My research and Graduate life

Kalani Hasanthika

Mixture models

Mixture models

- Probabilistic models for representing the presence of subpopulations within an overall population
- Generally, does not require knowing which subpopulation a data point belongs to
- Used to make statistical inferences about the properties of the sub-populations given only observations on the pooled population
- A form of unsupervised learning

Some real-world applications

Financial modeling

Machine Learning topics

- Speech recognition systems
- Handwriting recognition
- Tracking of multiple objects in a video sequence

Breast cancer prognosis

The two most common choices of distributions

Gaussian aka "normal" (for real-valued observations)

Categorical (for discrete observations)

- Poisson
- Negative Binomial

Gaussian mixture models

- z - Latent variable (hidden variable) with K groups
- For latent variable (hidden variable) group k,

$$P(z = k) = \pi_k$$

- Where $\sum_{k=1}^{K} \pi_k = 1$ and $\pi_k \geq 0$ for $k = \{1, \dots, K\}$
- Observations y for the k^{th} latent variable group has a Gaussian distribution with mean of μ_k and variance of σ_k^2

$$P(y|z = k) = N(y \mid \mu_k, \sigma_k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu_k}{\sigma_k}\right)^2}$$

Parameter estimation

- Joint distribution:

$$P(y) = \sum_{k=1}^{K} P(z = k) \times P(y \mid z = k) = \sum_{k=1}^{K} \pi_k N(y \mid \mu_k, \sigma_k)$$

- Log-likelihood:

$$l(\pi, \mu, \sigma) = \ln P(y \mid \pi, \mu, \sigma) = \ln \sum_{i=1}^{N} P(y_i \mid \pi, \mu, \sigma)$$

$$= \sum_{i=1}^{N} \ln \sum_{k=1}^{K} P(z_i = k) \times P(y_i \mid z_i = k, \mu_k, \sigma_k)$$

$$= \sum_{i=1}^{N} \ln \sum_{k=1}^{K} \pi_k N(y_i \mid \mu_k, \sigma_k)$$

Parameter estimation

- Optimize the log-likelihood function
- Note: We have a hidden variable z_i for every observation
- Cannot use method like maximum likelihood estimation (MLE)

- Expectation Maximization (EM) algorithm, which alternates between two steps:
- E-step: Calculate the probability of generating observation y_i from the k^{th} Gaussian sub-population

$$\gamma_k^i = P(z_i = k \mid y_i, \pi, \mu, \sigma)$$

- M-step: Calculate the MLEs for the parameters by,

$$\frac{\partial \ln P(y \mid \pi, \mu, \sigma)}{\partial \theta} = 0$$

$$\widehat{\mu_k} = \frac{1}{N_k} \sum_{i=1}^{N} \gamma_k^i y_i$$

- Where $N_k = \sum_{i=1}^{N} \gamma_k^i$

$$\widehat{\sigma_k} = \frac{1}{N_k} \sum_{i=1}^{N} \gamma_k^i (y_i - \widehat{\mu_k})^2$$

$$\widehat{\pi_k} = \frac{N_k}{N}$$

Negative binomial mixture regression model

- For sample $i, \varepsilon_j = 0$,

$$(y_{ij}|\varepsilon_j = 0) \sim NegBin(y_{ij}; \lambda_{ij0}, \phi_j)$$

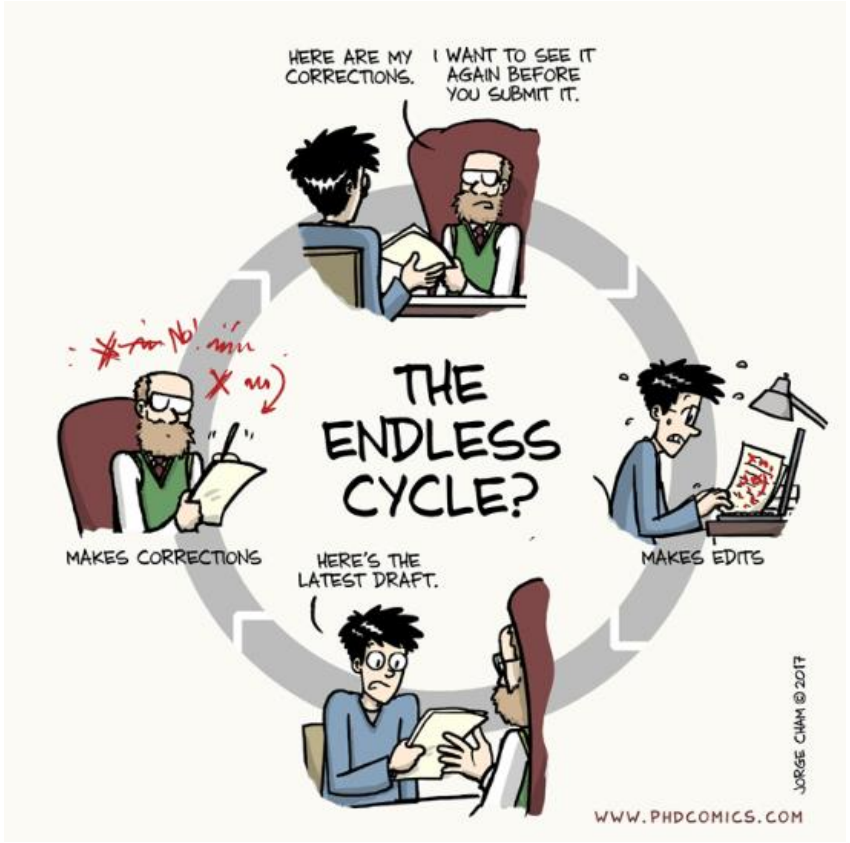
- For sample $i, \varepsilon_j = 1$,

$$(y(k)_{ij}|\varepsilon_j = 1, z_i = k) \sim NegBin(y_{ij}; \lambda_{ijk}, \phi_j)$$

- Where $i \in \{1, \dots, n\}$ and j $\in \{1, \dots, L\}$
- Where λ is the expectation and $\frac{1}{\phi}$ is the dispersion of the negative binomial distribution

ACADEMIC WORDLE

D	A	T	U	M
G	R	A	P	H
P	A	P	E	R
W	R	I	T	E
S	T	U	C	K
C	R	I	E	S



Graduate life

The format of learning may be different

Think through problems and puzzles in new ways

Produce your own research questions to tackle

Your classes will likely be heavier on reading, writing, and open discussions but lighter on the traditional lecture styles

Imposter syndrome is real - and wrong!

- Feeling of believing that they do not truly belong in their programs.
- Can be isolating
- Common among graduate students
- Others are probably feeling the same way you are and try to remind yourself that it is difficult for everyone.

Don’t be afraid to ask for help!

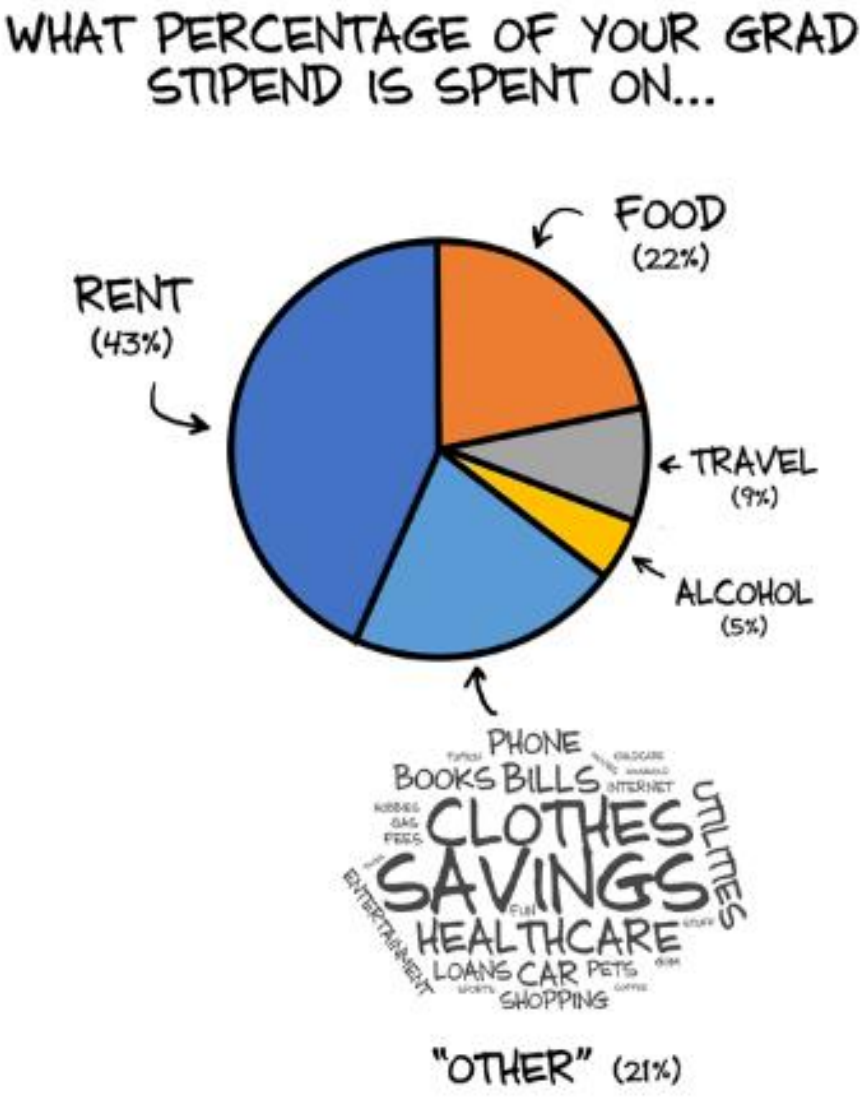
- It is okay to ask for help.
- Fear being the person that asks the “dumb” question.
- It is much better to ask for clarification than to trek on blindly and make mistakes
- Older graduate students and your cohort members are great resources
- Get comfortable

Learn to plan your time carefully

- Graduate studies are more intensive and more time-consuming
- Push yourself to develop good time management skills early on
- Set your own hours of work throughout the week
- Commit yourself to respecting those hours
- Make sure give yourself adequate breaks throughout the week.
- **Preventing burn out is important!**

Tips!

- Be aware of the financial commitment
- Get involved in research
- Consider going to therapy
- Think about the type of advisor you want to have
- Make sure you're passionate about the work
- Think about the future
- Use your practicum and internship opportunities wisely
- **Don't lose your sense of humor!**



Know that you’ve got this!

Questions

Thank you