Welcome to Day 17

Agenda: Q&A on $\boxed{\text{HW3}}$ due 4/1
Q&A on $\boxed{\text{CP3}}$ due Sun 4/5 (late deadline)
Q&A on Markov models

3 ways to ask:
(1) use chat on Zoom
(2) audio!
(3) Piazza live Q+A

HW3

What is entropy for vector r.v.?

$$\mathbb{E}_{q(z)}\left[-\log q(z)\right]$$ $z \in \text{onehot}(K)$

PMF

high: uncertainty

low: certain

0.0: total certainty

$$\sum_{z \in \Omega} q(z)\left[-\log q(z)\right]$$

is this entropy result scalar or vector?

PMF produce scalars!

$q(z) = q(z|\Gamma)$ Yes for HW3

$$q(z_k = 1 \mid r) = r_k$$

HW3 2c

Hint: Use Jensen's

$$\boxed{\log \mathbb{E}_{q(z)}[f(z)] \geq \mathbb{E}_{q(z)}[\log f(z)]}$$

f : any function produces scalar

q : any valid distribution over z

no explicit relation btwn f, q

$$\log \mathbb{E}[z^T z] \geq \mathbb{E}[\log z^T z] \qquad f(z) = z^T z$$

$$\log \mathbb{E}[q(z)^2] \geq \mathbb{E}[\log q(z)^2] \qquad f(z) = q(z)^2$$

Entropy:

$$\lim_{r \to 0} r \log r = 0.0$$

(prove via L'Hopital rule)

For Problem 2: which vector $r \in \Delta^K$ is 'max' entropy

looking for reasonable justification

NOT formal proofs

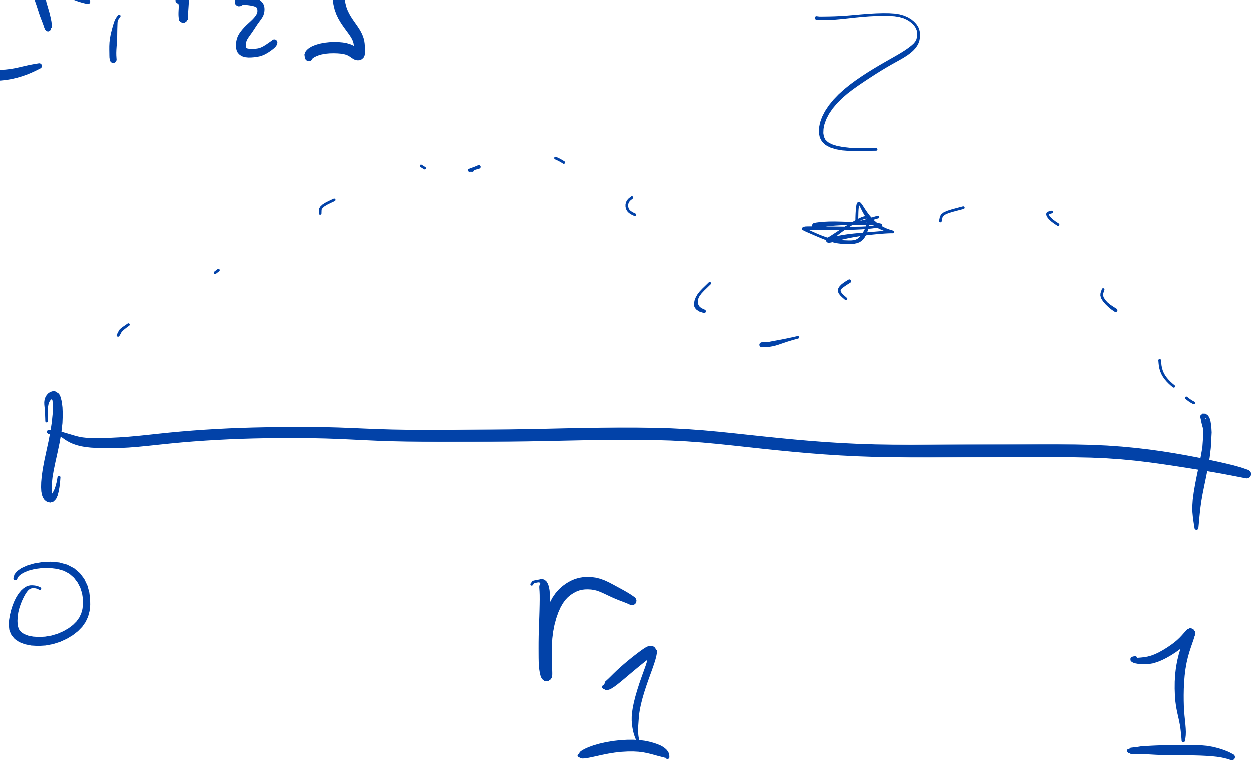
why would $\left[\frac{1}{2} \; \frac{1}{2}\right]$ be max entropy?

$K=2$

1 way: High entropy uncertain about outcomes

Make a plot K=2 case

$\vec{r} = [r_1, r_2]$



Any other HW3 ?s

Now: CP3 questions

penalty function is given

TODOs are all explicit in the code

#TODO must be in in function

---

assert in starter code commented out, but you should put them back in & debug that until they pass

EM loss has 3 terms
- negative E[log complete likelihood]
- negative entropy
- penalty term

EM loss

$$-\mathbb{E}_q\left[\log p(x,z)\right] - \left(\mathbb{E}_q\left[-\log q(z)\right]\right) + \text{penalty}(\pi)$$

neg. complete lik   neg entropy   penalty

See Day 16 lecture notes for expressions of

Why entropy term in EM loss?

(1) Bound on incomplete likelihood

$$\log p(x) \geq \mathbb{E}_q\left[\log p(x,z) - \log q(z)\right]$$

makes bound tighter

(2) Practical:
Be uncertain about latent z except where that hurts likelihood

Questions abt Markov Models

Day 17

- reviewed <u>dependencies</u> in time series
- Markov assumption

$$p(z_{t+1} \mid z_t, z_{t-1}, \ldots z_1)$$
$$= p(z_{t+1} \mid \underline{z_t})$$

Assume $z_t \in \{1, \ldots K\}$

- Markov Model
  PMF ~~that~~ generates $z_1, z_2 \ldots z_T$
- Hidden Markov Model
  $p(x_{1:T}, z_{1:T})$

HMM w/ Gaussian likelihoods

$$p(z_{1:T})\,\underbrace{p(x_{1:T} \mid z_{1:T})}$$

$x_t \sim$ iid given z_t

so if $z_t = k$

$$\boxed{p(x_t \mid z_t = k)} = N(\mu_k, \sigma_k)$$

HMM w/ Bern distrib.

$x_1, \ldots x_T$

$0, 0, 0, 1, 1, 0$

prob "Heads" under state k

$$p(x_t \mid z_t = k) = \text{Bern}(p_k)$$

Wed: Know $p(x_{1:T}, z_{1:T})$
given π, A, μ, σ

- compute $p(x_{1:T})$
  marginal likelihood
- estimate params
  $\underline{\pi, A, \mu, \sigma} = \max \log p(x_{1:T})$

EM!

K = # states
hyperparameter

GMM
G-HMM
Ber-HMM

CV, heldout validation set, "evidence"

Special Case:

HMM with initial prob π

trans prob $A = \begin{bmatrix} - & \pi & - \\ - & \pi & - \\ & \vdots & \\ - & \pi & - \end{bmatrix}$

with $p(x|z)$ params μ, Σ

equivalent (exactly) to GMM w/

weight π

parameter μ, Σ $p(x|z)$

Why? $\boxed{z_{1:3} = \{z_1, z_2, z_3\}}$ one $T = 3$

$$p^{HMM}(z_{1:T}) = \mathrm{Cat}(z_1 | \pi) \cdot \mathrm{Cat}(z_2 | \cancel{A}^{\pi}_{z_1}) \cdot \mathrm{Cat}(z_3 | \cancel{A}^{\pi}_{z_2})$$

several

$\boxed{\text{all rows of } A = \pi}$

exactly $p(z)$ for GMM

$$p(z_{1:T}) = p(z_1, z_2, \dots z_T)$$

compact

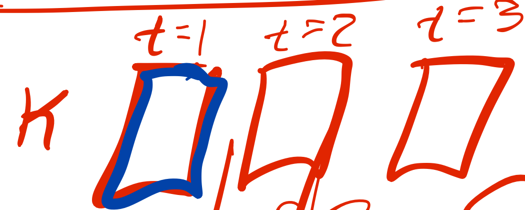
$[z_1, z_2, z_3] \dots z_{T-1}, z_T$

$$p(z_t \mid \underline{z_{1:t-1}})$$

$$p(z_{t+1} \mid \underline{z_{1:t-1}})$$

$z_1 \to z_2 \to z_3$

Markov dependent!



E step update $q(z)$ K^2

HMM: $q(z) = q(z_1)\prod_t q(z_t \mid z_{t-1})$

max $q(z \mid x)$

$$E_q[\log p(x, z) - \log q(z)]$$

M step update parameter

π, μ, σ

THANKS!

that's all folks