Ling 5801: Lecture Notes 14
Message Passing for Probability Models

Contents

14.1 Efficient inference by ‘message passing’

Most queries don’t need to calculate the full joint distribution (through 8,000,000 iterations):

$$\begin{aligned} \mathsf{P}_{\theta_{Sp}}(b) &= \sum_{p,v,f_0,f_1,f_2} \mathsf{P}_{\theta_{Sp}}(p,v,b,f_0,f_1,f_2) \\ &\stackrel{\text{def}}{=} \sum_{p,v,f_0,f_1,f_2} \mathsf{P}_{\theta_P}(p) \cdot \mathsf{P}_{\theta_V}(v \mid p) \cdot \mathsf{P}_{\theta_B}(b \mid p) \cdot \mathsf{P}_{\theta_{F_0}}(f_0 \mid v) \cdot \mathsf{P}_{\theta_{F_1}}(f_1 \mid b) \cdot \mathsf{P}_{\theta_{F_2}}(f_2 \mid b) \end{aligned}$$

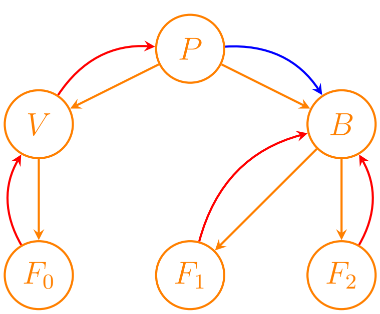
Instead, marginalize as we go, storing marginals (conditional probability tables) as ‘messages’:

$$\begin{aligned} \mathsf{P}_{\theta_{Sp}}(b) &= \sum_{p,v,f_0,f_1,f_2} \mathsf{P}_{\theta_{Sp}}(p,v,b,f_0,f_1,f_2) \\ &\stackrel{\text{def}}{=} \sum_{p} \Bigg(_{p} \mathsf{P}(p) \cdot \Bigg(_{p} \sum_{v} \mathsf{P}(v \mid p) \cdot \Bigg(_{v} \sum_{f_0} \mathsf{P}(f_0 \mid v) \Bigg)\Bigg)\Bigg) \cdot \mathsf{P}(b \mid p) \cdot \Bigg(_{b} \sum_{f_1} \mathsf{P}(f_1 \mid b) \Bigg) \cdot \Bigg(_{b} \sum_{f_2} \mathsf{P}(f_2 \mid b) \Bigg) \end{aligned}$$

(Re-arrangement of terms just comes from distributing products over sums in the full joint.)

Blue parens show ***forward messages***: distributions over free modeled variables (subscripts).

Red parens show ***backward messages***: likelihood fns over free conditioned-on variables (subscr).



Now just need space of a conditional probability distribution per variable!

14.2 Example

For example, to solve the following query (where variable F_0 is actually observed):

$$\mathsf{P}_{\theta_{Sp}}(b, f_0{=}12) = \sum_{p,v,f_1,f_2} \mathsf{P}_{\theta_{Sp}}(p, v, b, f_0{=}12, f_1, f_2)$$

$$\stackrel{\text{def}}{=} \sum_{p} \left(\mathsf{P}(p) \cdot \left(\sum_{v} \mathsf{P}(v \mid p) \cdot \mathsf{P}(f_0{=}12 \mid v)\right)_{p}\right)_{p} \cdot \mathsf{P}(b \mid p) \cdot \left(\sum_{f_1} \mathsf{P}(f_1 \mid b)\right)_{b} \cdot \left(\sum_{f_2} \mathsf{P}(f_2 \mid b)\right)_{b}$$

given the following models:

$\mathsf{P}_{\theta_P}(P) =$

/i/	/u/
.4	.6

$\mathsf{P}_{\theta_V}(V \mid P) =$

P	$+$	$-$
/i/	.8	.2
/u/	1	0

$\mathsf{P}_{\theta_{F_0}}(F_0 \mid V) =$

V	...	11	12	13	...
$+$	...	.04	.02	.01	...
$-$	...	.01	.01	.01	...

$\mathsf{P}_{\theta_B}(B \mid P) =$

P	$+$	$-$
/i/	0	1
/u/	.5	.5

we would generate the following messages:

from F_0 to V: $\mathsf{P}(F_0{=}12 \mid V) =$

V	12
$+$	.02
$-$	.01

from V to P: $\mathsf{P}(F_0{=}12 \mid P) =$

P	$F_0 = 12$
/i/	$\mathsf{P}_{\theta_{F_0}}(12 \mid +) \cdot \mathsf{P}_{\theta_V}(+ \mid \text{/i/}) + \mathsf{P}_{\theta_{F_0}}(12 \mid -) \cdot \mathsf{P}_{\theta_V}(- \mid \text{/i/})$ $= .02 \cdot .8 + .01 \cdot .2 = .018$
/u/	$\mathsf{P}_{\theta_{F_0}}(12 \mid +) \cdot \mathsf{P}_{\theta_V}(+ \mid \text{/u/}) + \mathsf{P}_{\theta_{F_0}}(12 \mid -) \cdot \mathsf{P}_{\theta_V}(- \mid \text{/u/})$ $= .02 \cdot 1 + .01 \cdot 0 = .020$

from P to B: $\mathsf{P}(P, F_0{=}12) =$

P=/i/, F_0=12	P=/u/, F_0=12
$\mathsf{P}_{\theta_P}(\text{/i/}) \cdot \mathsf{P}(F_0{=}12 \mid P{=}\text{/i/})$ $= .4 \cdot .018 = .0072$	$\mathsf{P}_{\theta_P}(\text{/u/}) \cdot \mathsf{P}(F_0{=}12 \mid P{=}\text{/u/})$ $= .6 \cdot .020 = .0120$

from F_1 to B: $\mathsf{P}(\text{any } F_1 \mid B) =$

B	any
$+$	1
$-$	1

from F_2 to B: $\mathsf{P}(\text{any } F_2 \mid B) =$

B	any
$+$	1
$-$	1

Product of model and three messages at B:

$\mathsf{P}(B, F_0{=}12) =$

$B{=}+, F_0{=}12$	$B{=}-, F_0{=}12$
$\mathsf{P}(P{=}/i/, F_0{=}12) \cdot \mathsf{P}_B(+ \mid /i/) \cdot 1 \cdot 1$	$\mathsf{P}(P{=}/i/, F_0{=}12) \cdot \mathsf{P}_B(- \mid /i/) \cdot 1 \cdot 1$
$+\, \mathsf{P}(P{=}/u/, F_0{=}12) \cdot \mathsf{P}_B(+ \mid /u/) \cdot 1 \cdot 1$	$+\, \mathsf{P}(P{=}/u/, F_0{=}12) \cdot \mathsf{P}_B(- \mid /u/) \cdot 1 \cdot 1$
$= .0072 \cdot 0 \cdot 1 \cdot 1 + .0120 \cdot .5 \cdot 1 \cdot 1 = .0060$	$= .0072 \cdot 1 \cdot 1 \cdot 1 + .0120 \cdot .5 \cdot 1 \cdot 1 = .0132$

Normalized:

$\mathsf{P}(B \mid F_0{=}12) =$

$B{=}+$	$B{=}-$
$\frac{.0060}{.0060+.0132} = .3125$	$\frac{.0132}{.0060+.0132} = .6875$

14.3 Example program

Find $\mathsf{P}(B \mid F_0 = 12)$ from Model `modP`, CondModels `modV`, `modB`, `modF0`, `modF1`, `modF2`:

```
bkwdF0 = {}
for v in modF0:    # obtain likelihood of observation given V (backward message)
  bkwdF0[v] = modF0[v]['12']
bkwdV = {}
for p in modV:     # marginalize or 'sum out' V to get likelihood given P (bkwd msg)
  for v in modV[p]:
    bkwdV[p] = bkwdV.get(p,0.0) + (modV[p][v] * bkwdF0[v])
fwrdP = {}
for p in modP:     # multiply prior over P by likelihood given P (backward message)
  fwrdP[p] = modP[p] * bkwdV[p]
...
```

Practice

Complete the above example.

14.4 Limits of message passing

Message passing degrades when network is not singly-connected.

For example, adding variable for height w. dependencies from P, to F_2, creates a 'diamond':

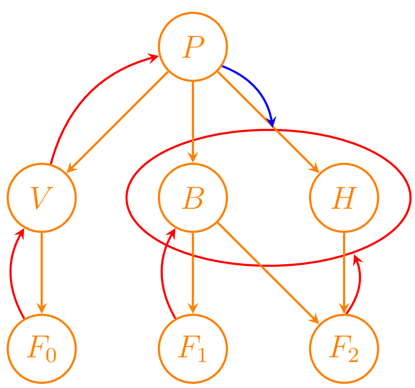
$$\langle P \times V \times B \times H \times F_0 \times F_1 \times F_2,\ 2^{P \times V \times B \times H \times F_0 \times F_1 \times F_2},\ \mathsf{P} \rangle$$
$$\text{where } P = \{/i/, /u/\}, V = \{+,-\}, B = \{+,-\}, H = \{+,-\}, F_0 = \mathbb{I}_0^{99}, F_1 = \mathbb{I}_0^{99}, F_2 = \mathbb{I}_0^{99}$$

$$\mathsf{P}_{\theta_P}(P) \stackrel{\text{def}}{=} \mathsf{P}(P)$$
$$\mathsf{P}_{\theta_V}(V \mid P) \stackrel{\text{def}}{=} \mathsf{P}(V \mid P)$$
$$\mathsf{P}_{\theta_B}(B \mid P, V) \stackrel{\text{def}}{=} \mathsf{P}(B \mid P)$$
$$\mathsf{P}_{\theta_H}(H \mid P, V, B) \stackrel{\text{def}}{=} \mathsf{P}(H \mid P)$$
$$\mathsf{P}_{\theta_{F_0}}(F_0 \mid P, V, B, H) \stackrel{\text{def}}{=} \mathsf{P}(F_0 \mid V)$$
$$\mathsf{P}_{\theta_{F_1}}(F_1 \mid P, V, B, H, F_0) \stackrel{\text{def}}{=} \mathsf{P}(F_1 \mid B)$$
$$\mathsf{P}_{\theta_{F_2}}(F_2 \mid P, V, B, H, F_0, F_1) \stackrel{\text{def}}{=} \mathsf{P}(F_2 \mid B, H)$$

This means some marginals will have multiple free variables (which makes them larger):

$$\begin{aligned}\mathsf{P}_{\theta_{Sp}}(b) &= \sum_{p,v,h,f_0,f_1,f_2} \mathsf{P}_{\theta_{Sp}}(p, v, b, h, f_0, f_1, f_2) \\ &\stackrel{\text{def}}{=} \sum_{p} \Bigg(_{\!p} \mathsf{P}(p) \cdot \Bigg(_{\!p} \sum_{v} \mathsf{P}(v \mid p) \cdot \ldots \Bigg)\Bigg) \cdot \mathsf{P}(b \mid p) \cdot \Bigg(_{\!b} \sum_{f_1} \mathsf{P}(f_1 \mid b) \Bigg) \cdot \sum_{h} \mathsf{P}(h \mid p) \cdot \Bigg(_{\!b,h} \sum_{f_2} \mathsf{P}(f_2 \mid b, h) \Bigg)\end{aligned}$$

Graphically, messages must pass through 'junctions' of joint variables:



Well, they're not full joints at least.