CS460/626 : Natural Language Processing/Speech, NLP and the Web

Lecture 34: A very useful Maximum Likelihood Function

Pushpak Bhattacharyya
CSE Dept.,
IIT Bombay

11th Nov, 2012

Observed variable

- #Observation = N

  $\mathrm{X} : \left\langle X_1, X_2, X_3 \cdots X_N \right\rangle$

  Each x_i is a categorical distribution of k outcomes

  $\mathrm{X_i} = \left\langle \mathrm{x_{i1}}, \mathrm{x_{i2}}, \mathrm{x_{i3}}, \ldots, \mathrm{x_{ij}}, \ldots \mathrm{x_{iK}} \right\rangle$

  $x_{ij} \in \{0,1\}, \sum_{j=1}^{M} x_{ij} = 1$

Hidden variable

- Each observation from M ‘sources’, giving rise to M values form a ‘categorised distribution’ for unobserved variables.

$$\mathrm{Z_i} = < \mathrm{z_{i1}}, \mathrm{z_{i2}}, \mathrm{z_{i3}}, \ldots, \mathrm{z_{ij}}, \ldots \mathrm{z_{im}} >$$

$$z_{ij} \in \{0,1\}, \sum_{j=1}^{M} z_{ij} = 1$$

Variables: The complete picture

- Observed

$$X: X_1, X_2, X_3, \ldots, X_i, \ldots, X_N$$

- Unobserved

$$Z: Z_1, Z_2, Z_3, \cdots, Z_i, \cdots, Z_N$$

- Complete data:
  - (X,Z)

Parameters

- $\pi_{\text{i}} = \text{prob. of choosing the ith 'source'}$

$$\sum_{i=1}^{M} \pi_i = 1$$

- $\text{p}_{\text{i1}}, \text{p}_{\text{i2}}, \text{p}_{\text{i3}}, \ldots, \text{p}_{\text{ij}}, \ldots, \text{p}_{\text{iK}}$ are probabilities of the 'outcomes' from the 'jth' source

$$\sum_{l=1}^{K} p_{jl} = 1$$

Example

- Sources are 10 dice
- Each dice has 6 outcomes
- M = 10
- K = 6
- Suppose #observations = 20
  - then N = 20

Likelihood formulation

- Maximum likelihood of complete data

$$l(\theta) = P(X, Z : \theta)$$

for the i^th item

$$P(X_i, Z_i : \theta) = \prod_{j=1}^{M} \left[\pi_j \left(\prod_{k=1}^{K} P_{jk}^{\,x_{ik}} \right) \right]^{z_{ij}}$$

$$\mathrm{p}_{j1}, \mathrm{p}_{j2}, \mathrm{p}_{j3}, \ldots, \mathrm{p}_{jl}, \ldots, \mathrm{p}_{jK}$$

$$\mathrm{X}_i = \langle \mathrm{x}_{i1}, \mathrm{x}_{i2}, \mathrm{x}_{i3}, \cdots, \mathrm{x}_{ij}, \cdots, \mathrm{x}_{iK} \rangle$$

$$\mathrm{Z}_i = \langle \mathrm{z}_{i1}, \mathrm{z}_{i2}, \mathrm{z}_{i3}, \cdots, \mathrm{z}_{ij}, \cdots, \mathrm{z}_{iK} \rangle$$

Bernoulli like trial

$$P(x_i, z_i : \theta) = \prod_{j=1}^{M} \pi_j \left(p_{j1}{}^{x_{i1}} \, p_{j2}{}^{x_{i2}} \cdots p_{jk}{}^{x_{ik}} \right)^{z_{ij}}$$

<u>MLE</u>

$$l(\theta) = P(X, Z : \theta) = \prod_{i=1}^{N} \prod_{j=1}^{M} \pi_j \left(\prod_{k=1}^{K} P_{jk}{}^{x_{ik}} \right)^{z_{ij}}$$

Variable definition

- *i* goes over observation
- *j* goes over source/observation
- *k* goes over outcome/source/observation

Log likelihood of MLE

$$ll(\theta) = \sum_{i=1}^{N} \sum_{j=1}^{M} \log \pi_j + z_{ij} \sum_{k=1}^{K} x_{ik} \log p_{jk}$$

Already proved:-
We have to maximize expectation wrt z of LL(θ)
[Consequence marginalization of P(X: θ) wrt z]

Expectation of log likelihood

$$E_z\left(LL(\theta)\right)$$

Now

$$E_z\left(f(z)\right) = f(E(z)) \text{ if f is linear in z}$$

$$E_z\left(f(z)\right) \stackrel{\Delta}{=} \sum_z P(z) f(z) = f\left[\sum_z P(z) z\right]$$

Optimization of expectation wrt constraints

$$E_z\left(LL(\theta)\right) = \sum_{i=1}^{N}\left[\sum_{j=1}^{M} E(z_{ij})\left(\log \pi_j + \sum_{k=1}^{K} x_{ik} \log P_{jk}\right)\right]$$

Maximize $E_z\left(LL(\theta)\right)$ subject to the constraints

$$\sum_{j=1}^{M} \pi_j = 1 \qquad (1)$$

$$\sum_{k=1}^{K} \log P_{ik} = 1 \qquad (2)$$

$$\sum_{j=1}^{M} z_{ij} = 1 \qquad (3)$$

Introduction of Lagrange multiplier

$$\lambda_1\left(\sum_{j=1}^{M}\pi_j - 1\right) \qquad \text{(A)}$$

$$\sum_{i=1}^{N}\lambda_{2i}\left(\sum_{j=1}^{M}z_{ij} - 1\right) \qquad \text{(B)}$$

$$\sum_{i=1}^{N}\lambda_{3\mathrm{i}}\left(\sum_{k=1}^{K}p_{ik} - 1\right) \qquad \text{(C)}$$

Maximization and expectation step

<u>M - step</u>

Maximize

$E_z\left(LL(\theta)\right) + (\text{A}) + (\text{B}) + (\text{C})$ with respect to π_j, p_{jk}

$j = 1,2\cdots\cdots \text{M}$

$\text{k} = 1,2\cdots\cdots \text{K}$

<u>E - step</u>

$$E(z_{ij}) = \frac{\pi_j \left(\prod_{k=1}^{K} P_{jk}^{\ x_{ik}} \right)^{z_{ij}}}{\sum_{j=1}^{M} \pi_j \left(\prod_{k=1}^{K} P_{jk}^{\ x_{ik}} \right)^{z_{ij}}}$$

Two coin example

$$P(z = z_i \mid x = x_i)$$

$$E(z_i) = \frac{P \times P_1^{x_i}(1 - P_1)^{(1-x_i)}}{P \times P_1^{x_i}(1 - P_1)^{(1-x_i)} + (1 - P) \times P_2^{x_i}(1 - P_2)^{(1-x_i)}}$$