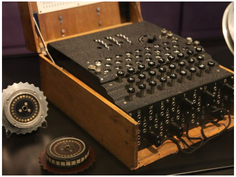
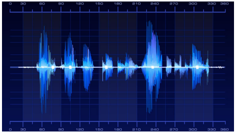
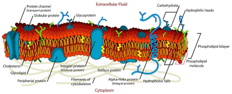



0 30 60 90 120 150 180 210 240 270 300 330 360


Extracellular Fluid
Protein channel
(transport protein)
Globular protein
Glycoprotein
Carbohydrate
Hydrophilic heads
Phospholipid bilayer
Phospholipid molecule
Cholesterol
Glycolipid
Integral protein
(Globular protein)
Surface protein
Peripherial protein
Filaments of cytoskeleton
Alpha-Helix protein
(Integral protein)
Hydrophobic tails
Cytoplasm

Skriti markovski modeli v časovnih vrstah

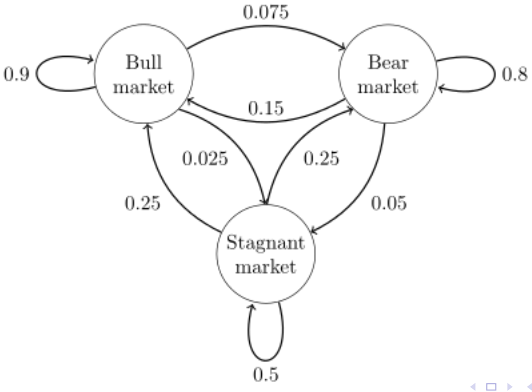
Fakulteta za matematiko in fiziko, Univerza v Ljubljani

10. december 2018

Martin Praček
Mentor: izr. prof. dr. Damjan Škulj

Markovski model

- Markovska lastnost je lastnost slučajnega procesa, da je njegova vrednost v času t odvisna le od njegove vrednosti v času $t-1$.
- Ločimo v celoti opazovan in delno opazovan ter
- Avtonomen in kotroliran sistem.

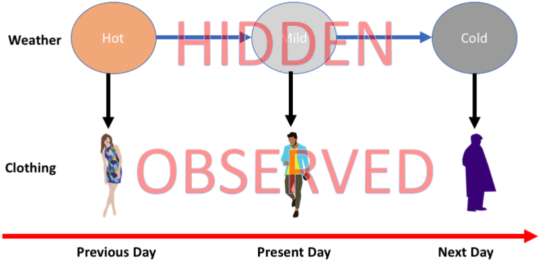


Delitev markovskih modelov

	V celoti opazovan	Le delno opazovan
Avtonomen	Markovska veriga	Skriti markovski model
Kontorliran	Markovski proces odločanja	Delno opazovalen proces odločanja

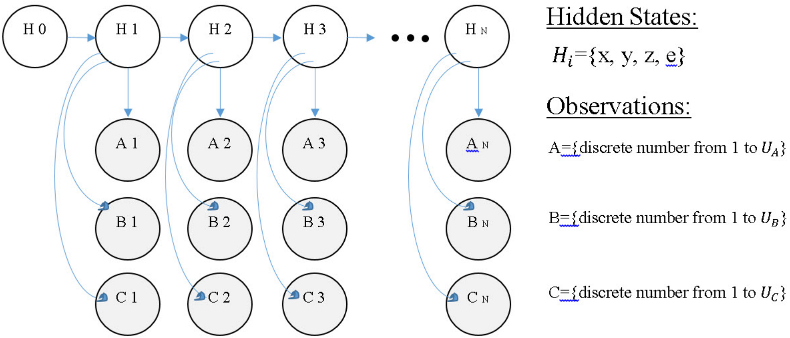
Skriti markovski model

Skriti markovski model je statistični markovski model, kjer predpostavljamo, da je modelirani sistem markovski proces z skritimi stanji.
Gre torej za tip modela, kjer lahko razberemo rezultat, ne moremo pa ugotoviti, kakšna je bila funkcija, ki nam ga je dala.



Zahteve

- Markovska lastnost
- Enaki razdeljeni časi signalov
- Sistem ima N stanj, vsako določa slučajna spremenljivka S



- Slučajnih spremenljivk v nobenem času ne poznamo, poznamo pa slučajni proces Q, ki predstavlja signale
- Porazdelitveno funkcijo vsakega stanja i označimo z $b_i(x)$
- Vektor začetnih stanj je Π
- Prehodna matrika A, ki je neodvisna od časa

Porazdelitvena funkcija

- Gaussova mešanica
- $b_i = \sum_{j=1}^{M} c_{ij} N(x; \mu_j, \sigma_j^2)$
- Število mešanic M
- Matrika Γ, μ_{ij} predstavlja pričakovano vrednost mešanice j v stanju i
- Matrika Σ, kjer σ_{ij} predstavlja varianco mešanice j v stanju i
- Matrika C, koeficienti c_{ij} iz Gaussove mešanice

Osnovanje modela

Da bomo lahko delali z našim modelom, moramo najprej izvesti t.i. trening modela.

- šop k-povprečij
- Akaikov informacijski kriterij

Izračun $P(O|\lambda)$

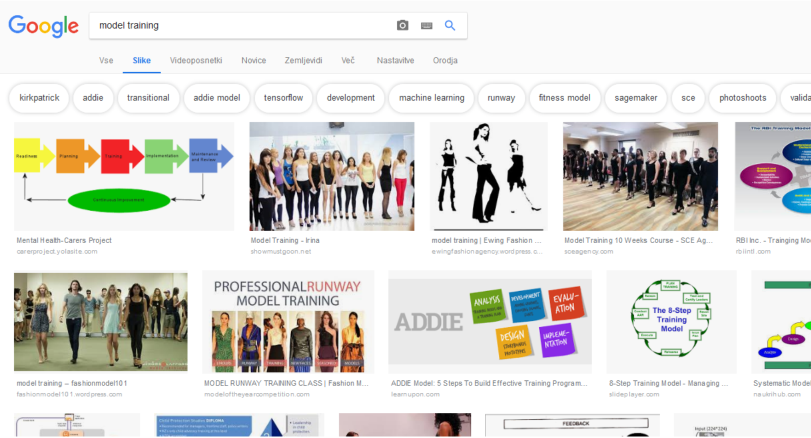
Z λ označimo vse ostale parametre $\lambda = (\Pi, A, C, \Gamma, \Sigma)$. Za učinkovit izračun le teh delamo *naprej* $-$ *nazaj*.

$$\alpha_1(i) = \pi_i b_i(O_1) = \pi_i b_i = \sum_{k=1}^{M} c_{ik}$$

$$\Downarrow$$

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

Trening modela

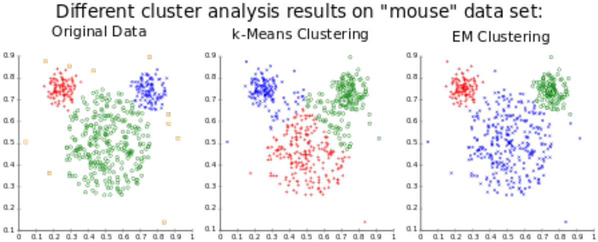


Trening modela

Da bomo lahko $P(O|\lambda)$ maksimizirali, potrebujemo začetne ocene parametrov. Za to ne poznamo analitičnega postopka, lahko pa lokalno maksimiziramo, z naprimer, Baum-Welchovim algoritmom. Pri tem nas ne skrbijo ocene za A ter Π, kjer moramo paziti le na neničelnost le teh. Več problemov nam povzročajo C, Σ in Γ.

C, Σ in Γ

Za dobro začetno oceno si lahko pomagamo z šopom k-povprečij. Za C to pomeni, da bo element $c_{ij} = 1/k$ za vsak par ij, kjer je k število šopov povprečij. Pričakovane vrednosti in variance nato pridobimo iz vrednosti teh šopov povprečij.



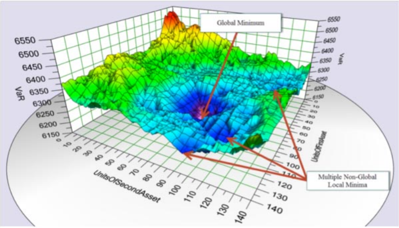
Lokalno maksimiziranje

Za potrebe lokalne maksimizacije določimo dodatne funkcije:

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_t) \beta_{t+1}(j)}{P(O|\lambda)}$$

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)}$$

$$\gamma_t(j,k) = \gamma_t(j) \frac{c_{jk} N(x; \mu_{jk}, \sigma_{jk}^2)}{\sum_{m=1}^{M} c_{jm} N(x; \mu_{jm}, \sigma_{jm}^2)}$$



Lokalno maksimiziranje

Prek dodatnih funkcij definiramo iteracijske postopke za naše spremenljivke:

$$\overline{\Pi_i} = \gamma_1(i)$$

$$\overline{a_{ij}} = \frac{\sum_{t=1}^{T-1} \xi_t}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

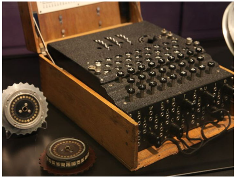
$$\overline{c_{jk}} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)}{\sum_{t=1}^{T} \sum_{m=1}^{M} \gamma_t(j,m)}$$

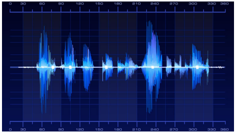
$$\overline{\mu_{jk}} = \frac{\sum_{t=1}^{T} \gamma_t(j,k) O_t}{\sum_{t=1}^{T} \gamma_t(j,k)}$$

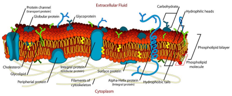
Začetno stanje in Viterbijev algoritem

Za delo s tem algoritmom moramo definirati $\delta_t(i)$, ki za vsako stanje i vrne največjo verjetnost vzdolž poti v času t. Prek $\delta_t(i)$ nato induktivno izvedemo algoritem.
Viterbijev algoritem nam vrne $p*$, ki je največja verjetnost in q_T*, ki nam pove stanje v času T, ki nam to verjetnost vrne.

Uporaba




0 30 60 90 120 150 180 210 240 270 300 330 360


Extracellular Fluid
Protein channel
(transport protein)
Globular protein
Glycoprotein
Carbohydrate
Hydrophilic heads
Phospholipid bilayer
Phospholipid
molecule
Cholesterol
Glycolipid
Peripherial protein
Integral protein
(Globular protein)
Surface protein
Filaments of
cytoskeleton
Alpha-Helix protein
(Integral protein)
Hydrophobic tails
Cytoplasm

V dolgi predstavitvi se bom bolj posvetil sami finančni analizi, ki sem jo tokrat zaenkrat pustil pri miru. V prihodnje bom tudi sam poizkusil določiti skriti markovski model na svojem setu podatkov.