Μηχανική Μάθηση

Μιχάλης Τίτσιας

Διάλεξη 10ή

Markov models και *Hidden Markov Models*

Περιεχόμενα

- Ακολουθιακά δεδομένα
- Markov models
- Εφαρμογή σε επεξεργασία φυσικής γλώσσας
- Hidden Markov models
- Εφαρμογές των Hidden Markov Models
- Ο αλγόριθμος EM για HMMs
- Ο αλγόριθμος forward-backward
- Παράδειγμα υπολογισμού των α και β μηνυμάτων
- Ο αλγόριθμος Viterbi
- Παράδειγμα εφαρμογής του EM αλγορίθμου και εύρεσης βέλτιστου μοναπατιού

Ακολουθιακά δεδομένα

- Ως τώρα (σε όλα τα προβλήματα μάθησης που έχουμε δει) έχουμε βασιστεί στην υποθέση ότι τα διαθέσιμα δεδομένα $X = \{\mathbf{x}_n\}_{n=1}^N$ είναι ανεξάρτητα μεταξύ τους
- Αυτό μας επέτρεπε να γράφουμε την από κοινού κατανομή/πιθανοφάνεια των δεδομένων ως το γινόμενο

$$p(X) = \prod_{n=1}^{N} p(\mathbf{x}_n)$$

- Ωστόσο σε πάρα πολλές εφαρμογές έχουμε πρόσβαση σε δεδομένα τα οποία είναι στατιστικά εξαρτημένα

$$p(X) \neq \prod_{n=1}^{N} p(\mathbf{x}_n)$$

- Η σημαντικότερη περίπτωση τέτοιων δεδομένων είναι τα ακολουθιακά δεδομένα (sequence data)

Ορισμός ακολουθιακών δεδομένων

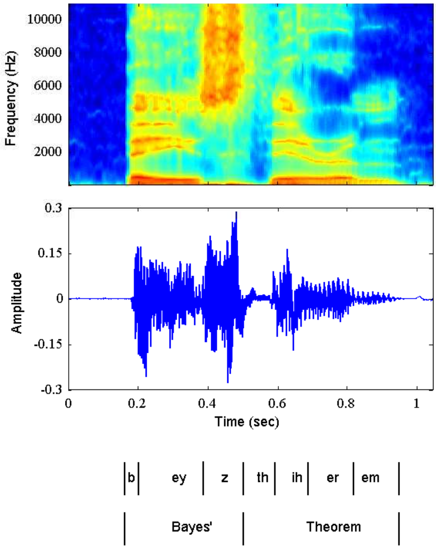
- Έστω δεδομένα $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N-1}, \mathbf{x}_N)$. Θα ονομάζονται ακολουθιακά όταν η σειρά με την οποία εμφανίζονται έχει σημασία
  - $\Rightarrow$ τα δεδομένα είναι κατά φυσικό τρόπο ταξινομημένα υπό την έννοια ότι τα γειτονικά του $\mathbf{x}_n$ είναι το $\mathbf{x}_{n-1}$ και $\mathbf{x}_{n+1}$

Ακολουθιακά δεδομένα

Παραδείγματα ακολουθιακών δεδομένων

- Χρονοσειρές:
  - Ημερήσια δεδομένα που περιγράφουν το καιρό, ημερήσιες τιμές μετοχών ή ισοτιμίες νομισμάτων κτλ
  - Ακολουθιακά δεδοδένα ήχου $\Rightarrow$ speech recognition
  - Δεδομένα εικόνων σε μορφή video
  - Εβδομαδιαία αποτελέσματα αγώνων ποδοσφαίρου
- Χωρικά εξαρτώμενα δεδομένα:
  - DNA δεδομένα: Π.χ. οι βάσεις κατά μήκος του DNA έχουν χωρική εξαρτησία
  - Φυσική γλώσσα: οι χαρακτήρες ενός κειμένου έχουν συσχέτιση μεταξύ τους

Κυματομορφή ήχου για την φράση Bayes' theorem



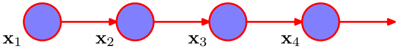
Markov models

- Έστω ακολουθιακά δεδομένα $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N-1}, \mathbf{x}_N)$
- Ένας απλοικός τρόπος μοντελοποίησης των δεδομένων αυτών είναι να τα θεωρήσουμε ως ανεξάρτητα (δηλ. να υποθέσουμε ότι η από κοινού κατανομή τους είναι $\prod_{n=1}^{N} p(\mathbf{x}_n)$)
  - $\Rightarrow$ αγνοεί την εξάρτηση, π.χ. αν $\mathbf{x}_{N+1}$ είναι το άγνωστο αποτέλεσμα του άγωνα ποδοσφαίρου της ερχόμενης εβδομάδος το παραπάνω μοντέλο θα το προέβλεπε μόνο βάσει της συχνότητας με την οποία κερδίζει μια ομάδα (αγνοώντας τα συγκεκριμένα γνωστά αποτελέσματα των προηγουμένων εβδομάδων)
- Σε αντίθεση το πιο γενικό μοντέλο θα είχε την μορφή

$$p(X) = p(\mathbf{x}_1) \prod_{n=2}^{N} p(\mathbf{x}_n | \mathbf{x}_{n-1}, \mathbf{x}_{n-2}, \ldots, \mathbf{x}_1)$$

- Markov models απλουστεύουν το παραπάνω υποθέτοντας ότι η κάθε υπό συνθήκη κατανομή εξαρτάται μόνο από το προηγούμενο δεδομένο

Markov models



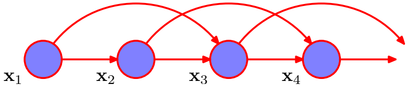
- Markov models (πρώτης τάξης)

$$p(X) = p(\mathbf{x}_1) \prod_{n=2}^{N} p(\mathbf{x}_n | \mathbf{x}_{n-1})$$

- Αν η συναρτησιακή μορφή της κάθε υπό συνθηκή κατανομής δεν αλλάζει με το χρόνο τότε το μοντέλο ονομάζεται ομοιογενές ή στατικό (homogeneous or stationary)
- Όταν κάθε $\mathbf{x}_n$ παίρνει διακριτές τιμές, δηλ. $\mathbf{x}_n \in \{1, \ldots, K\}$ το παραπάνω μοντέλο ονομάζεται αλυσίδα Markov (Markov chain)
- Παρακάτω θα μας αποσχολήσουν ομοιογενή μοντέλα τα οποία αποτελούν αλυσίδες Markov

Markov models

(Παρένθεση: μπορούμε γενικότερα να ορίσουμε Markov models υψηλότερης τάξης)



- 2-τάξης Markov model

$$p(X) = p(\mathbf{x}_1)p(\mathbf{x}_2|\mathbf{x}_1)\prod_{n=3}^{N} p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{x}_{n-2})$$

Markov models

- Έστω ομοιογενές Markov model/chain $p(X) = p(\mathbf{x}_1)\prod_{n=2}^{N} p(\mathbf{x}_n|\mathbf{x}_{n-1})$ όπου κάθε $\mathbf{x}_n$ παίρνει K τιμές (βάσει 1 of K encoding)
- Η αρχική κατανομή καθορίζεται από ένα διάνυσμα

$$\pi_k = p(\mathbf{x}_{1k} = 1), \quad \sum_{k=1}^{K} \pi_k = 1$$

- Η κατανομή μετάβασης (transition density) $p(\mathbf{x}_n|\mathbf{x}_{n-1})$ καθορίζεται από $K \times K$ πίνακα μετάβασης

$$A_{jk} = p(\mathbf{x}_{nk} = 1|\mathbf{x}_{(n-1)j} = 1), \quad \sum_{k=1}^{K} A_{jk} = 1, \ j = 1, \ldots, K$$

- (π, A) αποτελούν τις παραμέτρους που καθορίζουν το μοντέλο
  - $\Rightarrow$ εκπαίδευση με μέγιστη πιθανοφάνεια

Markov models

- Δοθέντος μιας ακολουθίας δεδομένων $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ η πιθανοφάνεια είναι

$$
\begin{aligned}
p(X|\boldsymbol{\pi}, A) &= p(\mathbf{x}_1) \prod_{n=2}^{N} p(\mathbf{x}_n|\mathbf{x}_{n-1}) \\
&= \left(\prod_{k=1}^{K} \pi_k^{x_{1k}} \right) \prod_{n=2}^{N} \left(\prod_{k=1}^{K} \prod_{j=1}^{K} A_{jk}^{x_{(n-1)j} x_{nk}} \right)
\end{aligned}
$$

- Η μεγιστοποίηση της πιθανοφάνειας (με το συνήθη τρόπο, δηλ. παίρνωντας λογάριθμο κτλ) δίνει για το πίνακα μετάβασης A

$$
A_{jk} = \frac{\sum_{n=2}^{N} x_{(n-1)j} x_{nk}}{\sum_{l=1}^{K} \sum_{n=2}^{N} x_{(n-1)j} x_{nl}}
$$

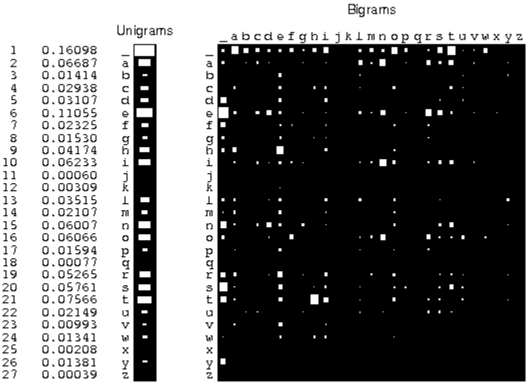
που απλά είναι ο κανονικοποιημένος αριθμός των φορών που είχαμε μετάβαση από την κατάσταση j στην k

Εφαρμογή σε επεξεργασία φυσικής γλώσσας

- Έστω κάθε δεδομένο $\mathbf{x}_n$ αντιπροσωπεύει μια λέξη ή γράμμα/σύμβολο από ένα κείμενο φυσικής γλώσσας
- Μοντέλα γλώσσας αποτελούν πιθανοτικές κατανομές ως προς τις ακολουθίες των λέξεων ή συμβόλων
- Unigram: Αν χρησιμοποιήσουμε ένα μοντέλο που υποθέτει ότι οι λέξεις παράγονται ανεξάρτητα μεταξύ τους, δηλ. $p(X) = \prod_{n=1}^{N} p(\mathbf{x}_n)$ (0-τάξης Markov)
  - $\Rightarrow$ η πιθανότητα $p(x_{nk} = 1)$ ονομάζεται unigram
- Bigram: Αν υποθέσουμε ένα Markov model πρώτης τάξης $p(X) = p(\mathbf{x}_1) \prod_{n=2}^{N} p(\mathbf{x}_n | \mathbf{x}_{n-1})$
  - $\Rightarrow$ η πιθανότητα $p(x_{nk} = 1 | x_{(n-1)j} = 1)$ ονομάζεται bigram
- Πιο γενικά κάποιος μπορεί να ορίσει ένα n-gram μοντέλο, δηλ. n-τάξης μοντέλο Markov

Εφαρμογή σε επεξεργασία φυσικής γλώσσας

Εφαρμογή unigram και bigram στο βιβλίο του Darwin “On The Origin of Species” για τα σύμβολα (a,. . . ,z, -) όπου - συμβολίζει το κενό



Το μέγεθος των άσπρων κουτιών είναι ανάλογο της τιμής (πιθανότητας) του αντίστοιχου στοιχείου του διάνυσματος ή του πίνακα

Hidden Markov Models

- Εκπαίδευση ενός Markov μοντέλου πρώτης τάξης απαιτεί τον καθορισμό $O(K^2)$ παραμέτρων (τιμές του πίνακα μετάβασης A)
- Μοντέλα M-τάξης απαιτούν τον καθορισμό $O(K^{M+1})$ παραμέτρων
  - δεν είναι εφικτό για μεγάλο M
  - ακόμα και αν ορίζαμε M-τάξης μοντέλα με λίγες παραμέτρους, πάλι θα περιοριζόμασταν από το γεγονός ότι το μοντέλο είναι M-τάξης
- Θα θέλαμε να ορίσουμε μοντέλα που δεν περιορίζονται από κάποια τάξη αλλά συγχρόνως διατηρούν τον αριθμό των δεδομένων στο ελάχιστο, δηλ. $O(K^2)$
  - τέτοια μοντέλα είναι τα Hidden Markov Models

Hidden Markov Models

Ιδέα: Hidden Markov Models (HMMs) μπορούν να οριστούν ώς μια γενίκευση του μοντέλου μίξης κατανομών

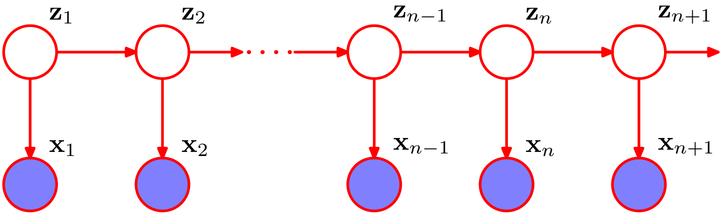
- Στις μίξεις κατανομών, είχαμε ανεξάρτητες κρυμμένες μεταβλήτες $\{\mathbf{z}_n\}_{n=1}^N$ που καθόριζαν την παραγωγή των δεδομένων $\{\mathbf{x}_n\}_{n=1}^N$

$$p(X, Z) = \prod_{n=1}^{N} p(\mathbf{x}_n|\mathbf{z}_n)p(\mathbf{z}_n) = \left(\prod_{n=1}^{n} p(\mathbf{z}_n)\right)\left(\prod_{n=1}^{N} p(\mathbf{x}_n|\mathbf{z}_n)\right)$$

- Σε ένα HMM η παραπάνω μίξη τροποποιείται ώστε οι κρυμμένες μεταβλητές ακολουθούν μια αλυσίδα Markov, δηλ.

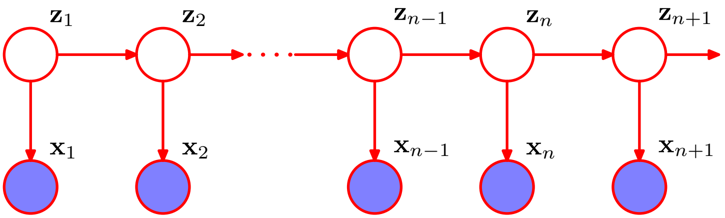
$$p(Z) = p(\mathbf{z}_1)\prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1})$$

Hidden Markov Models



$$P(X, Z) = \left(p(\mathbf{z}_1) \prod_{n=2}^{N} p(\mathbf{z}_n | \mathbf{z}_{n-1}) \right) \left(\prod_{n=1}^{N} p(\mathbf{x}_n | \mathbf{z}_n) \right)$$

Hidden Markov Models

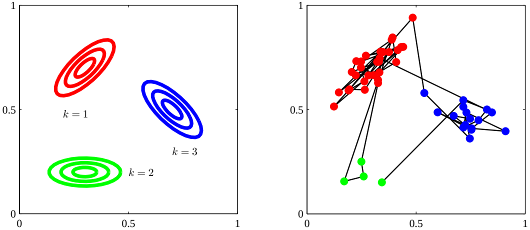


- $p(z_{1k} = 1) = \pi_k \Rightarrow$ αρχική κατανομή
- $p(z_{nk} = 1 | z_{(n-1)j} = 1) = A_{jk} \Rightarrow$ πίνακας μετάβασης
- $p(\mathbf{x}_n | \mathbf{z}_n)$ ονομάζεται κατανομή εκπομπής (emission density). Όπως και στις μίξεις μπορεί να είναι π.χ. Gaussian

$$p(\mathbf{x}_n | z_{nk} = 1) = \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \Sigma_k)$$

και όποτε έχουν K διαφορετικές Gaussian κατανομές που αντιστοιχούν στις K τιμές της κρυμμένης μεταβλητής

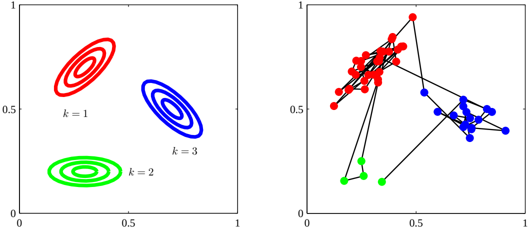
Hidden Markov Models



$$P(X,Z) = \left(\prod_{n=1}^{N} p(\mathbf{x}_n|\mathbf{z}_n)\right)\left(p(\mathbf{z}_1)\prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1})\right)$$
$$= p(\mathbf{z}_1)p(\mathbf{x}_1|z_1)p(\mathbf{z}_2|\mathbf{z}_1)p(\mathbf{x}_2|z_2)\ldots$$

- Ένα HMM υποθέτει ότι τα δεδομένα παράγονται ακολουθιακά
  - Έπελεξε πρώτα την ομάδα $\mathbf{z}_1$ του πρώτου δεδομένου βάσει της $p(\mathbf{z}_1)$ και έπειτα το δεδομένο το ίδιο $\mathbf{x}_1$
  - Έπελεξε την ομάδα $\mathbf{z}_2$ του δεύτερου δεδομένου βάσει της $p(\mathbf{z}_2|\mathbf{z}_1)$ και έπειτα την τιμή του $\mathbf{x}_2$
  - κτλ

Hidden Markov Models



$$P(X,Z) = p(\mathbf{z}_1)p(\mathbf{x}_1|z_1)p(\mathbf{z}_2|\mathbf{z}_1)p(\mathbf{x}_2|z_2)\ldots$$

- Στο Σχήμα οι κατανομές εκπομπής έχουν υποτεθεί Gaussian

$$p(\mathbf{x}_n|z_{nk}=1) = \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \Sigma_k), \quad k=1,2,3$$

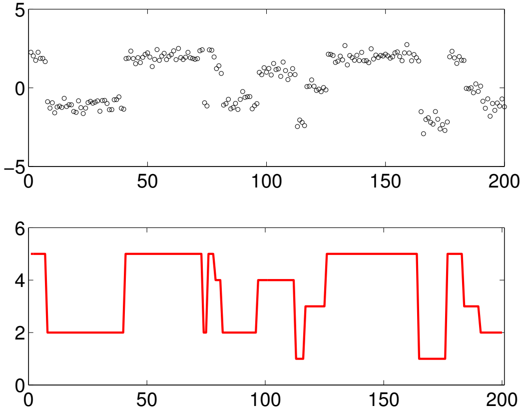
- Δοθέντος ότι οι τιμές της διαγώνιου του πίνακα μετάβασης (δηλ. οι τιμές A_{kk}, $k=1,2,3$) είναι μεγάλες το HMM παρουσιάζει "sticky" συμπεριφορά
  - δηλ. παραμονή στην ίδια ομάδα/κατάσταση για σειρά δεδομένων

Hidden Markov Models

- Διαισθητικά το HMM υποθέτει ότι η παραγωγή των δεδομένων ακολουθεί διαφορετικές φάσεις ή περιόδους πράγμα το οποίο μοντελοποιείται μέσω της αλυσίδας Markov που ακολουθούν οι κρυμμένες μεταβλητές $\mathbf{z}_n$
  - Π.χ. αν τα δεδομένα είναι δυαδικά και καταγράφουν το αν έβρεξε ή δεν έβρεξε την κάθε μέρα, η κρυμμένη μεταβλητή μπορεί να αντιστοιχεί στην εποχή τους έτους ώστε ανάλογα με την εποχή η πιθανότητα να βρέξει αλλάζει
- Σε αντίθεση τα Markov Models δεν υποθέτουν κρυμμένες φάσεις, αλλά υποθέτουν ότι η τιμή του τρέχοντος δεδομένου εξαρτάται από ένα παράθυρο προηγούμενων τιμών

Hidden Markov Models

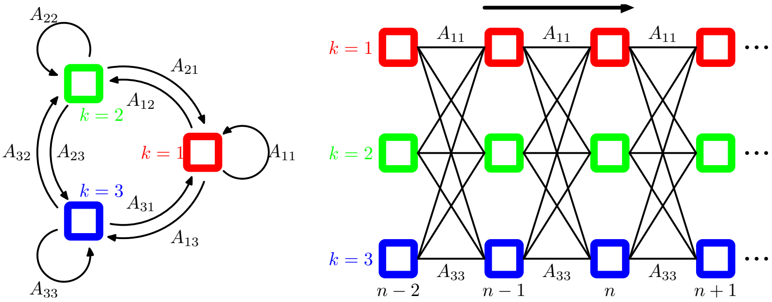
Παράδειγμα



Τα δεδομένα του σχήματος (πάνω εικόνα) έχουν παραχθεί με τεχνητό τρόπο από $K = 5$ καταστάσεις. Το μονοπάτι (δηλ. η ακολουθία των κρυμμένων μεταβλητών/καταστάσεων $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N$) που έχει παράγει τα δεδομένα φαίνεται με την κόκκινη γραμμή (κάτω εικόνα)

Παρατηρούμε ότι τα δεδομένα παρουσιάζουν "sticky" συμπεριφορά (δηλ. έχουμε παραμονή στην ίδια ομάδα/κατάσταση για σειρά δεδομένων)

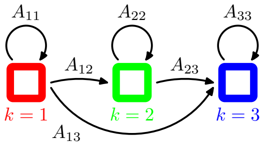
Hidden Markov Models

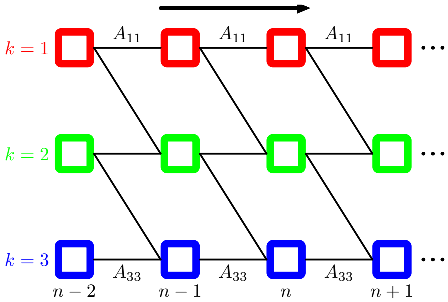


Οπτικοποίηση της δομής του πίνακα μετάβασης A (που καθορίζει το πως ακολουθιακά αλλάζουν οι ομάδες/καταστάσεις) γίνεται με

- διαγράμματα που έχουν την μορφή αυτόματου (δεξιά στην εικόνα)
- ή διαγράμματα τύπου trellis (αριστερά στην εικόνα)

Hidden Markov Models





- Σε πολλά HMMs ο πίνακας μετάβασης μπορεί να εμπεριέχει περιορισμούς ώστε κάποιες μεταβάσεις να μην επιτρέπονται
- Στο Σχήμα φαίνονται δύο παραδείγματα πινάκων μετάβασης για left-to-right HMMs όπου το μοντέλο ξεκινά από την κατάσταση $k = 1$ και έπειτα ακολουθιακά είτε μεταπηδά σε μια νέα κατάσταση ή παραμένει στην τρέχουσα

Εφαρμογές των Hidden Markov Models

- Automatic speech recognition: Κάθε δεδομένο $\mathbf{x}_n$ αναπαριστά χαρακτηριστικά από ακουστικό σήμα φωνής ενώ η κρυμμένη μεταβλητή $\mathbf{z}_n$ υποδεικνύει την λέξη στην οποία το ακουστικό σήμα φωνής αντιστοιχεί. Ο πίνακας μετάβασης $p(\mathbf{z}_n|\mathbf{z}_{n-1})$ ουσιαστικά είναι το μοντέλο φυσικής γλώσσας (δηλ. πως μεταπηδούμε από λέξη σε λέξη κατά την ροή του λόγου) ενώ η κάθε κατανομή εκπομπής $p(\mathbf{x}_n|\mathbf{z}_n)$ μοντελοποιεί το σήμα που καταγράφεται/παρατηρείται
- Activity recognition: Κάθε δεδομένο $\mathbf{x}_n$ αναπαριστά χαρακτηριστικά από ένα frame ενός video, ενώ η κρυμμένη μεταβλητή $\mathbf{z}_n$ αντιπροσωπεύει την ασχολία/ενέργεια με την οποία καταπιάνεται ένα άτομο που φαίνεται στο video. Για παράδειγμα το κάθε $\mathbf{z}_n$ μπορεί να παιρνεί τιμές από το σύνολο (τρέξιμο, περπάτημα, ενέργεια κλοπής κτλ)
- Part of speech tagging: Κάθε δεδομένο $\mathbf{x}_n$ αναπαριστά λέξη σε ένα κείμενο, ενώ το $\mathbf{z}_n$ αναπαριστά το τι μέρος του λόγου είναι αυτή η λέξη, π.χ. υποκείμενο, ρήμα, αντικείμενο
- Gene finding: Κάθε δεδομένο $\mathbf{x}_n$ αναπαριστά μια από τις 4 βάσεις του DNA (A,C,G,T), ενώ η $\mathbf{z}_n$ παίρνει δύο τιμές που αναπαριστούν αν μια βάση βρίσκεται εντός ή εκτός περιοχής που κωδικοποιεί ένα γονίδιο (gene-coding region)
- Πολλές άλλες εφαρμογές σε βιοπληροφορική, επεξεργασία κειμένου, χρονοσειρές στα οικονομικά κα

Ο αλγόριθμος EM για HMMs

- Έχουμε δεδομένα $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ για τα οποία υποθέτουμε ένα HMM
- Από κοινού κατανομή των δεδομένων και των κρυμμένων μεταβλητών (δοθέντος των παραμέτρων $\boldsymbol{\theta} = (\boldsymbol{\pi}, A, \phi)$):

$$\begin{aligned} p(X, Z|\boldsymbol{\theta}) &= p(\mathbf{z}_1|\boldsymbol{\pi}) \prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1}, A) \prod_{n=1}^{N} p(\mathbf{x}_n|\mathbf{z}_n, \phi) \\ &= \prod_{k=1}^{K} \pi_k^{z_{1k}} \prod_{n=2}^{N} \prod_{k=1}^{K} \prod_{j=1}^{K} A_{jk}^{z_{(n-1)j} z_{nk}} \prod_{n=1}^{N} \prod_{k=1}^{K} p(\mathbf{x}_n|\phi_k)^{z_{nk}} \end{aligned}$$

  αυτή είναι η complete πιθανοφάνεια

- Αφού Z είναι κρυμμένες μεταβλητές, για να υπολογίσουμε την πιθανοφάνεια θα πρέπει να εφαρμόσουμε τον κανόνα αθροίσματος

$$p(X|\boldsymbol{\theta}) = \sum_{Z} p(X, Z|\boldsymbol{\theta})$$

  την πιθανοφάνεια αυτή θα θέλαμε να μεγιστοποιήσουμε ως προς $\boldsymbol{\theta}$

Ο αλγόριθμος EM για HMMs

- Complete πιθανοφάνεια

$$p(X, Z|\boldsymbol{\theta}) = \prod_{k=1}^{K} \pi_k^{z_{1k}} \prod_{n=2}^{N} \prod_{k=1}^{K} \prod_{j=1}^{K} A_{jk}^{z_{(n-1)j} z_{nk}} \prod_{n=1}^{N} \prod_{k=1}^{K} p(\mathbf{x}_n|\phi_k)^{z_{nk}}$$

- Πιθανοφάνεια

$$p(X|\boldsymbol{\theta}) = \sum_{Z} p(X, Z|\boldsymbol{\theta})$$

- Αφού έχουμε να κάνουμε με κρυμμένες μεταβλητές μπορούμε να εφαρμόσουμε τον αλγόριθμο EM που βασίζεται στην μεγιστοποίηση της expected λογαριθμικής complete πιθανοφάνειας

Ο αλγόριθμος EM για HMMs

- Complete πιθανοφάνεια

$$p(X,Z|\boldsymbol{\theta}) = \prod_{k=1}^{K} \pi_k^{z_{1k}} \prod_{n=2}^{N}\prod_{k=1}^{K}\prod_{j=1}^{K} A_{jk}^{z_{(n-1)j}z_{nk}} \prod_{n=1}^{N}\prod_{k=1}^{K} p(\mathbf{x}_n|\phi_k)^{z_{nk}}$$

- Λογαριθμική complete πιθανοφάνεια

$$\mathcal{L}_c(\boldsymbol{\theta}) = \sum_{k=1}^{K} z_{1k}\log\pi_k + \sum_{n=2}^{N}\sum_{k=1}^{K}\sum_{j=1}^{K} z_{(n-1)j}z_{nk}\log A_{jk} + \sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}\log p(\mathbf{x}_n|\phi_k)$$

- Expected λογαριθμική complete πιθανοφάνεια

$$\begin{aligned} Q(\boldsymbol{\theta}) &= \sum_{k=1}^{K}\gamma(z_{1k})\log\pi_k + \sum_{n=2}^{N}\sum_{k=1}^{K}\sum_{j=1}^{K}\xi(z_{(n-1)j},z_{nk})\log A_{jk} \\ &+ \sum_{n=1}^{N}\sum_{k=1}^{K}\gamma(z_{nk})\log p(\mathbf{x}_n|\phi_k) \end{aligned}$$

Ο αλγόριθμος EM για HMMs

- Expected λογαριθμική complete πιθανοφάνεια

$$Q(\boldsymbol{\theta}) = \sum_{k=1}^{K} \gamma(z_{1k}) \log \pi_k + \sum_{n=2}^{N} \sum_{k=1}^{K} \sum_{j=1}^{K} \xi(z_{(n-1)j}, z_{nk}) \log A_{jk}$$
$$+ \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \log p(\mathbf{x}_n | \phi_k)$$

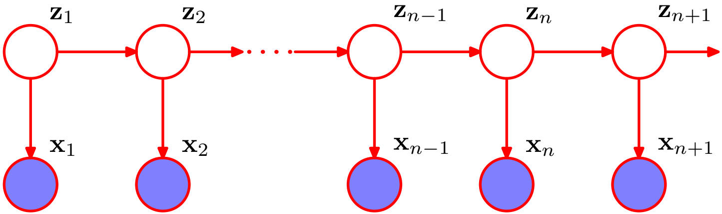
όπου $\gamma(\mathbf{z}_n) = p(\mathbf{z}_n | X, \boldsymbol{\theta}^{old})$ και $\xi(\mathbf{z}_{n-1}, \mathbf{z}_n) = p(\mathbf{z}_{n-1}, \mathbf{z}_n | X, \boldsymbol{\theta}^{old})$ είναι οι εκ των υστέρων πιθανότητες του υπολογίζονται στο E βήμα δοθέντος των προηγούμενων τιμών των παραμέτρων (αυτό απαιτεί dynamic programming όπως θα δούμε σε λίγο)

- Στο M βήμα μεγιστοποιούμε ως προς τις παραμέτρους:

$$\pi_k = \frac{\gamma(z_{1k})}{\sum_{j=1}^{K} \gamma(z_{1j})} = \gamma(z_{1k}), \quad A_{jk} = \frac{\sum_{n=2}^{N} \xi(z_{(n-1)j}, z_{nk})}{\sum_{l=1}^{K} \sum_{n=2}^{N} \xi(z_{(n-1)j}, z_{nl})},$$

ενώ η μεγιστοποίηση ως προς ϕ_k εξαρτάται από τη μορφή της κατανομής εκπομπής και οδηγεί σε πανομοιότυπες αντίστοιχες εξισώσεις με τις μίξεις κατανομών

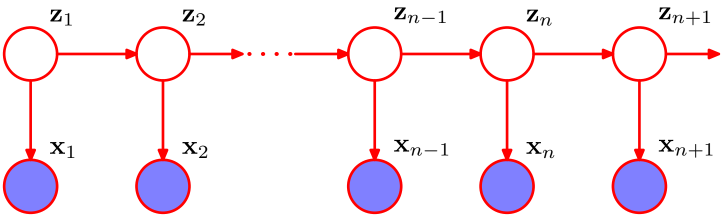
Ο αλγόριθμος forward-backward



- Ο αλγόριθμος EM για HMMs είναι πολύ όμοιος με τον αλγόριθμο EM για μίξεις κατανομών
    - Ειδικά το M βήμα είναι σχέδον πανομοιότυπο
- Ωστόσο το E βήμα διαφέρει σημαντικά λόγω του ότι ο υπολογισμός των εκ των υστέρων πιθανοτήτων $\gamma(\mathbf{z}_n)$ και $\xi(\mathbf{z}_{n-1}, \mathbf{z}_n)$ είναι πιο περίπλοκος
    - παρατήρησε ότι για μίξεις κατανομών υπάρχει μόνο η αντίστοιχη πιθανότητα $\gamma(\mathbf{z}_n)$

Ας επικεντρωθούμε στον υπολογισμό του $\gamma(z_n)$

Ο αλγόριθμος forward-backward

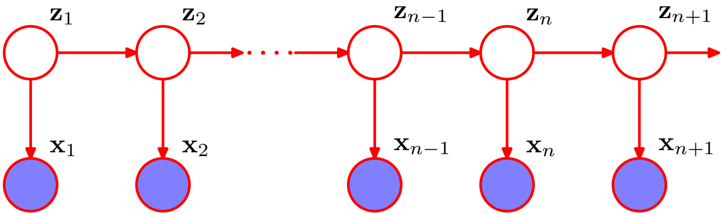


- $\gamma(\mathbf{z}_n) = p(\mathbf{z}_n|X)$, δηλ. **εξαρτάται από τα όλα τα δεδομένα και όχι μόνο από το $\mathbf{x}_n$!**
- Χρησιμοποιώντας τον κανόνα του Bayes και τον κανόνα αθροίσματος των πιθανοτήτων θα μπορούσαμε να υπολογίσουμε το $p(\mathbf{z}_n|X)$ βάσει

$$p(\mathbf{z}_n|X) = \frac{p(X, \mathbf{z}_n)}{p(X)} = \frac{\sum_{\mathbf{z}_1,\ldots,\mathbf{z}_{n-1},\mathbf{z}_{n+1},\ldots,\mathbf{z}_N} p(X, \mathbf{z}_1, \ldots, \mathbf{z}_N)}{p(X)}$$

  το οποίο απαιτεί ένα άθροισμα $O(K^N)$ όρων $\Rightarrow$ μη εφικτό υπολογιστικά
- Η ίδεα του forward-backward αλγόριθμου είναι να πραγματοποιήσει τον παραπάνω υπολογισμό με έξυπνο τρόπο, χρησιμοποιώντας την δομή του γράφου και παίρνωντας χρόνο μόνο $O(K^2 N)$!

Ο αλγόριθμος forward-backward



- **Ιδεά**: Χρησιμοποίησε την ιδιότητα Markov $\Rightarrow$ δοθέντος του παρόντος, το παρελθόν και το μέλλον είναι ανεξάρτητα μεταξύ τους
  - Για το HMM αυτό σημαίνει ότι δοθέντος μιας κρυμμένης μεταβλητής $\mathbf{z}_n$, τα προηγούμενα δεδομένα και τα επόμενα είναι μεταξύ τους ανεξάρτητα
  - Πιο ακριβέστερα ισχύει το εξής

$$
\begin{aligned}
p(X|\mathbf{z}_n) &= p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{z}_n) \\
&= p(\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{z}_n) p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{z}_n)
\end{aligned}
$$

Ο αλγόριθμος forward-backward

- Οπότε έχουμε

$$p(\mathbf{z}_n|X) = \frac{p(X|\mathbf{z}_n)p(\mathbf{z}_n)}{p(X)}$$

Ο αλγόριθμος forward-backward

- Οπότε έχουμε

$$\begin{aligned} p(\mathbf{z}_n|X) &= \frac{p(X|\mathbf{z}_n)p(\mathbf{z}_n)}{p(X)} \\ &= \frac{p(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n)p(\mathbf{z}_n)}{p(X)} \end{aligned}$$

Ο αλγόριθμος forward-backward

- Οπότε έχουμε

$$\begin{aligned} p(\mathbf{z}_n|X) &= \frac{p(X|\mathbf{z}_n)p(\mathbf{z}_n)}{p(X)} \\ &= \frac{p(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n)p(\mathbf{z}_n)}{p(X)} \\ &= \frac{p(\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{z}_n)p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n)p(\mathbf{z}_n)}{p(X)} \end{aligned}$$

Ο αλγόριθμος forward-backward

- Οπότε έχουμε

$$
\begin{aligned}
p(\mathbf{z}_n|X) &= \frac{p(X|\mathbf{z}_n)p(\mathbf{z}_n)}{p(X)} \\
&= \frac{p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{z}_n)p(\mathbf{z}_n)}{p(X)} \\
&= \frac{p(\mathbf{x}_1, \ldots, \mathbf{x}_n|\mathbf{z}_n)p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{z}_n)p(\mathbf{z}_n)}{p(X)} \\
&= \frac{p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_n)p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{z}_n)}{p(X)}
\end{aligned}
$$

Ο αλγόριθμος forward-backward

- Οπότε έχουμε

$$
\begin{aligned}
p(\mathbf{z}_n|X) &= \frac{p(X|\mathbf{z}_n)p(\mathbf{z}_n)}{p(X)} \\
&= \frac{p(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n)p(\mathbf{z}_n)}{p(X)} \\
&= \frac{p(\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{z}_n)p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n)p(\mathbf{z}_n)}{p(X)} \\
&= \frac{p(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{z}_n)p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n)}{p(X)} \\
&= \frac{\alpha(\mathbf{z}_n)\beta(\mathbf{z}_n)}{p(X)}
\end{aligned}
$$

Ο αλγόριθμος forward-backward

$$p(\mathbf{z}_n|X) = \frac{\alpha(\mathbf{z}_n)\beta(\mathbf{z}_n)}{p(X)}$$

όπου

$$\alpha(\mathbf{z}_n) = p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_n)$$

$$\beta(\mathbf{z}_n) = p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{z}_n)$$

Μπορεί να δειχθεί ότι το «μήνυμα» $\alpha(\mathbf{z}_n)$ ικανοποιεί την παρακάτω forward αναδρομή (με αρχικοποίηση $\alpha(\mathbf{z}_1) = p(\mathbf{x}_1|\mathbf{z}_1)p(\mathbf{z}_1)$):

$$\alpha(\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n)\sum_{\mathbf{z}_{n-1}}\alpha(\mathbf{z}_{n-1})p(\mathbf{z}_n|\mathbf{z}_{n-1})$$

και το μήνυμα $\beta(\mathbf{z}_n)$ την backward αναδρομή (με αρχικοποίηση $\beta(\mathbf{z}_N) = 1$)

$$\beta(\mathbf{z}_n) = \sum_{\mathbf{z}_{n+1}}\beta(\mathbf{z}_{n+1})p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n)$$

(Οπότε με ένα forward και ένα backward πέρασμα μπορούμε να υπολογίσουμε όλα τα α και β μηνύματα)

Ο αλγόριθμος forward-backward

Forward αναδρομή (με αρχικοποίηση $\alpha(\mathbf{z}_1) = p(\mathbf{x}_1|\mathbf{z}_1)p(\mathbf{z}_1)$):

$$\alpha(\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} \alpha(\mathbf{z}_{n-1}) p(\mathbf{z}_n|\mathbf{z}_{n-1})$$

Backward αναδρομή (με αρχικοποίηση $\beta(\mathbf{z}_N) = 1$)

$$\beta(\mathbf{z}_n) = \sum_{\mathbf{z}_{n+1}} \beta(\mathbf{z}_{n+1}) p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1}) p(\mathbf{z}_{n+1}|\mathbf{z}_n)$$

Οι δύο αναδρομές είναι ανεξάρτητες μεταξύ τους οπότε θα μπορούσαν να γίνουν παράλληλα. Επίσης σε κάθε βήμα μιας αναδρομής, π.χ. κατά τον υπολογισμό του $\alpha(\mathbf{z}_n)$, απαιτούνται $O(K^2)$ πράξεις. Συγκεκρίμενα, αφού το κάθε $\mathbf{z}_n$ παίρνει K τιμές ο υπολογισμός του $\alpha(\mathbf{z}_n)$ αντιστοιχεί στον υπολογισμό ενός διανύσματος διάστασης K, δηλ. των τιμών:

$$\alpha(z_{nk} = 1) = p(\mathbf{x}_n|z_{nk} = 1) \sum_{j=1}^{K} \alpha(z_{(n-1)j} = 1) p(z_{nk} = 1|z_{(n-1)j} = 1), \quad k = 1, \ldots, K$$

το οποίο απαιτεί $O(K^2)$ πράξεις. Επίσης η κάθε αναδρομή είναι γραμμική ως προς τον αριθμό των δεδομένων N, οπότε η συνολική πολυπλοκότητα χρόνου της forward (και το ίδιο και για την backward) αναδρομή είναι $O(K^2 N)$

Ο αλγόριθμος forward-backward

Έχοντας υπολογίζει τα $\alpha(\mathbf{z}_n)$ και $\beta(\mathbf{z}_n)$ για $n = 1, \ldots, N$ (τα οποία στην υλοποίηση στον υπολογιστή μπορούν να αποθηκευτούν σε δύο επιμέρους πίνακες διάστασης $N \times K$) το E βήμα του EM ολοκληρώνεται με τις πράξεις

$$\gamma(\mathbf{z}_n) = \frac{\alpha(\mathbf{z}_n)\beta(\mathbf{z}_n)}{p(X)}$$

(δηλ. υπολογισμό των τιμών $\gamma(z_{nk} = 1) = \frac{\alpha(z_{nk}=1)\beta(z_{nk}=1)}{p(X)}, \ k = 1, \ldots, K$)

για κάθε n και όπου

$$p(X) = \sum_{\mathbf{z}_N} \alpha(\mathbf{z}_N) = \sum_{\mathbf{z}_n} \alpha(\mathbf{z}_n)\beta(\mathbf{z}_n), \quad n = 1, \ldots, N-1$$

είναι η πιθανοφάνεια των δεδομένων και η οποία μπορεί να υπολογιστεί με N ισοδύναμους τρόπους (πολύ χρήσιμη ιδιότητα για de-bugging του κώδικα υπολογισμού των α και β μηνυμάτων!)

Ομοίως μπορεί εύκολα να δειχθεί ότι

$$\xi(\mathbf{z}_{n-1}, \mathbf{z}_n) = \frac{\alpha(\mathbf{z}_{n-1})p(\mathbf{x}_n|\mathbf{z}_n)p(\mathbf{z}_n|\mathbf{z}_{n-1})\beta(\mathbf{z}_n)}{p(X)}$$

Παράδειγμα υπολογισμού των α και β μηνυμάτων

Έστω ένα HMM ως προς 5 δεδομένα $X = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5)$

Η από κοινού κατανομή των δεδομένων $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5)$ και των κρυμμένων μεταβλητών $(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4, \mathbf{z}_5)$ είναι

$$p(\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)p(\mathbf{z}_3|\mathbf{z}_2)p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_1|\mathbf{z}_1)p(\mathbf{x}_2|\mathbf{z}_2)p(\mathbf{x}_3|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)$$

Παράδειγμα υπολογισμού των α και β μηνυμάτων

Έστω ένα HMM ως προς 5 δεδομένα $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5)$

Η από κοινού κατανομή των δεδομένων $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5)$ και των κρυμμένων μεταβλητών $(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4, \mathbf{z}_5)$ είναι

$$p(\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)p(\mathbf{z}_3|\mathbf{z}_2)p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_1|\mathbf{z}_1)p(\mathbf{x}_2|\mathbf{z}_2)p(\mathbf{x}_3|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)$$

ή

$$p(\mathbf{z}_1)p(\mathbf{x}_1|\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)p(\mathbf{x}_2|\mathbf{z}_2)p(\mathbf{z}_3|\mathbf{z}_2)p(\mathbf{x}_3|\mathbf{z}_3)p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)$$

Παράδειγμα υπολογισμού των α και β μηνυμάτων

$$p(\mathbf{z}_1)p(\mathbf{x}_1|\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)p(\mathbf{x}_2|\mathbf{z}_2)p(\mathbf{z}_3|\mathbf{z}_2)p(\mathbf{x}_3|\mathbf{z}_3)p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)$$

Έστω τώρα ότι θα θέλαμε να υπολογίσουμε την εκ των υστέρων πιθανότητα $p(\mathbf{z}_3|X)$ η οποία βάσει του κανόνα του Bayes δίνεται από

$$p(\mathbf{z}_3|X) = \frac{p(\mathbf{z}_3, X)}{p(X)} = \frac{p(\mathbf{z}_3, X)}{\sum_{\mathbf{z}_3} p(\mathbf{z}_3, X)}$$

Είναι προφανές ότι για να υπολογίσουμε την ποσότητα αυτή αρκεί να υπολογίσουμε την περιθώρια κατανομή $p(\mathbf{z}_3, X)$. Χρησιμοποιώντας τον κανόνα αθροίσματος η $p(\mathbf{z}_3, X)$ ισούται με

$$\sum_{\mathbf{z}_1,\mathbf{z}_2,\mathbf{z}_4,\mathbf{z}_5} p(\mathbf{z}_1)p(\mathbf{x}_1|\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)p(\mathbf{x}_2|\mathbf{z}_2)p(\mathbf{z}_3|\mathbf{z}_2)p(\mathbf{x}_3|\mathbf{z}_3)p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)$$

Ο forward-backward αλγόριθμος υπολογίζει αυτό το άθροισμα με μια ορισμένη σείρα ώστε ο συνολικός αριθμός των υπολογισμών να είναι ο ελάχιστος

Παράδειγμα υπολογισμού των α και β μηνυμάτων

Θέλουμε να υπολογίσουμε

$$\sum_{\mathbf{z}_1,\mathbf{z}_2,\mathbf{z}_4,\mathbf{z}_5} p(\mathbf{z}_1)p(\mathbf{x}_1|\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)p(\mathbf{x}_2|\mathbf{z}_2)p(\mathbf{z}_3|\mathbf{z}_2)p(\mathbf{x}_3|\mathbf{z}_3)p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)$$

Μπορούμε να κατανείμουμε τους υπολογισμούς ως εξής

$$\left(\sum_{\mathbf{z}_2}\left(\sum_{\mathbf{z}_1} p(\mathbf{z}_1)p(\mathbf{x}_1|\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)\right) p(\mathbf{x}_2|\mathbf{z}_2)p(\mathbf{z}_3|\mathbf{z}_2)\right) p(\mathbf{x}_3|\mathbf{z}_3)$$
$$\times\left(\sum_{\mathbf{z}_4} p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)\left(\sum_{\mathbf{z}_5} p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)\right)\right)$$

Όπου τα εσωτερικά αθροίσματα θα υπολογιστούν πρώτα και έπειτα τα εξωτερικά

Παράδειγμα υπολογισμού των α και β μηνυμάτων

$$\left(\sum_{\mathbf{z}_2}\left(\sum_{\mathbf{z}_1} p(\mathbf{z}_1)p(\mathbf{x}_1|\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)\right)p(\mathbf{x}_2|\mathbf{z}_2)p(\mathbf{z}_3|\mathbf{z}_2)\right)p(\mathbf{x}_3|\mathbf{z}_3)$$
$$\times\left(\sum_{\mathbf{z}_4} p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)\left(\sum_{\mathbf{z}_5} p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)\right)\right)$$

Θέτοντας (και υπολογίζοντας πρώτα) $\alpha(\mathbf{z}_1) = p(\mathbf{z}_1)p(\mathbf{x}_1|\mathbf{z}_1)$ έχουμε

$$\left(\sum_{\mathbf{z}_2}\left(\sum_{\mathbf{z}_1} \alpha(\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)\right)p(\mathbf{x}_2|\mathbf{z}_2)p(\mathbf{z}_3|\mathbf{z}_2)\right)p(\mathbf{x}_3|\mathbf{z}_3)$$
$$\times\left(\sum_{\mathbf{z}_4} p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)\left(\sum_{\mathbf{z}_5} p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)\right)\right)$$

Παράδειγμα υπολογισμού των α και β μηνυμάτων

$$\left(\sum_{\mathbf{z}_2}\left(\sum_{\mathbf{z}_1} p(\mathbf{z}_1)p(\mathbf{x}_1|\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)\right)p(\mathbf{x}_2|\mathbf{z}_2)p(\mathbf{z}_3|\mathbf{z}_2)\right)p(\mathbf{x}_3|\mathbf{z}_3)$$
$$\times\left(\sum_{\mathbf{z}_4} p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)\left(\sum_{\mathbf{z}_5} p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)\right)\right)$$

Θέτοντας (και υπολογίζοντας πρώτα) $\alpha(\mathbf{z}_1) = p(\mathbf{z}_1)p(\mathbf{x}_1|\mathbf{z}_1)$ έχουμε

$$\left(\sum_{\mathbf{z}_2}\left(\sum_{\mathbf{z}_1} \alpha(\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)\right)p(\mathbf{x}_2|\mathbf{z}_2)p(\mathbf{z}_3|\mathbf{z}_2)\right)p(\mathbf{x}_3|\mathbf{z}_3)$$
$$\times\left(\sum_{\mathbf{z}_4} p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)\left(\sum_{\mathbf{z}_5} p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)\right)\right)$$

2ή φάση: $\alpha(\mathbf{z}_2) = \left(\sum_{\mathbf{z}_1} \alpha(\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)\right)p(\mathbf{x}_2|\mathbf{z}_2)$ και έχουμε

$$\left(\sum_{\mathbf{z}_2} \alpha(\mathbf{z}_2)p(\mathbf{z}_3|\mathbf{z}_2)\right)p(\mathbf{x}_3|\mathbf{z}_3)\left(\sum_{\mathbf{z}_4} p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)\left(\sum_{\mathbf{z}_5} p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)\right)\right)$$

Παράδειγμα υπολογισμού των α και β μηνυμάτων

$$\left(\sum_{\mathbf{z}_2}\left(\sum_{\mathbf{z}_1} p(\mathbf{z}_1)p(\mathbf{x}_1|\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)\right)p(\mathbf{x}_2|\mathbf{z}_2)p(\mathbf{z}_3|\mathbf{z}_2)\right)p(\mathbf{x}_3|\mathbf{z}_3)$$
$$\times\left(\sum_{\mathbf{z}_4} p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)\left(\sum_{\mathbf{z}_5} p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)\right)\right)$$

Θέτοντας (και υπολογίζοντας πρώτα) $\alpha(\mathbf{z}_1) = p(\mathbf{z}_1)p(\mathbf{x}_1|\mathbf{z}_1)$ έχουμε

$$\left(\sum_{\mathbf{z}_2}\left(\sum_{\mathbf{z}_1} \alpha(\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)\right)p(\mathbf{x}_2|\mathbf{z}_2)p(\mathbf{z}_3|\mathbf{z}_2)\right)p(\mathbf{x}_3|\mathbf{z}_3)$$
$$\times\left(\sum_{\mathbf{z}_4} p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)\left(\sum_{\mathbf{z}_5} p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)\right)\right)$$

2ή φάση: $\alpha(\mathbf{z}_2) = \left(\sum_{\mathbf{z}_1} \alpha(\mathbf{z}_1)p(\mathbf{z}_2|\mathbf{z}_1)\right)p(\mathbf{x}_2|\mathbf{z}_2)$ και έχουμε

$$\left(\sum_{\mathbf{z}_2} \alpha(\mathbf{z}_2)p(\mathbf{z}_3|\mathbf{z}_2)\right)p(\mathbf{x}_3|\mathbf{z}_3)\left(\sum_{\mathbf{z}_4} p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)\left(\sum_{\mathbf{z}_5} p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)\right)\right)$$

3ή φάση: $\alpha(\mathbf{z}_3) = \left(\sum_{\mathbf{z}_2} \alpha(\mathbf{z}_2)p(\mathbf{z}_3|\mathbf{z}_2)\right)p(\mathbf{x}_3|\mathbf{z}_3)$ και έχουμε

$$\alpha(\mathbf{z}_3)\left(\sum_{\mathbf{z}_4} p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)\left(\sum_{\mathbf{z}_5} p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)\right)\right)$$

Παράδειγμα υπολογισμού των α και β μηνυμάτων

$$\alpha(\mathbf{z}_3)\left(\sum_{\mathbf{z}_4} p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)\left(\sum_{\mathbf{z}_5} p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)\right)\right)$$

Δουλεύοντας τώρα backwards, υπολογίζουμε $\beta(\mathbf{z}_4) = \sum_{\mathbf{z}_5} p(\mathbf{z}_5|\mathbf{z}_4)p(\mathbf{x}_5|\mathbf{z}_5)\beta(\mathbf{z}_5)$ (όπου $\beta(\mathbf{z}_5)$ τέτοιο ώστε $\beta(\mathbf{z}_5) = 1$ για κάθε τιμή του $\mathbf{z}_5$!)

$$\alpha(\mathbf{z}_3)\left(\sum_{\mathbf{z}_4} p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)\beta(\mathbf{z}_4)\right)$$

Έπειτα υπολογίζουμε $\beta(\mathbf{z}_3) = \sum_{\mathbf{z}_4} p(\mathbf{z}_4|\mathbf{z}_3)p(\mathbf{x}_4|\mathbf{z}_4)\beta(\mathbf{z}_4)$

$$\alpha(\mathbf{z}_3)\beta(\mathbf{z}_3)$$

το οποίο ισούται με την ποσότητα $p(\mathbf{z}_3, X)$, δηλ.

$$p(\mathbf{z}_3, X) = \alpha(\mathbf{z}_3)\beta(\mathbf{z}_3)$$

Ο αλγόριθμος Viterbi

Ένας επιπλέον αλγόριθμος που χρησιμοποιείται στα HMMs (αφού πρώτα έχει γίνει εκπαίδευση με τον EM) είναι ο Viterbi που βρίσκει εκείνο το μοναπάτι των κρυμμένων μεταβλητών Z^* που έχει την μεγαλύτερη εκ των υστέρων πιθανότητα

$$Z^* = \arg\max_Z p(Z|X) = \arg\max_Z p(X,Z) = \arg\max_Z \log p(X,Z)$$

Ο δυναμικός αλγόριθμος που χρησιποποιείται είναι σχεδόν πανομοιότυπος με το forward α πέρασμα

- Η διαφορά έγκειται στο ότι ο τελεστής αθροίσματος $\sum$ αντικαθίσταται από τον τελεστή μεγίστου max

Ο αλγόριθμος Viterbi

- Συγκεκριμένα ο αλγόριθμος Viterbi προωθεί ένα μήνυμα το οποίο αρχικοποιείται ως

$$\omega(\mathbf{z}_1) = \log p(\mathbf{x}_1|z_1) + \log p(\mathbf{z}_1)$$

  και υπολογίζεται αναδρομικά βάσει

$$\omega(\mathbf{z}_n) = \log p(\mathbf{x}_n|\mathbf{z}_n) + \max_{\mathbf{z}_{n-1}} [\omega(\mathbf{z}_{n-1}) + \log p(\mathbf{z}_n|\mathbf{z}_{n-1})]$$

$$\delta(\mathbf{z}_n) = \mathbf{z}^*_{n-1}$$

  όπου $\delta(\mathbf{z}_n)$ είναι ένα βοηθητικό μήνυμα που απλώς αποθηκεύει την τιμή του $\mathbf{z}_{n-1}$ για την οποία επιτυγχάνεται το μέγιστο

- Όταν το forward πέρασμα ολοκληρωθεί, έχουμε υπολογίζει και το τελευταίο μήνυμα $\omega(\mathbf{z}_N)$ (καθώς και όλα τα βοηθητικά δ μηνύματα). Έπειτα υπολογίζουμε

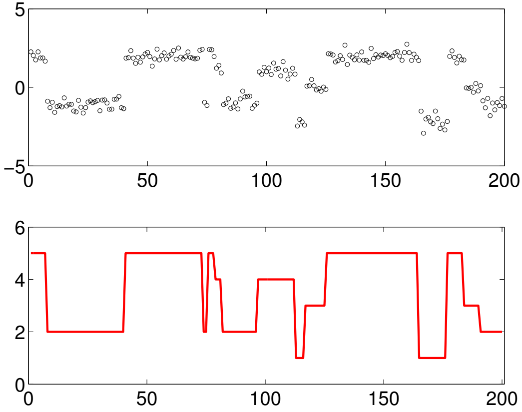
$$\mathbf{z}^*_N = \arg\max_{\mathbf{z}_N} [\omega(\mathbf{z}_N)]$$

  και ξεκινώντας από $\mathbf{z}^*_N$ οπισθοδρομούμε αναδρομικά (backtrack) βάσει

$$\mathbf{z}^*_{n-1} \leftarrow \delta(\mathbf{z}^*_n)$$

  που βρίσκει το βέλτιστο μονοπάτι $Z^* = (\mathbf{z}^*_1, \ldots, \mathbf{z}^*_N)$

Παράδειγμα εφαρμογής του EM αλγορίθμου και εύρεσης βέλτιστου μονοπατιού



Έστω τα δεδομένα του σχήματος (πάνω εικόνα) όπου έχουν παραχθεί με τεχνητό τρόπο από $K = 5$ καταστάσεις. Το μονοπάτι που έχει παράγει τα δεδομένα φαίνεται με την κόκκινη γραμμή (κάτω εικόνα)

Παράδειγμα εφαρμογής του EM αλγορίθμου και εύρεσης βέλτιστου μονοπατιού

Επίσης προφανώς γνωρίζουμε και τις παραμέτρους που έχουν παράγει τα δεδομένα

- Αρχική κατανομή: $\pi_k = \frac{1}{5},\ k = 1, \ldots, K$
- Πίνακας μετάβασης:

$$A = \begin{vmatrix} 0.9500 & 0.0125 & 0.0125 & 0.0125 & 0.0125 \\ 0.0125 & 0.9500 & 0.0125 & 0.0125 & 0.0125 \\ 0.0125 & 0.0125 & 0.9500 & 0.0125 & 0.0125 \\ 0.0125 & 0.0125 & 0.0125 & 0.9500 & 0.0125 \\ 0.0125 & 0.0125 & 0.0125 & 0.0125 & 0.9500 \end{vmatrix}$$

- Gaussian κατανομές εκπομπής:

$$p(x|k) = \mathcal{N}(x|\mu_k, \sigma_k^2), \quad k = 1, \ldots, 5$$

$$\mu_1 = -2, \mu_2 = -1, \mu_3 = 0, \mu_4 = 1, \mu_5 = 2$$

$$\sigma_k^2 = 0.09, \quad k = 1, \ldots, 5$$

Παράδειγμα εφαρμογής του EM αλγορίθμου και εύρεσης βέλτιστου μονοπατιού

Εφαρμόζουμε EM (υποθέτοντας ότι δεν γνωρίζουμε τις παραμέτρους και το γεγονός ότι τα δεδομένα είναι τεχνητώς κατασκευασμένα!)

- Εκτίμηση αρχικής κατανομής:
  $(\pi_1 = 0, \pi_2 = 0, \pi_3 = 0, \pi_4 = 0, \pi_5 = 1)$
- Εκτίμηση πίνακα μετάβασης:

$$A = \begin{vmatrix} 0.8750 & 0.0000 & 0.0625 & 0 & 0.0625 \\ 0.0000 & 0.9491 & 0.0000 & 0.0169 & 0.0339 \\ 0 & 0.0625 & 0.8750 & 0 & 0.0625 \\ 0.0531 & 0.0530 & 0.0000 & 0.8940 & 0.0000 \\ 0.0111 & 0.0224 & 0.0112 & 0.0114 & 0.9438 \end{vmatrix}$$

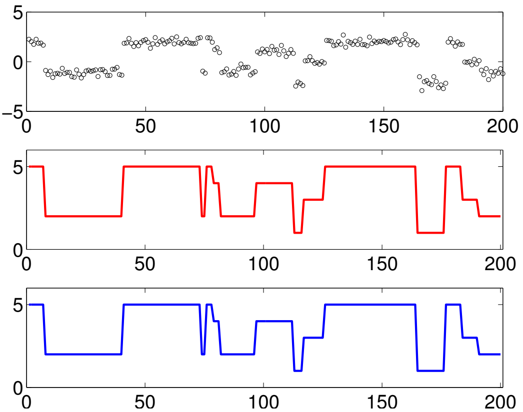
- Εκτίμηση Gaussian κατανομών εκπομπής:

$$p(x|k) = \mathcal{N}(x|\mu_k, \sigma_k^2), \quad k = 1, \dots, 5$$

$$\mu_1 = -2.2050, \mu_2 = -1.0619, \mu_3 = -0.0063, \mu_4 = 1.0793, \mu_5 = 1.9874$$

$$\sigma_1^2 = 0.1387, \sigma_2^2 = 0.0965, \sigma_3^2 = 0.0385, \sigma_4^2 = 0.0777, \sigma_5^2 = 0.0720$$

Παράδειγμα εφαρμογής του EM αλγορίθμου και εύρεσης βέλτιστου μοναπατιού



Επίσης με τον αλγόριθμο Viterbi μπορούμε να εκτιμήσουμε το βέλτιστο μονοπάτι (δηλ. τις τιμές των κρυμμένων μεταβλήτων $\mathbf{z}_1, \dots, \mathbf{z}_N$ που εξηγούν τα δεδομένα με την μεγαλύτερη εκ των υστέρων πιθανότητα) το οποίο φαίνεται με το μπλέ χρώμα στην τελευταία εικόνα

Παρατηρούμε ότι στην προκειμένη περίπτωση έχουμε προσεγγίσει τέλεια την αλήθεια (δηλ. το κόκκινο μονοπάτι)

Επίλογος

- Διάβασμα για το σπίτι: Bishop: κεφάλαιο 13