extension to jointly estimate model and calibrate against observed aggregate data