Problem 1:

a) Maximum likelihood value of λ_A

$$f(x) = \sum_{i=1}^{N} \frac{\lambda_A^{x_i} e^{-\lambda_A}}{x_i!}$$

where x_i denotes the length of chapter i

$$\log(f(x)) = \log\left(\sum_{i=1}^{N} \frac{\lambda_A^{x_i} e^{-\lambda_A}}{x_i!}\right)$$

$$= \sum_{i=1}^{N} \log\left(\frac{\lambda_A^{x_i} e^{-\lambda_A}}{x_i!}\right)$$

$$= \sum_{i=1}^{N} \left[\log\left(\lambda_A^{x_i} e^{-\lambda_A}\right) - \log(x_i!)\right]$$

$$= \sum_{i=1}^{N} \left[\log(\lambda_A^{x_i}) + \log(e^{-\lambda_A}) - \log(x_i!)\right]$$

$$= \sum_{i=1}^{N} \left[x_i \log(\lambda_A) + -\lambda_A - \log(x_i!)\right]$$

$$\cancel{= \sum_{i=1}^{N} x_i \log \lambda_A - \log \lambda_A}$$

$$= \sum_{i=1}^{N} x_i \log \lambda_A - \sum_{i=1}^{N} \lambda_A - \sum_{i=1}^{N} \log(x_i!)$$

DATE
PAGE

$$= \log \lambda_A \sum_{i=1}^{N} x_i - N\lambda_A - \sum_{i=1}^{N} \log(x_i!)$$

Taking derivative based on λ

$$\frac{dy}{d\lambda} = \left(\log \lambda_A \sum_{i=1}^{N} x_i\right)\frac{dy}{d\lambda} - (N\lambda_A)\frac{dy}{d\lambda} - \left(\sum_{i=1}^{N} \log x_i!\right)\frac{dy}{d\lambda}$$

$$\frac{\sum_{i=1}^{N} x_i}{\lambda_A} - N - 0 = 0$$

$$\frac{1}{\lambda_A} \sum_{i=1}^{N} x_i = N$$

$$\frac{\sum_{i=1}^{N} x_i}{N} = \lambda_A$$

$$\therefore \boxed{\lambda_A = \frac{\sum_{i=1}^{N} x_i}{N}}$$

b) Generative Story

Probability of a novel being generated by by Thackeray $= \eta$

Therefore, Probability of a novel being generated by Trollope $= (1-\eta)$

Hence, p(y) can be modeled as a Bernoulli distribution

(1) $\therefore y \sim \text{Bernoulli}(\eta)$

Based on the output of the bernoulli distribution

p(x|y) can be modeled as poisson dist.

$$\therefore p(x|y=\text{Thackeray}) = \sum_{i=1}^{N} \frac{\lambda_T^{x_i} e^{-\lambda_T}}{x_i!}$$

$$p(x|y=\text{Trollope}) = \sum_{i=1}^{N} \frac{\lambda_R^{x_i} e^{-\lambda_T}}{x_i!}$$

After modeling p(y) - class priors and p(x|y) we can then use bayes rule to derive the posterior distribution on y given x

$$\text{argmax}\ p(y|x) = \underset{y}{\text{argmax}} \frac{p(x|y)p(y)}{p(x)}$$

Scanned by CamScanner

DATE
PAGE

Therefore

$$\boxed{\arg\max p(y|x) = \arg\max p(x|y)\,p(y)}$$

This will allow us to make our predictions

c) After modeling the generative story through bernoulli (priors) and poisson we will be able to cluster the issues by using Bayes theorem.

we will be able to use a classification algorith such as Logistic Regression or Naive Bayes.

(i) Suppose we model it using Naive Bayes

$$\therefore \; p(w|y) = \frac{n_w + 1}{|C_y| + |V|}$$

where $w \rightarrow$ A chapter of length w

$\therefore |C_y| \rightarrow$ No. of chapters of ~~length~~ any length conditioned on label.

$|V| \rightarrow$ Total no of chapters irrespective of class.

ii) Suppose we use Logistic Regression

$$\sigma(x) = \frac{1}{1 + \exp(-w^T x - w_0)}$$

The feature vector (x) can be
the length of chapters in a novel
or may be
the length of chapters belonging to a particular author.

- weight vector can be initialised to 0 & if it is present in the novel or if an author has a chapter of that length, a weight can be assigned & ~~delta~~ delta (δ) can be calculated.

- How we model our weight vector will depend on how we model our feature vector.

DATE
PAGE

d) Update rule

We have 3 parameters in our modeling

η - Bernoulli

λ_T, λ_R (Poisson)

$$Q(\Theta', \Theta) = \sum_{i=1}^{m} \left[\alpha_i \log\left(P(T_h, x^{(i)} | \Theta) \cdot \eta + (1-\alpha_i) \log\left(P(T_t, x^{(i)} | \Theta) \cdot (1-\eta)\right)\right)\right]$$

$$= \sum_{i=1}^{m} \left[\alpha_i \log\left[\prod_{j=1}^{n} \frac{\lambda_{Th}^{x_j} \cdot e^{-\lambda_{Th}}}{x_j!} \cdot \eta \right] + (1-\alpha_i) \log\left[\prod_{j=1}^{n} \frac{\lambda_{Tt}^{x_j} \cdot e^{-\lambda_{Tt}}}{x_j!} (1-\eta) \right] \right]$$

$$= \sum_{i=1}^{m} \alpha_i \left[\sum_{j=1}^{n} \left[\log\left(\frac{\lambda_{Th}^{x_j} \cdot e^{-\lambda_{Th}}}{x_j!} \right) + \log(\eta) \right] + (1-\alpha_i) \left[\sum_{j=1}^{n} \left[\left(\frac{\lambda_{Tt}^{x_j} \cdot e^{-\lambda_{Tt}}}{x_j!} \right) + \log(1-\eta) \right] \right] \right] \quad (1)$$

Maximising based on η i.e. $dy/d\eta$

$$\sum_{i=1}^{m} \left[\alpha_i \sum_{j=1}^{n} \left[0 + \frac{1}{\eta} \right] + (1-\alpha_i) \sum_{j=1}^{n} \left[0 + \frac{(-1)}{(1-\eta)} \right] \right] = 0$$

$$\sum_{i=1}^{m} \left[\frac{\alpha_i \times n}{\eta} - \frac{(1-\alpha_i) \times n}{(1-\eta)} \right] = 0$$

$$\sum_{i=1}^{m} \frac{\alpha_i n}{\eta} = \sum_{i=1}^{m} \frac{(1-\alpha_i) \cdot n}{(1-\eta)}$$

$$\frac{n}{\eta}\sum_{i=1}^{m}\alpha_i = \frac{n}{(1-\eta)}\sum_{i=1}^{m}(1-\alpha_i)$$

$$\frac{n}{\eta}\sum_{i=1}^{m}\alpha_i = \frac{n\cdot\left(m-\sum_{i=1}^{m}\alpha_i\right)}{(1-\eta)}$$

$$\frac{(1-\eta)}{\eta} = \frac{\cancel{n}\cdot\left(m-\sum_{i=1}^{m}\alpha_i\right)}{\cancel{n}\sum_{i=1}^{m}\alpha_i}$$

$$\frac{1}{\eta}-1 = \frac{m}{\sum_{i=1}^{m}\alpha_i}-1$$

$$\therefore \frac{1}{\eta} = \frac{m}{\sum_{i=1}^{m}\alpha_i}$$

$$\boxed{\eta = \frac{\sum_{i=1}^{m}\alpha_i}{m}}$$

DATE
PAGE

$$= \sum_{i=1}^{m} \left[\alpha_i \sum_{j=1}^{n} \left[\log\left(\frac{\lambda_{Th}^{x_j} e^{-\lambda_{Th}}}{x_j!} \right) + \log(\eta) \right] + (1-\alpha_i) \left[\log\left(\frac{\lambda_{Tl}^{x_j} e^{-\lambda_{Tl}}}{x_j!} \right) + \log(1-\eta) \right] \right]$$

— Maximising based on λ_{Th}

$$= \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{n} \left[\log(\lambda_{Th}^{x_j}) + \log(e^{-\lambda_{Th}}) - \log(x_j!) + \log(\eta) \right]$$

$$+$$

$$(1-\alpha_i) \left[\log(\lambda_{Tl}^{x_j}) + \log(e^{-\lambda_{Tl}}) - \log(x_j!) + \log(1-\eta) \right]$$

$\frac{dy}{d\lambda_{Th}}$ — Taking derivative

$$= \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{n} \left[\frac{x_j}{\lambda_{Th}} - 1 \right] = 0$$

$$\sum_{i=1}^{m} \alpha_i \left(\frac{\sum_{j=1}^{n} x_j}{\lambda_{Th}} - n \right) = 0$$

$$\frac{\sum_{i=1}^{m} \alpha_i \sum_{j=1}^{n} x_j}{\lambda_{Th}} = \sum_{i=1}^{m} \alpha_i \cdot n$$

$$\therefore \boxed{\lambda_{Th} = \frac{\sum_{i=1}^{m} \alpha_i \sum_{j=1}^{n} x_j}{\sum_{i=1}^{m} \alpha_i \cdot n}}$$

$$= \sum_{i=1}^{m}\left[\alpha_i \sum_{j=1}^{n}\left[\log\left(\frac{\lambda_{Th}^{x_j} e^{-\lambda_{Th}}}{x_j!}\right) + \log(\eta)\right] + (1-\alpha_i)\sum_{j=1}^{n}\left[\log\left(\frac{\lambda_{Tr}^{x_j} e^{-\lambda_{Tr}}}{x_j!}\right) + \log(1-\eta)\right]\right]$$

Maximising based on λ_{TR}

$$= \sum_{i=1}^{m}\Big[\alpha_i \sum_{j=1}^{n} x_j \log(\lambda_{Th}) + \log(e^{-\lambda_{Th}}) - \log(x_j!) + \log(\eta)$$

$$+$$

$$(1-\alpha_i)\sum_{j=1}^{n} x_j \log(\lambda_{Tr}) + \log(e^{-\lambda_{Tr}}) - \log(x_j!) + \log(1-\eta)\Big]$$

$\frac{dy}{d\lambda_{Tr}}$ - Taking derivative

$$\sum_{i=1}^{m}(1-\alpha_i)\left[\sum_{j=1}^{n}\frac{x_j}{\lambda_{TR}} - 1\right] = 0$$

$$= \sum_{i=1}^{m}(1-\alpha_i)\left(\frac{\sum_{j=1}^{n} x_j}{\lambda_{TR}} - n\right) = 0$$

$$= \frac{\sum_{i=1}^{m}(1-\alpha_i)\sum_{j=1}^{n} x_j}{\lambda_{TR}} = \sum_{i=1}^{m}(1-\alpha_i)\cdot n$$

$$\boxed{\lambda_{TR} = \frac{\sum_{i=1}^{m}(1-\alpha_i)\sum_{j=1}^{n} x_j}{\sum_{i=1}^{m}(1-\alpha_i)\cdot n}}$$

e) Pseudo code

$$\boldsymbol{m} \rightarrow number\ of\ novels, \boldsymbol{n} \rightarrow number\ of\ chapters, \boldsymbol{\theta} = \{\eta, \lambda_{Th}, \lambda_{Tr}\}$$

Repeat Until Convergence {

(E-Step)

$$\mathbb{Q}(\theta, \theta') = \sum_{i=1}^{m} \left[\alpha_i * \log\left(P\left(Th, X^i \middle| \theta\right) * \eta\right) + (1 - \alpha_i) * \log\left(P\left(Tr, X^i \middle| \theta\right) * (1 - \eta)\right)\right]$$

$$\mathbb{Q}(\theta, \theta') = \sum_{i=1}^{m} \left[\alpha_i * \log\left(\prod_{j=1}^{n} \left(\frac{\lambda_{Th}^{x_j} * e^{-\lambda_{Th}}}{x_j!}\right) * \eta\right) + (1 - \alpha_i) * \log\left(\prod_{j=1}^{n} \left(\frac{\lambda_{Tr}^{x_j} * e^{-\lambda_{Tr}}}{x_j!}\right) * (1 - \eta)\right)\right]$$

$$\mathbb{Q}(\theta, \theta') = \sum_{i=1}^{m} \left[\begin{array}{c} \alpha_i * \left(\sum_{j=1}^{n} \log\left(\frac{\lambda_{Th}^{x_j} * e^{-\lambda_{Th}}}{x_j!}\right) + \log(\eta)\right) \\ + \\ (1 - \alpha_i) * \log\left(\sum_{j=1}^{n} \log\left(\frac{\lambda_{Th}^{x_j} * e^{-\lambda_{Th}}}{x_j!}\right) + \log(1 - \eta)\right) \end{array}\right]$$

(M-Step)

$$\eta = \frac{\sum_{i=1}^{n} \alpha_i}{m}$$

$$\lambda_{TH} = \frac{\left(\sum_{i=1}^{m} \alpha_i * \sum_{j=1}^{n} x_j\right)}{\sum_{i=1}^{m} \alpha_i * n}$$

$$\lambda_{TR} = \frac{\left(\sum_{i=1}^{m} (1 - \alpha_i) * \sum_{j=1}^{n} x_j\right)}{\sum_{i=1}^{m} (1 - \alpha_i) * n}$$

}

Convergence condition:

EM improves log likelihood of the parameters at each iteration. The convergence condition can be as follows:

If the difference between the log likelihood of two iterations is less than some tolerance/threshold parameter, we can declare that EM is improving too slowly and hence we can stop.

Problem 2:

f) The algorithm runs for constant number of iterations i.e. 10. May be taking a threshold/tolerance parameter would have improved the accuracy. The learning rate could have been dynamic and changed after each iteration.

g) Confusion Matrix

	2.0	6.0	
2.0	43	2	**43 / 45 = 0.9555555555555556**
6.0	3	28	**28 / 31 = 0.903225806451613**

Accuracy: 71 / 76 = 0.934210526315789

h) Interpretation
There were 45 documents, which actually belonged to label 2.0. The algorithm predicted 43 of them correct.
There were 31 documents, which actually belonged to label 6.0. The algorithm predicted 28 of them correct.