(mô tả đề bài tam giáo)

Mô hình quan hệ trong hình xxx (mô hình T2) có thể được định nghĩa lại dưới dạng mathematical form quan hệ tương quan các tham số (tô đậm) như sau:

$$logit(p) = \**alpha** + (\**beta\_{VT}** + \**beta\_{VB}** VB + \**beta\_{VC}** VC) VT + \**beta\_{AVT}** AVT$$

trong đó ký hiệu tương ứng:

C: Confucianism; T: Taoism; B: Buddhism; P: Popular; S: Secular

V of C: Confucianism; T: Taoism; B: Buddhism; P: Popular; S: Secular

AV of C: Confucianism; T: Taoism; B: Buddhism; P: Popular; S: Secular

ký hiệu “∼” có nghĩa là “is distributed as.”

Phân tích Bayesian analysis phân bổ lại độ tin cậy giữa các giá trị tham số trong tập hợp của các khả năng được xác định bởi mô hình đã chọn.

So với các bài toán hồi quy tuyến tính ở các chương trước, bài toán này có công thức trên tương đối phức tạp. Nếu mô tả dạng lưới quan hệ ta sẽ có dạng:

(hình quan hệ bài tam giáo)

* Lưới này dễ hiểu hơn nhiều

Một số lưới hồi quy:

|  |  |
| --- | --- |
| Công thức hồi quy | Lưới |
| $$y ~ bernoulli(theta\_y)$$ |  |
| $$y\_{dist} = \alpha\_{speed} + \beta\_{speed} x\_{speed}$$ |  |
| $$y\_{burden} = \alpha + \beta\_{res} \* x\_{res} + \beta\_{insured} x\_{insured}$$ |  |

* Cách tiếp cận VL xây dựng lưới quan hệ trước, không tiếp cận công thức hồi quy trước
* Công thức toán do bayesvl tự tạo
* Code stan hồi quy bayesvl tự tạo

Các lưới trong bảng trên được dựng trên bayesvl

|  |  |
| --- | --- |
| Công thức hồi quy | Code bayesvl |
| $$y ~ bernoulli(theta\_y)$$ | dag <- bvl\_addNode(dag, "y", "bern","beta(1, 1)") |
| $$y\_{dist} = \alpha\_{speed} + \beta\_{speed} x\_{speed}$$ |  |
| $$y\_{burden} = \alpha + \beta\_{res} \* x\_{res} + \beta\_{insured} x\_{insured}$$ |  |

Khái niệm node “trans” và “dummy”:

Để mô hình hoá bài toán thành các node trong lưới quan hệ, ta phân giải công thức xxx trên thành dạng đơn giản:

$$logit(p) = \**alpha** + \**beta\_{VT}**\* VT + \**beta\_{VB}** \* VB\* VT + \**beta\_{VC}** \* VC\*VT + \**beta\_{AVT}** \* AVT$$

Trên công thức này nếu so với các mô hình hồi quy tuyến tính đã học, ta nhận thấy có 1 số node “kép” các cụm: VB\* VT và VC\*VT

Như vậy phương trình này không còn là tuyến tính nữa mà đã là công thức phức hợp dạng phi tuyến. Vậy ta mô hình hoá các node này như thế nào? Cách tiếp cận bayesl cho phép chúng ta tạo ra các node trung gian “dummy” để tính trước các cụm phi tuyến dạng:

$$node\_{VB\*VT} = VB \* VT$$

$$node\_{VC\*VT} = VC \* VT$$

Như vậy công thức của chúng ta sẽ trở thành:

$$logit(p) = \**alpha** + \**beta\_{VT}**\* VT + \**beta\_{VB}** \* node\_{VB\*VT} + \**beta\_{VC}** \* node\_{VC\*VT} + \**beta\_{AVT}** \* AVT$$

Nếu vẽ lại cây quan hệ ở dạng này ta có:

Ta có 2 cách để xây dựng các node trung gian này, cách thứ nhất là tạo tính trước từ dữ liệu observations tạo ra 2 node “ảo” VB\*VT và VC\*VT rồi thực hiện bài toán. Hai ra đưa vào công thức hồi quy tính các node trung gian này trong quá trình hồi quy.

Đây là điểm khác nhau giữa node trung gian loại trans và dummy. Ta có thể nhận ra sự khác nhau khi dùng 2 loại node này trong code stan do bayesvl tạo ra dưới đây.

Mô hình bài tam giáo ở trên khi dựng bằng bayesvl sử dụng node trans

|  |
| --- |
| # Model using transformed data nodes  model <- bayesvl()  model <- bvl\_addNode(model, "T", "binorm")  model <- bvl\_addNode(model, "VB", "binorm")  model <- bvl\_addNode(model, "VC", "binorm")  model <- bvl\_addNode(model, "VT", "binorm")  model <- bvl\_addNode(model, "AVT", "binorm")  model <- bvl\_addNode(model, "Grp1", "trans")  model <- bvl\_addNode(model, "Grp2", "trans")  model <- bvl\_addArc(model, "AVT", "T", "slope")  model <- bvl\_addArc(model, "VT", "T", "slope")  model <- bvl\_addArc(model, "Grp1", "T", "slope")  model <- bvl\_addArc(model, "Grp2", "T", "slope")  model <- bvl\_addArc(model, "VB", "Grp1", "\*")  model <- bvl\_addArc(model, "VT", "Grp1", "\*")  model <- bvl\_addArc(model, "VC", "Grp2", "\*")  model <- bvl\_addArc(model, "VT", "Grp2", "\*") |

Code stan của mô hình:

|  |
| --- |
|  |

Nếu sử dụng node dummny:

|  |
| --- |
| # Model using dummy nodes  model <- bayesvl()  model <- bvl\_addNode(model, "T", "binorm")  model <- bvl\_addNode(model, "VB", "binorm")  model <- bvl\_addNode(model, "VC", "binorm")  model <- bvl\_addNode(model, "VT", "binorm")  model <- bvl\_addNode(model, "AVT", "binorm")  model <- bvl\_addNode(model, "Grp1", "dummy")  model <- bvl\_addNode(model, "Grp2", "dummy")  model <- bvl\_addArc(model, "AVT", "T", "slope")  model <- bvl\_addArc(model, "Grp2", "T", "+")  model <- bvl\_addArc(model, "VB", "Grp1", "slope")  model <- bvl\_addArc(model, "VC", "Grp1", "slope")  model <- bvl\_addArc(model, "VT", "Grp2", "\*")  model <- bvl\_addArc(model, "Grp1", "Grp2", "\*") |

Code stan của mô hình:

|  |
| --- |
|  |

Nếu so sánh 2 code này ta có thể nhận ra sự khác biệt. Các node trans sẽ được tính trong khối lệnh transformed data của stan, trong khi đó các node dummy được tính luôn trong khối lệnh parameters của mô hình.



Kết quả bài tam giáo