Lecture 6/20 : - Mixture of Binomial Dist → HW3 PB (Cons)

- Mixt of Poissons Dist

Next

Wed 6/25

Demo HW 3

- Other Gen Models

- Sampling techniques

Binomial Dist $P[X=k] = \binom{n}{k} p^k (1-p)^{n-k}$

$\rightarrow$ $\{Pr[tail]\}^{\#tails}$

probab to see k heads in n trials.

choice of k positions in n

$\{Pr[head]\}^{\#heads}$

p = coin fixed bias = Pr[Head]

n = # trials = # flips

X = R.v = # heads observed

Binom Th $\displaystyle\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1 \Rightarrow$ valid distribution over possible outcomes $\{X=0, X=1, X=2, \ldots X=n\}$

Fit dist to data

x_i = # heads $\longrightarrow$ $p \simeq \dfrac{x_i}{n}$ chance of head

Multiple Coins $k=1$ $k=2$ . . . $\boxed{k=K}$ 6 (# coins not heads)

p_k = bias for coin k = pr[head | coin k]

Data $x_1, x_2, \dots x_M$ # of heads in n trials ; M experiments each with a fixed unknown coin.

$M=20$ experiments. same coin 100

Mixture $P(x_i) = \sum_{k=1}^{K} w_k \cdot Bin(x_i \mid n, p_k)$

E STEP

Membership π_{ik} = prob that coin k generated x_i = $p(\text{coin } k \mid x_i)$

$= \dfrac{pr[x_i \mid \text{coin } k] \cdot pr[\text{coin } k]}{\text{normalized over all } k \text{ numerators}} = \dfrac{Bin(x_i \mid p_k, n) \cdot w_k}{\text{normalized}}$

w_k fixed

$= \dfrac{w_k \cdot p_k^{x_i} (1-p_k)^{n-x_i} \cancel{\binom{M}{x_i}}}{\text{normalized} \ \cancel{\binom{M}{x_i}}}$

$\prod_{i=1}^{M} \left[\prod_{k=1}^{K} \cdot \underbrace{\left[pr\binom{\text{coin}}{\text{select}} \cdot p(x_i \mid \text{coin } k) \right]}_{pr(\text{coin } k,\ x_i \text{ heads})}^{\pi_{ik}} \right]$ $\pi_{ik} \to 1$, $\to 0$

for $i=1 : M = 20$ experiments

- pick coin $k \in \{1, 2, \dots K\}$ prob w_k
- flip coin $n = 100$ times
- record x_i = # heads.

M step

$$\text{log likelihood} = \log\left[\prod_{i=1}^{M}\prod_{k=1}^{K}\left[w_k \underbrace{\operatorname{Bin}(x_i \mid p_k, n)}_{\binom{n}{x_i} p_k^{x_i} (1-p_k)^{n-x_i}}\right]^{\pi_{ik}}\right]$$

filter ↗ 1
R.V ↘ 0

n = # Flips for each experiment $i = 1, 2, \ldots M$

avg (π_{ik})

expected over π_{ik}

$$\mathrm{ELL} = \sum_{i=1}^{M}\sum_{k=1}^{K}\langle \pi_{ik}\rangle\left[\underbrace{\log w_k}_{\text{ignore}} + \underbrace{\boxed{\log\binom{n}{x_i}}}_{\text{ignore}} + \log p_k^{x_i} + \log(1-p_k)^{n-x_i}\right]$$

$E[\pi_{ik}]$ prob

$$\frac{\partial \mathrm{ELL}}{\partial p_k} = \frac{\partial}{\partial p_k}\sum_{i=1}^{M}\langle \pi_{ik}\rangle\left[x_i \log p_k + (n-x_i)\log(1-p_k)\right]$$

$$= \sum_{i} \pi_{ik} x_i \cdot \frac{1}{p_k} - \sum_{i=1}^{M} \pi_{ik}\frac{n-x_i}{1-p_k}$$

$$= S_k \cdot \frac{1}{p_k} - nR_k\frac{1}{1-p_k} + S_k\frac{1}{1-p_k} \overset{\text{want}}{=} 0$$

$$S_k = \sum_i \pi_{ik} x_i$$

$$R_k = \sum_i \pi_{ik}$$

$\Big|\ \cdot\ p_k \cdot (1-p_k)$

$$\underline{S_k(1-p_k)} - \boxed{nR_k \cdot p_k} + \underline{S_k \cdot p_k} \overset{?}{=} 0$$

$$S_k \overset{?}{=} \quad nR_k \cdot p_k$$

$$p_k = \frac{S_k}{n \cdot R_k} = \frac{\sum_i \langle \pi_{ik} \rangle \cdot x_i}{\sum_i \langle \pi_{ik} \rangle} = \text{weighted AVG } \{x_i\} \text{ weights } \langle \pi_{ik} \rangle$$

For w_k: Lagrangian $= \sum_i \sum_k \pi_{ik} \log(w_k)$ subj to $\sum_k w_k = 1$

identical to mixt gauss

$$\Rightarrow w_k = \frac{1}{N} \sum_{i=1}^{M} \pi_{ik}$$

Poisson Distribution # events in fixed interval

- events independent
- average occurrance rate $= \lambda$

$$\frac{\text{Poisson}}{\text{dist}} = \text{limit of } \frac{\text{Binomial}}{\text{dist}} \; Bin(\underset{\uparrow \text{ success prob}}{p}, \underset{\uparrow \text{ \# trials}}{n}) \Rightarrow \begin{array}{l} \text{avg rate} \quad \text{binomial} \\ \lambda = np = E[X] \\ p = \frac{\lambda}{n} \end{array}$$

$$P[X = \#\text{events}] = \lim_{n \to \infty} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \lim_{n \to \infty} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$\boxed{\begin{array}{l} \left(1+\frac{1}{n}\right)^n \to e \\ \left(1-\frac{1}{n}\right)^n \to \frac{1}{e} \end{array}}$$

$$= \lim_{n} \frac{n!}{x!(n-x)!} \cdot \frac{\lambda^x}{n^x} \cdot \left(1 - \frac{\lambda}{n}\right)^{n \cdot \frac{n-x}{n}} \qquad e^{-\lambda}$$

$$= \boxed{\lim_{n}} \; \underset{1}{\boxed{\frac{n!}{(n-x)!\, n^x}}} \cdot \frac{\lambda^x}{x!} \cdot \boxed{\left[\left(1 - \frac{\lambda}{n}\right)^n\right]}^{\boxed{\frac{n-x}{n}} \Rightarrow 1} = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$\underset{pr}{\text{Poisson}}(x) \simeq \frac{\lambda^x}{x!} \qquad \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \underset{\text{Series}}{\overset{\text{Taylor}}{=\!=}} e^{\lambda}$$

$$pr(x) = \frac{\lambda^x / x!}{\text{normalized over all } x \in \{0, 1, 2, \ldots \infty\}} = \frac{\lambda^x / x!}{\sum \text{numerators}} = \frac{\lambda^x / x!}{e^{\lambda}}$$

Mixture of Poisson distributions : K coins flipped ∞ times.

$$P(x_i) = \sum_{k=1}^{K} w_k \frac{\lambda_k^{x_i}}{x_i!} e^{-\lambda_k}$$

$$\sum_{k=1}^{K} w_k = 1$$

prob of x_i loads in fixed interval with avg λ_k

fixed selection probab.

E step π_{ik} = membership prob $\left(\begin{smallmatrix}k\\ \text{coin}\end{smallmatrix} \middle| x_i\right) = \dfrac{pr(k\text{ coin}) \cdot p(x_i | \text{coin } k)}{\text{normalized}}$

$$= \frac{w_k \cdot \lambda_k^{x_i} / x_i! \cdot e^{-\lambda_k}}{\text{normalized over coins } k = 1, 2, \ldots K}$$

M step w_k = same as before

$$= \frac{1}{m} \cdot \sum_{i=1}^{M} \pi_{ik}$$

Prior prob of selecting K coin

M step for λ_k

$$P(x_i) = \sum_{k=1}^{K} w_k \frac{\lambda_k^{x_i}}{x_i!} e^{-\lambda_k}$$

$$E_{[\pi_{ik}]} \text{LogLikelihood} = \sum_{i=1}^{M} \sum_{k=1}^{K} \underbrace{\boxed{\langle \pi_{ik} \rangle}}_{E[\text{filler}]} \left[\log w_k + x_i \log \lambda_k - \lambda_k - \log(x_i!) \right]$$

for λ_k: maximize $\sum_{i=1}^{M} \pi_{ik} [x_i \log \lambda_k - \lambda_k]$

$$\frac{\partial}{\partial \lambda_k} = \sum_{i=1}^{M} \langle \pi_{ik} \rangle x_i \cdot \frac{1}{\lambda_k} - \sum_{i=1}^{M} \langle \pi_{ik} \rangle \overset{\text{want}}{=} 0$$

$$\lambda_k = \frac{\sum \langle \pi_{ik} \rangle x_i}{\sum \langle \pi_{ik} \rangle}$$

weighted AVG (x_i) $\frac{\text{\# events}}{\text{hour}}$

$= \overset{W}{\text{AVG}}\left(\frac{\text{\# events per}}{\text{hour}}\right)$

Gen. Models Overview

- Gauss Discriminant Analysis ✓
- Naive Bayes ✓
- GMM (EM of mixt_gaussians) ✓
  - EM for Mixt_Binom
    Mixt_Poisson
- Hidden Markov Models → next time

Markov Process
sequence

hidden states — unknown: $Z_1 \; Z_2 \; \dots \; Z_T$

observation: $X_1 \; X_2 \; \dots \; X_T$

- stationary state dist $[\pi_1 \; \pi_2 \; \dots \; \pi_k]$
- transition probab $P_{ij} = pr[Z_i \to Z_j]$
- emission prob $Z_i \to X_j$

• Latent Dirichlet Allocation "Topic Models" for text.

doc

d = dist over topics

K topic dists.

Φ_1

Φ_2

Φ_K

topic

Φ = dist over words

• SAMPLING (unsup ML)

basics

(next)

• (**) VARIATIONAL ENCODER

VARIATIONAL: var inference, alternative to EM

ENCODER: NN encoder