Roman Shaikh

Student number: 18300989

CS7DS3-Applied Statistical Modelling

Short-Assignment-2

Question 1:

Suppose y follows a Poisson distribution so that $y \sim (\theta)$ and

$$P(Y = y) = \frac{1}{y!}\theta^{y}\exp(-\theta)$$

A. General Exponential family of distribution form is given by:

$$\boxed{P(x \mid \eta) = h(x)\exp(\eta^T\, t(x) - a(\eta))}$$

Where η is called "**natural parameter**"

$t(x)$ is called "**sufficient static**"

$h(x)$ is called "**underlying measure**"

$a(\eta)$ is called "**log normalizer**"

Given, $P(Y = y) = \frac{1}{y!}\exp(-\theta)\,\theta^{y}$

$$\boxed{= \frac{1}{y!}\,\exp(y\log\theta - \theta)}$$

Therefore, on comparing above equation with standard form, we get

$$\boldsymbol{a(\eta) = \theta}$$
$$\boldsymbol{\eta^T = \log\theta}$$
$$\boldsymbol{h(x) = \frac{1}{y!}}$$
$$\boldsymbol{t(x) = y}$$

When we have enough samples / observations $y_i \quad i = 1 \dots n$, this PDF becomes as follows,

$$P\left(\frac{y}{\theta}\right) = \prod_{i=1}^{n}\frac{1}{y_i!}\exp(-\theta)\,\theta^{y_i}$$

$$= \frac{1}{\prod_{i=1}^{n} y_i!}\exp(-n\theta)\ \theta^{\sum y_i}$$

$$\boxed{= \frac{1}{\prod_{i=1}^{n} y_i!}\exp(\sum y_i log\ \theta - n\theta)}$$

Here,

$\eta(\theta) = log\ \theta$ is **Natural parameter.($\boldsymbol{\phi(\theta)}$)**

$a(\theta) = \exp(-n\theta)$ is the **link function / log normalizer**. **($\boldsymbol{g(\theta)}$)**

$h(y) = \frac{1}{\prod_{i=1}^{n} y_i!}$ Is **the normalizing constant. ($\boldsymbol{h(y)}$)**

$t(x) = \sum y_i$ is the **sufficient static. ($\boldsymbol{s(y)}$)**

B. We build a Poisson regressing model using

$$\boxed{\log(y) = \alpha + \beta X + \varepsilon}$$

Here y is a count Response variable.

We can also consider y/t (i.e. the incident / response), where t is the time interval or space or grouping.

X is the exploratory variable.

Therefore, we consider y/t as our outcome.

We require, the No. of times a customer defaults a loan - ? | Given some variables.

For this, We first group customers based on defaulter count.
We consider variables like the annual income & No. of Loans of a customer.

We can group this variable in different brackets like:

- **Annual income {**Low income (0 to 30k), High income (30 to 50k)}
- **No. of Loans** {<2, 2-5, >5}

The customers with same values of feature variable $X_1, X_2 \dots X_n$ belong to same group. i.e. defaulter or not defaulter.

E.g. 0 – No defaulter & 1 - Yes defaulter.

Total Loan	Income group	Defaulter status
<2	High	0
3-5	Low	1
>5	Low	1
>5	High	0
<2	Low	0
…	…	…

Therefor as from above example we can identify a defaulting customer based on the No. Of loans and the income group.

Effects on analysis:

Interpretation of parameter estimates

$$\log(\mu) = \alpha + \beta X$$

$$\mu = \exp(\, \alpha + \beta X)$$

$$= \exp(\alpha)\exp(\beta X)$$

i. When $\beta = 0$, we get $\exp(\alpha) = \mu$ i.e. mean of μ [expected count of y is e^{α}]. This implies X is not related to y

ii. If $\beta > 0$, then $\exp(\beta) > 1$ and expected count $E(y|x)$ is $\exp(\beta)$ times larger than when $X = 0$

iii. Similarly, if $\beta < 0$, then $\exp(\beta) < 1$ and expected count $E(y|x)$ is $\exp(\beta)$ times smaller than when $X = 0$

If the time t is not considered, then we can apply linear regression instead of Poisson.

Question 2:

A.

Latent Variable:

α = Parameter of the Dirichlet prior on the pre-document topic distribution

β = Dirichlet prior on the pre-topic word distribution.

y_i = topic distribution of document i

z_{ij} = topic of word j in document i

Parameters:

w_{ij} = specific word

Latent Dirichlet allocation (LDA):

- LDA is a generative model in statics which allows set of observed values to be explained by unobserved groups.
- LDA is an example of topic model, widely used in NLP for classifying words in a document and identifying topics.

The PDF of Dirichlet distribution is given by

$$Dir(\theta\ |\alpha\ ) = \frac{1}{\beta(\alpha)} \prod_{i=1}^{K} \theta_i^{\alpha_i - 1} \text{ , where } \beta(\alpha) = \frac{\prod_{i=1}^{k} \tau(\alpha_i)}{\tau(\sum_{i=1}^{k} \alpha_i)}$$

and $\alpha = (\alpha_1 \ldots \alpha_k)$

Graph for Latent Dirichlet allocation (LDA):

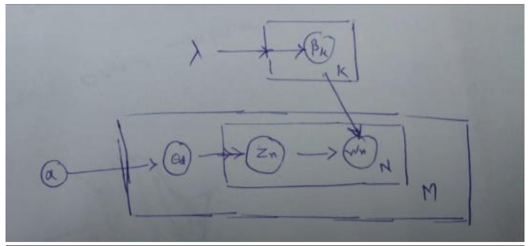
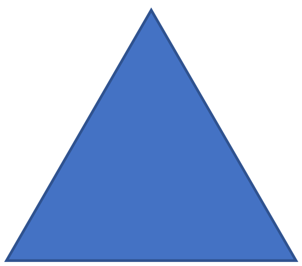


Plate notation for LDA with Dirichlet-distributed topic-word distributions

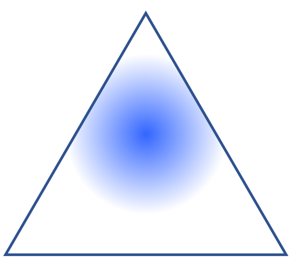
Explanation:

i. First, we draw multinomial distribution for each topic $y_i \dots y_k$
ii. Then for each document we generate a multinomial distribution $y_i \sim D\ (\alpha)$
iii. Follow this by selecting for each work a multinomial distribution parametrized by Y

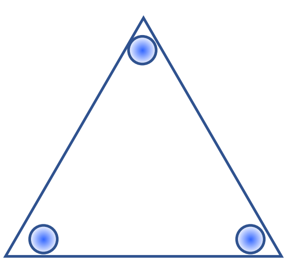
Below we find the visualization of Dirichlet distribution with varying α:



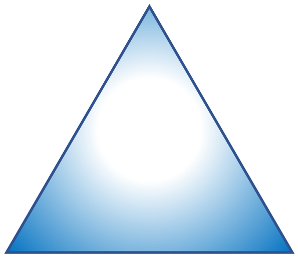
→ We suppose uniform mean across m =$\frac{1}{k}$,equal probability distribution for all multinomial distribution.



→ If α is increased to a larger value, this concentrates the probability distribution around mean i.e. in center of the simplex.



→ If mean m is different and α is adjusted, then the mean can be concentrated around three corners



→ when m = no. of topic concentrated around edges with very less probability, to belong to any of the topics.

Mixture model:

- Mixture model is a probabilistic approach for representing the normally distributed subpopulations within an overall population.
- They generally don't require knowing which subpopulation a data point belongs to , allowing the model to learn the subpopulation automatically.

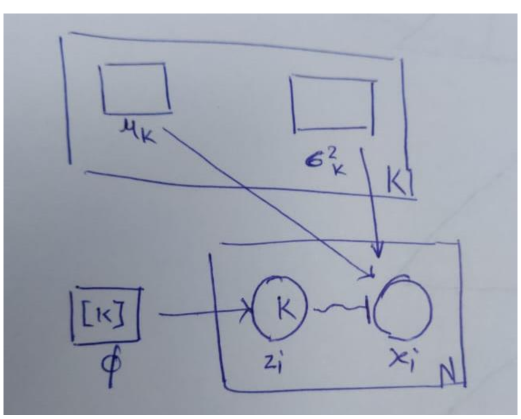
One-dimensional Model is given by:

$$p(x) = \sum_{i=1}^{K} \phi_i N(x|\mu_i, \sigma_i)$$

$$N(x|\mu_i, \sigma_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right)$$

$$\sum_{i=1}^{K} \phi_i = 1$$

Graph for Mixture Model (Non-Bayesian GMM):



Non-Bayesian Gaussian mixture model using plate notation.

Where,

K → Number of mixture components

N → number of observations

$z_i = 1 \ldots N$ → component of observation i

$x_i = 1 \ldots N$ → observation i

$\theta_i = 1 \ldots K$ → $\{\mu_i, i = 1 \ldots K; \sigma_i^2, i = 1 \ldots K\}$

$\mu_i = 1 \ldots K$ → mean component of i

$\sigma_i^2 = 1 \ldots K$ → Variance component of i

$x_i = 1 \ldots N$ → $N(\mu_{zi}, \sigma^2_{zi)})$

<u>Differences between these two models:</u>

LDA:

- LDA is not a purely mixture model but a **admixture** model.
- Has more than one latent variable.
- Used to model documents and topics
- For each topic we calculate the Multinomial distribution

Mixture Model:

- Mixture models can be either Gaussian mixture model, multivariate gaussian mixture model OR a categorical mixture model
- MMs use single latent variable
- This are modeled over number of observations.
- There are two main parts in a mixture model a kernel with parameter θ represented by $P(\alpha \mid \theta)$ and a mixing distribution $P(\theta)$