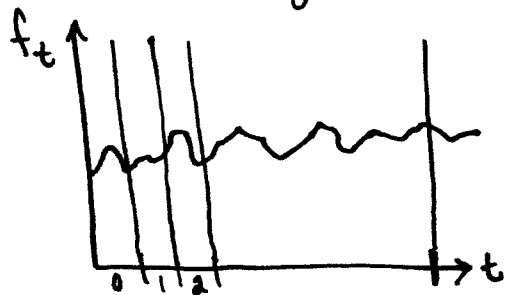


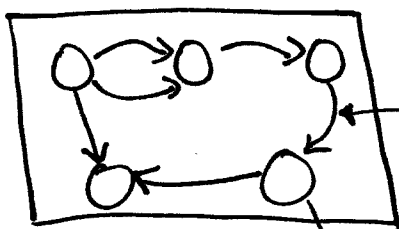
EXAMPLES OF HMMAaron Kemp
akemp@ucsd.edu2. Speech Recognition

- discretize time

$$t = 0, 1, \dots, T-1$$

- classify each time slice as being one of M different output types
(vector quantization, spectral method) $M \sim 256$ - generates output sequence $O_0, O_1, \dots, O_{T-1}$ taking values in $\{1, 2, \dots, M\}$ Suppose possible words are $W_1, \dots, W_k$

Associate hidden Markov model with each word.

Associate $\lambda = (a, b, \pi)$ with each word. ^{Given λ} gives $\lambda_1, \dots, \lambda_k$. Pick value of $\{\lambda_i\}_{i=1}^k$ that maximizes $P^{\lambda_i}(O) = P(O | \text{model has parameter } \lambda_i)$ π = initial distribution of MC q .

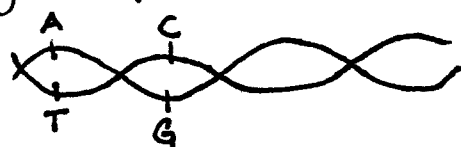
$$a = (a_{ij})_{i,j=1}^N$$

$$P(O_t = j | q_t = i) = b_{ij}$$

3. Biological Sequence Analysis

(Durbin, Eddy, Krogh, Mitchison)

Locating CpG islands

DNA

A, C, G, T

Focus on one strand

Pair CG: ^{normally} Where this occurs, there is a high probability that C will mutate to T. P.2

For biologically important reasons, this mutation is suppressed in short stretches of the genome (e.g., around "start" regions of many genes).

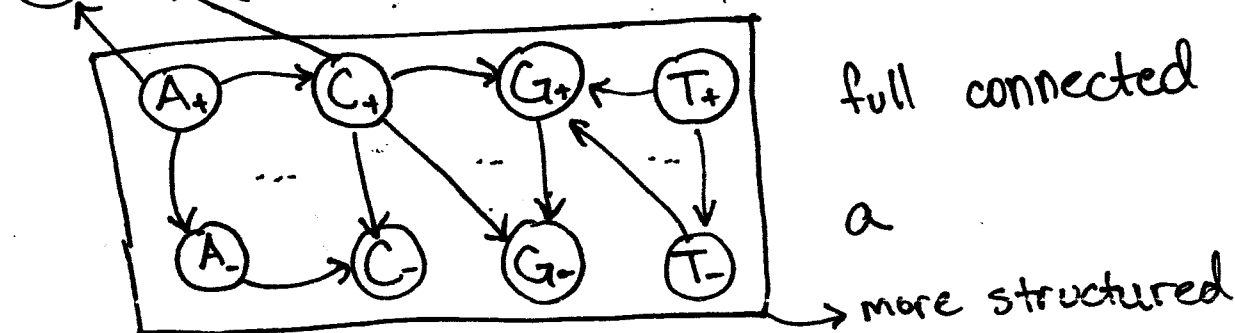
In these regions, see many more CpG pairs than elsewhere (called CpG islands).

Qn: Given a long piece of sequence, how can we find CpG islands?

Output set = $\{A, C, G, T\}$

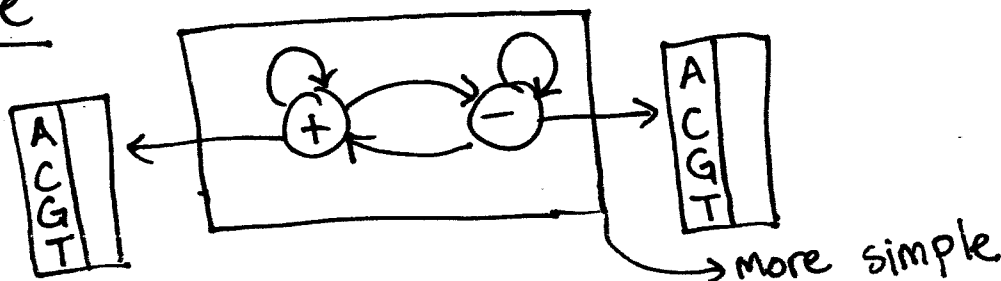
States for HMM $\{A_+, C_+, G_+, T_+, A_-, C_-, G_-, T_-\}$

C_+ (+) - indicates in CpG island
 C_- (-) - " not in CpG island



Alternative

Here prob. of transitioning from $\oplus$ to $\ominus$ cannot depend on letter output.



(p.8)

Three main problems for HMMs

1. Training (system identification)

Fix N, M

Variable parameter $\lambda = (a, b, \pi)$

$$\begin{cases} \sum_j a_{ij} = 1, a_{ij} \geq 0 \\ \sum_k b_{ik} = 1, b_{ik} \geq 0 \\ \sum_i \pi_i = 1, \pi_i \geq 0 \end{cases}$$

Given training data (sample output sequences) for W different "words" (entities)

Let $\mathcal{W}_i$ be a collection of output sequences $(O_0, O_1, \dots, O_{T-1})$ for "word" i , $i=1, \dots, W$.

Choose λ_i (parameter for "word" i) to maximize $\sum_{Q_j \in \mathcal{W}_i} S^{\lambda_i}(Q_j)$

where $S^{\lambda_i}(Q_j) = \log P^{\lambda_i}(Q_j)$

For CpG island example, also know (from annotated data) where $Q_j = (q_0^j, q_1^j, \dots, q_{T-1}^j)$ ^{is, then} ~~use~~
 $S^{\lambda_i}(O_j, Q_j) = \log P^{\lambda_i}(O_j, Q_j)$ in place of $S^{\lambda_i}(Q_j)$

2. Decoding (fix λ) P.4

Identify the hidden sequence $(q_0, q_1, \dots, q_{T-1})$ that best generates the observation sequence $O = (O_0, O_1, \dots, O_{T-1})$.

Find $q^* = (q_0^*, \dots, q_{T-1}^*)$: maximize $P^\lambda(O, q)$

Equivalent to find
 $\underset{q}{\operatorname{argmax}} P^\lambda(q|O)$

$$q^* = \underset{q}{\operatorname{argmax}} P^\lambda(O, q)$$

$$P^\lambda(O, q) = \underbrace{P^\lambda(q|O)} P^\lambda(O)$$

3. Recognition (Classification)

Given O , consider a finite set of HMMs indexed by $\lambda_1, \dots, \lambda_W$

$$\lambda^* = \underset{i=1}{\overset{W}{\operatorname{argmax}}} P^{\lambda_i}(O)$$

In training and recognition, need to be able to compute

$$P^\lambda(O) \quad O = (O_0, O_1, \dots, O_{T-1})$$

efficiently

Blind approach: $P^\lambda(O) = \sum_{Q=(q_0, \dots, q_{T-1})} P^\lambda(O, Q) =$

$$\sum_Q \underbrace{\pi_{q_0}}_{1 \text{ calc}} \underbrace{a_{q_0 q_1} \dots a_{q_{T-2} q_{T-1}}}_{T-1 \text{ calcs}} \underbrace{b_{q_0 O_0} \dots b_{q_{T-1} O_{T-1}}}_{T-1 \text{ calcs}}$$

of calculations: N^T terms in sum ; $2T$ Multiplications per summand

p.5

$$\begin{aligned} \# \text{ mult} &= N^T \cdot 2T \\ \# \text{ addns} &= N^T - 1 \end{aligned} \quad \# \text{ operations order } \underbrace{TN^T}_{\text{large}}$$

Forward Method - efficient.

Fix $O^* = (O_0^*, O_1^*, \dots, O_{T-1}^*)$, λ

Define $\alpha_t^\lambda(i) = P^\lambda(Q_t = i, O_0 = O_0^*, \dots, O_t = O_t^*)$
 $t = 0, 1, \dots, T-1$
 $i = 1, \dots, N$

(i) $\alpha_0^\lambda(i) = P^\lambda(Q_0 = i, O_0 = O_0^*)$
 $= \pi_i b_{i, O_0^*}; i = 1, \dots, N$

(ii) $\alpha_{t+1}^\lambda(j) = P^\lambda(Q_{t+1} = j, O|_{t+1} = (O_0^*, \dots, O_{t+1}^*))$
 $= \sum_{i=1}^N P^\lambda(Q_{t+1} = j, Q_t = i, O|_{t+1} = (O_0^*, \dots, O_{t+1}^*))$
 $= \sum_{i=1}^N P^\lambda(Q_{t+1} = j, O_{t+1} = O_{t+1}^* | Q_t = i, O|_t = O|_t^*)$
 $\quad * P^\lambda(Q_t = i, O|_t = O|_t^* | t)$

$= \sum_{i=1}^N (a_{ij} b_{j, O_{t+1}^*}) (\alpha_t^\lambda(i)) \quad \begin{matrix} j = 1, \dots, N \\ t = 0, 1, \dots, T-2 \end{matrix}$

$P^\lambda(O^*) = \sum_{i=1}^N \underbrace{\alpha_{T-1}^\lambda(i)}_{P^\lambda(Q_{T-1} = i, O = O^*)}$ # of calcs is order $N^2 T$