

Les 11 Markov processen en Markov modellen

We hebben in het kader van de Lineaire Algebra in Wiskunde 1 een aantal voorbeelden van systemen gezien, die zich door overgangsmatrices laten beschrijven. Voorbeelden hiervan waren:

- Populaties die zich volgens overgangen tussen de verschillende generaties ontwikkelen.
- De verspreiding van de Euro munten over de verschillende landen.

Iets algemener gesproken hebben we het hierbij over systemen, die gekarakteriseerd zijn door: (i) mogelijke *toestanden* van het systeem; en (ii) *overgangen* tussen deze toestanden.

We zullen het in deze (en de volgende) les over dit soort systemen hebben, waarbij we vooral naar het belangrijke geval kijken, dat de overgangen door kansverdelingen worden beschreven.

11.1 Markov processen

Als de overgangen tussen de mogelijke toestanden van een systeem door kansverdelingen gegeven zijn, spreekt men meestal van *Markov processen*. Een Markov proces wordt door de volgende gegevens gekarakteriseerd:

- een aantal mogelijke toestanden S_1, S_2, \dots, S_N , die we *states* noemen;
- op elke tijdstip $t = 0, 1, 2, \dots$ een state $q_t \in \{S_1, \dots, S_N\}$, waarin het systeem zich op dit tijdstip bevindt;
- gegeven de states q_0, q_1, \dots, q_{t-1} op de tijdstippen $0, 1, \dots, t-1$, de kansverdeling dat het systeem op tijdstip t in de state S_j terecht komt, d.w.z. de voorwaardelijke kansen

$$p(q_t = S_j \mid q_{t-1} = S_{i_{t-1}}, \dots, q_1 = S_{i_1}, q_0 = S_{i_0}).$$

Het probleem zit in de exponentiële groei van het aantal mogelijke states op de tijdstippen $0, 1, \dots, t-1$, dit zijn er namelijk N^t . Voor elke van deze mogelijkheden moeten we in principe een aparte kansverdeling bepalen, maar dat is natuurlijk voor grotere waarden van t ondoenlijk.

Als we nog eens naar het voorbeeld van de taalherkenning middels letterfrequenties kijken, kunnen we dit zien als een Markov proces waarbij de states de verschillende letters zijn. In dit geval zouden voor elk beginstuk van t letters een kansverdeling voor de daarop volgende letter moeten bepalen. Voor een beginstuk van 8 letters zijn dit bijvoorbeeld $27^8 = 282429536481$ verdelingen, en die kunnen we nog bepalen nog opslaan.

Maar dit voorbeeld wijst ook al een mogelijke oplossing aan: We kunnen ervan uitgaan dat de kansverdeling voor de 9-de letter niet erg verandert als we de eerste letter q_0 veranderen en waarschijnlijk zal ook de letter op tijdstip $t = 1$ nog geen grote betekenis voor q_8 hebben. Dit leidt tot het idee, de kansverdeling

voor de states op tijdstip t te benaderen door de kansverdeling die alleen maar met de k voorafgaande states rekening houdt, d.w.z.:

$$p(q_t = S_j \mid q_{t-1} = S_{i_{t-1}}, \dots, q_0 = S_{i_0}) \approx p(q_t = S_j \mid q_{t-1} = S_{i_{t-1}}, \dots, q_{t-k} = S_{i_{t-k}}).$$

De rechterzijde van deze vergelijking geeft de kansverdeling van een Markov proces, waarbij de kans op de state op tijdstip t alleen maar van de k voorafgaande states afhangt, en dit noemt men een *Markov proces van orde k* . Merk op dat we hierbij ook nog veronderstellen dat de kansen niet van het tijdstip t afhangen, maar alleen maar van de rij voorafgaande states.

Bij een **Markov proces van orde 0** speelt de geschiedenis helemaal geen rol, de states worden alleen maar volgens een kansverdeling op de states voortgebracht. Zo'n Markov proces krijgen we bijvoorbeeld, als we (zo als in de laatste les) alleen maar de relatieve frequenties van de letters in een taal bepalen en vervolgens letters volgens deze kansverdeling produceren. De relatieve frequentie van de letters zal dan wel kloppen, maar bijvoorbeeld de relatieve frequenties van paren van letters niet meer. Hiervoor hebben we een Markov proces van orde 1 nodig.

Een **Markov proces van orde 1** is gekarakteriseerd door de overgangskansen $a_{ij} := p(q_t = S_j \mid q_{t-1} = S_i)$. Omdat we veronderstellen dat deze kansen onafhankelijk van t zijn, kunnen we de kansen in een *overgangsmatrix* $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ invullen. Voor deze overgangsmatrix A geldt dat $a_{ij} \geq 0$ en dat $\sum_{j=1}^n a_{ij} = 1$ voor alle $i = 1, \dots, n$, omdat het systeem vanuit state S_i naar een van de states S_j moet overgaan.

Een handige eigenschap van de overgangsmatrix A is dat we met behulp van de machten van A makkelijk kunnen berekenen wat er over een aantal stappen gebeurt: Het element (i, j) van A^k geeft de kans aan, dat het systeem in (precies) k stappen van state S_i naar S_j gaat.

Een Markov proces van orde 1 laat zich ook overzichtelijk door een *graaf* of *state diagram* representeren: De states zijn punten en de overgangen zijn pijltjes tussen de states, met de kans voor de overgangen als labels.

Als we bijvoorbeeld het weer als een (eenvoudige) Markov proces willen beschrijven, zouden we misschien de drie states

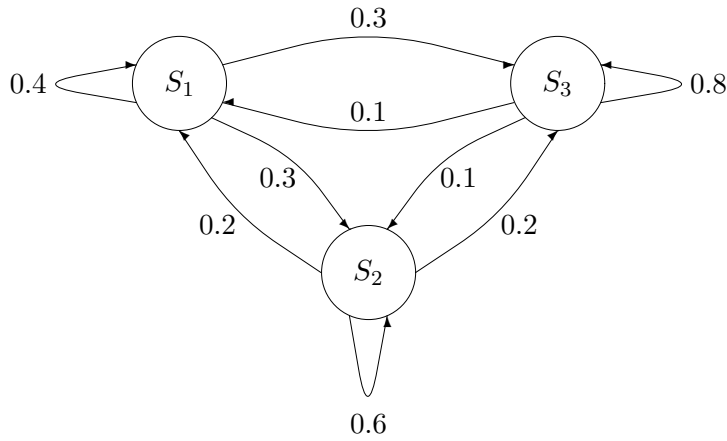
$$S_1 = \text{regen}, \quad S_2 = \text{bewolkt}, \quad S_3 = \text{zonnig}$$

kunnen kiezen. Als overgangsmatrix veronderstellen we de (erg optimistische) matrix

$$A = \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}.$$

Dan heeft deze Markov proces het state diagram uit Figuur II.2.

Aan de hand van de overgangskansen kunnen we een aantal vragen makkelijk beantwoorden:



Figuur II.2: Markov proces voor het weer.

- (1) Wat is de kans op drie dagen zon gevolgd van een dag regen?
- (2) Wat is de kans dat het weer precies d dagen hetzelfde blijft?

Bij vraag (1) willen we de kans op de rij $O = S_3 S_3 S_3 S_1$ van states weten. Maar de overgang $S_3 \rightarrow S_3$ heeft kans $a_{33} = 0.8$ en de overgang $S_3 \rightarrow S_1$ heeft kans $a_{31} = 0.1$, dus is de kans op deze rij $0.8 \cdot 0.8 \cdot 0.1 = 0.064$. Hierbij veronderstellen we wel, dat de vraag op een dag gesteld wordt, waar het al zonnig is, dus waar we al in state S_3 zitten.

Vraag (2) gaat over een rij $O = \underbrace{S_i S_i \dots S_i}_d S_j$ van states, waarbij er precies d keer de state S_i voorkomt en de state S_j verschillend is van S_i . Maar de kans dat we van state S_i naar een state verschillend van S_i gaan is $1 - a_{ii}$, dus is de kans $p(O)$ van deze rij states $p(O) = a_{ii}^{d-1} \cdot (1 - a_{ii})$.

We kunnen nu zelfs de verwachtingswaarde voor het aantal dagen d die we in state S_i blijven, berekenen, er geldt:

$$E[d] = \sum_{d=1}^{\infty} d \cdot a_{ii}^{d-1} \cdot (1 - a_{ii}) = \frac{1}{1 - a_{ii}}.$$

Dit zien we als volgt in: Voor de meetkundige reeks geldt $\sum_{d=0}^{\infty} x^d = \frac{1}{1-x}$ als $|x| < 1$. Maar $\sum_{d=1}^{\infty} dx^{d-1} = (\sum_{d=0}^{\infty} x^d)'$, omdat we in dit geval termsgewijs mogen afleiden. Aan de andere kant is $(\frac{1}{1-x})' = \frac{1}{(1-x)^2}$, en dus is $\sum_{d=1}^{\infty} d \cdot a_{ii}^{d-1} \cdot (1 - a_{ii}) = \frac{1}{(1-a_{ii})^2} (1 - a_{ii}) = \frac{1}{1-a_{ii}}$.

In ons optimistisch model van het weer is de kans dat het blijft regenen $a_{11} = 0.4$, dus is de verwachtingswaarde voor het aantal regendagen achter elkaar gelijk aan $\frac{1}{1-0.4} = \frac{1}{0.6} \approx 1.67$. Net zo krijgen we voor het verwachte aantal bewolkte dagen achter elkaar de waarde $\frac{1}{1-0.6} = 2.5$ en voor het aantal zonnige dagen achter elkaar hebben we de verwachtingswaarde $\frac{1}{1-0.8} = 5$.

11.2 Stochastische automaten

Bij de Markov processen zijn we ervan uitgegaan dat het systeem op tijdstippen $t = 0, 1, \dots$ van een state naar een andere state overgaat. In sommige samenhangen wordt zo'n overgang veroorzaakt door een input aan het systeem. Maar dan is het plausibel dat de overgangskansen ook van de mogelijke inputs kunnen afhangen. Dit betekent, dat er bij een Markov proces van orde 1 voor elke mogelijke input een aparte overgangsmatrix is. Zo'n systeem noemt men ook een *stochastische automaat*.

We bekijken dit aan de hand van het voorbeeld van een *emotionele robot*. Stel de robot heeft drie mogelijke states, namelijk

$$S_1 = \text{gelukkig}, \quad S_2 = \text{bedroefd}, \quad S_3 = \text{mal}$$

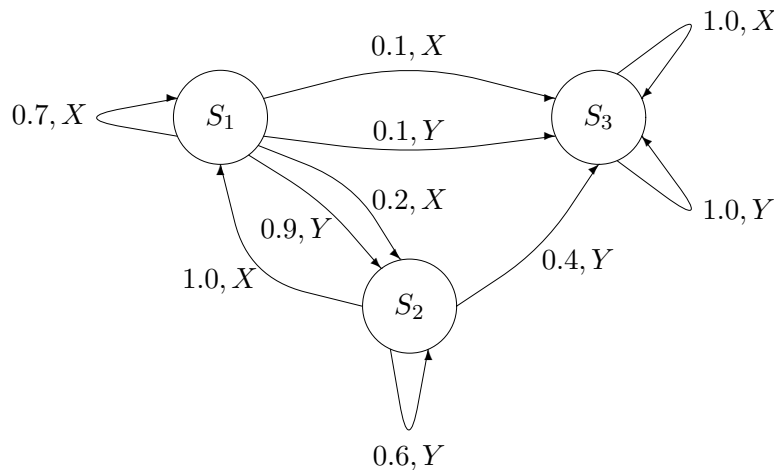
en er zijn de twee mogelijke inputs

$$X = \text{'hallo schat'} \quad \text{en} \quad Y = \text{'oude roestdoos'}$$

dan horen bij deze twee inputs misschien de overgangsmatrices

$$A_X = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 1.0 & 0 & 0 \\ 0 & 0 & 1.0 \end{pmatrix} \quad \text{en} \quad A_Y = \begin{pmatrix} 0 & 0.9 & 0.1 \\ 0 & 0.6 & 0.4 \\ 0 & 0 & 1.0 \end{pmatrix}.$$

Ook een stochastische automaat kunnen we door een state diagram beschrijven, waarbij de labels de input en de kans voor de overgang bij deze input bevatten. In Figuur II.3 is het state diagram voor onze emotionele robot te zien.



Figuur II.3: Stochastische automaat voor een emotionele robot.

Merk op dat in een state diagram voor elke state de som van de kansen op de uitgaande pijltjes voor eenzelfde input gelijk aan 1 moet zijn.

11.3 Markov modellen

We hebben tot nu toe het standpunt ingenomen, dat door een Markov proces een rij states wordt voortgebracht. Maar we kunnen de opzet van een

Markov proces ook opvatten als een model voor het mechanisme dat de rij van states voortbrengt. Uit deze perspectief noemt men het stelsel van states en overgangskansverdelingen een *Markov model*. Het idee hierachter is dat een onbekend proces de states produceert, maar dat we veronderstellen dat dit proces zich gedraagt als een Markov proces en de states en overgangskansen dus een model voor het proces zijn.

Het probleem dat we hierbij moeten oplossen is, uit waarnemingen van rijen van states de overgangskansen tussen de states te schatten. Dit gaan we aan het voorbeeld van de letterfrequenties bekijken:

Voor een Markov model van orde 0 hebben we alleen maar de kansverdeling van de states nodig, dus de kansverdeling van de letters, en die krijgen we als relatieve frequenties van de letters in een (grote) achtergrond tekst (training tekst).

Voor een Markov model van orde 1 moeten we de overgangskansen $a_{ij} := p(q_t = S_j \mid q_{t-1} = S_i)$ bepalen. Maar er geldt voor de voorwaardelijke kans a_{ij} dat

$$p(q_t = S_j \mid q_{t-1} = S_i) = \frac{p(q_t = S_j, q_{t-1} = S_i)}{p(q_{t-1} = S_i)},$$

dus kunnen we de a_{ij} op een training tekst bepalen als quotiënt van de frequentie f_{ij} van letterparen met S_i als eerste letter en de totale frequentie f_i van de letter S_i . Hierbij hoeven we frequenties f_i van de enkele letters niet eens expliciet te bepalen, want er geldt $f_i = \sum_j f_{ij}$ omdat we elk voorkomen van S_i hebben geteld als we alle paren met S_i in de eerste plaats hebben geteld. (Voor de letter op de laatste plaats in de training tekst klopt dit natuurlijk niet, maar deze fout kunnen we verwaarlozen.) We krijgen dus de overgangskansen a_{ij} heel makkelijk als

$$a_{ij} = \frac{f_{ij}}{f_i}.$$

In de aanpak met de relatieve frequenties bestaat er een klein probleem met de *zeldzame gebeurtenissen*. Voor een paar gebeurtenissen met een kleine kans kan het gebeuren dat dit paar in het training materiaal helemaal niet voorkomt. Maar het is in het algemeen niet verstandig aan een overgang de kans 0 toe te kennen (behalve als het een *verboden* overgang is), omdat dit betekent dat men deze overgang nooit kan produceren en het model aan een rij states waarin deze overgang wel voorkomt de kans 0 geeft.

Een simpele (maar vaak voldoende) oplossing van dit probleem is, de teller voor de frequenties van de paren niet met 0 maar met 1 te initialiseren, dus te veronderstellen dat elk gebeurtenis wel een keer is gezien (zo iets als éénmaal is geen maal). Maar er zijn ook subtielere en ingewikkeldere methoden ontwikkeld, dit valt onder het begrip van *gladmaken* van kansverdelingen.

We hebben gezien dat Markov processen en Markov modellen in principe twee zijden van eenzelfde munt zijn: Uit waarnemingen van een onbekend proces maken we een Markov model, en we zeggen dat het Markov model het proces

goed beschrijft als de Markov proces die bij het Markov model hoort een rij gebeurtenissen produceert dat goed met de waarnemingen overeen komt.

11.4 Toepassingen van Markov modellen

De twee perspectieven om naar Markov modellen te kijken geven ook de meest belangrijke typen van toepassing: Simulatie met behulp van Markov processen en classificatie (of toetsing) met behulp van Markov modellen.

Simulatie

Hierbij gebruiken we een Markov model om een rij gebeurtenissen voort te brengen, waarop bijvoorbeeld andere modellen getoetst kunnen worden.

Als we bijvoorbeeld voor de rijen van letters in verschillende talen een Markov model van orde 1 bouwen, kunnen we (onzinnige) teksten produceren, die niettemin typische elementen van de taal laten zien.

Voor de talen Nederlands (NL), Engels (EN), Duits (DU) en Fins (FI) krijgen we zo bijvoorbeeld *teksten* als de volgende (met 160 letters):

NL EVEFOOE OVORER KET DESTS NDEFT MELL CEN HEN ET MEDE ENIJFEBE
HEPGE G IN JEN VOONDEDE HE ESTETETE DE HE DER COROPEETLL
NFFTE LENG MHT VOT HET EUDE DERANLODENGEMH

EN S COTHENN CHENCTHER BEN THXS INTHABJ IT EUPAUS ISTHANTEN
CIOPE WAGESON IN M CONA ATHEDEDED AN JON DERENN T RTH THEPLE
UES PTAD TIONTHAT ERO OR FFION TTUNEROCTHE

DU RELEMM FT DELLATIT APTZERKO TUN ASER WOPF KPEH RARINTOKEN IG
W MT BURER MENGS URHEM ZICHAAT KAHED URIIENSP ENTEN ERT
ZUNAUN SIONG D SE VERZUR HUMAN TSER DIE ASC

FI VA EN MMA LLEN LILIOD TOS IHTORON ATUN MISA VUN KA
OROLUSAMUJA POKUNITUSIM M DOSTOTA HAITTANEMINTISISON URECD
KOMI HTI KUOHOONTOULI T OUUUSKARIS OP SSEHJOITAVU

Deze teksten laten verschillende elementen zien, die typisch voor de talen zijn, zo als de dubbele OO en de IJ in het Nederlands of de TH in het Engels.

Als men in plaats van een Markov model van orde 1 hiervoor een Markov model van orde 2 neemt, dus de relatieve frequenties van tripels van letters telt, worden de verschillen nog veel duidelijker. Maar er zijn $27^3 = 19683$ verschillende tripels van letters, en om hiervoor een redelijke kansverdeling te krijgen, zou men een training tekst van een paar miljoen letters nodig hebben. Voor de $27^2 = 729$ paren van letters hebben we training teksten van slechts ongeveer 50000 letters gebruikt.

Classificatie/Toetsing

We veronderstellen dat we bij een classificatie taak voor elke klasse een Markov model gebouwd hebben dat de elementen van de klasse goed beschrijft. Voor

de verschillende Markov modellen berekenen we nu de kans dat een rij waarnemingen door dit model voortgebracht is en kiezen dan de klas van hetgeen Markov model dat de hoogste kans oplevert. De onderbouwing hiervoor is als volgt:

Bij een Markov proces van orde 0 met N mogelijke uitkomsten S_1, \dots, S_N kunnen we de kans op de rij $x_1 x_2 \dots x_n$ van uitkomsten eenvoudig berekenen als $p(x_1 x_2 \dots x_n) = p(x_1) \cdot p(x_2) \cdot \dots \cdot p(x_n)$.

We kijken nu naar de stochast X van een Markov proces met *echte* kansen $p_i = p(X = S_i)$ en beschrijven deze door een Markov model met kansen $q_i = p'(q_t = S_i)$.

Als n groot is, is het aantal van uitkomsten x_i in de rij ongeveer gelijk aan $n \cdot p_i$. Dan krijgen we

$$p(x_1 x_2 \dots x_n) = \prod_{i=1}^N p(X = S_i)^{n \cdot p_i} = \prod_{i=1}^N q_i^{n \cdot p_i} \text{ voor de juiste kansen en}$$

$$p'(x_1 x_2 \dots x_n) = \prod_{i=1}^N p'(q_t = S_i)^{n \cdot p_i} = \prod_{i=1}^N q_i^{n \cdot p_i} \text{ voor de benadering in het model.}$$

Om rijen van verschillende lengten te kunnen vergelijken moeten we hieruit nog de n -de wortel trekken, dit geeft

$$p(x_1 x_2 \dots x_n)^{\frac{1}{n}} = \prod_{i=1}^N p_i^{p_i} \quad \text{en} \quad p'(x_1 x_2 \dots x_n)^{\frac{1}{n}} = \prod_{i=1}^N q_i^{p_i}.$$

Als we van deze vergelijkingen de logaritme (met basis 2) nemen, krijgen we een verband met een oude bekende uit de laatste les, namelijk

$$H(X) = - \sum_{i=1}^N p_i \cdot {}^2\log(p_i) = - \frac{1}{n} \cdot {}^2\log(p(x_1 x_2 \dots x_n))$$

$$\leq - \sum_{i=1}^N p_i \cdot {}^2\log(q_i) = - \frac{1}{n} \cdot {}^2\log(p'(x_1 x_2 \dots x_n)) =: H.$$

In de limiet $n \rightarrow \infty$ geeft dus $H := - \frac{1}{n} \cdot {}^2\log(p(x_1 x_2 \dots x_n))$ een schatting voor de entropie $H(X)$ van de kansverdeling van de stochast X en deze schatting is beter als H een lagere waarde heeft, want we weten dat het minimum bereikt wordt als Q de juiste kansverdeling van X is.

Wat we net hebben gezien geldt ook voor algemene Markov processen, namelijk dat voor de entropie van een stochast X geldt dat

$$H(X) = \lim_{n \rightarrow \infty} - \frac{1}{n} \cdot {}^2\log(p(x_1 x_2 \dots x_n))$$

waarbij $p(x_1 x_2 \dots x_n)$ de juiste kansverdeling voor de stochast X aangeeft. Als we nu dezelfde kans met de kansen uit een Markov model berekenen, wordt deze kans lager als het Markov model de stochast beter beschrijft.

Een andere manier om hierna te kijken is als volgt: Voor de entropie $H(X)$ is $2^{H(X)}$ het gemiddelde aantal alternatieven dat men voor de stochast X verwacht, met andere woorden, de stochast X is net zo moeilijk als een uniforme verdeling met $2^{H(X)}$ alternatieven. Maar we weten dat voor $H = -\frac{1}{n} \log(p'(x_1 x_2 \dots x_n))$ steeds geldt dat $H \geq H(X)$ en dus ook $2^H \geq 2^{H(X)}$. We kunnen dus zeggen, dat de beschrijving van de stochast X met het Markov model met kansverdeling Q net zo moeilijk is als een uniforme verdeling met 2^H alternatieven, en natuurlijk is degene beschrijving het beste waarvoor 2^H minimaal is.

We passen dit idee nu op korte testteksten toe, waarvoor we de taal willen bepalen. We nemen aan dat we een Markov model van orde 1 hebben met overgangskansen a_{ij} van state S_i naar state S_j en met kans $b_i = \sum_{j=1}^N a_{ij}$ voor state S_i . Met zo'n model berekenen we de kans van een rij $x_1 x_2 \dots x_n$ van letters door

$$\begin{aligned} p(x_1 x_2 \dots x_n) &= p(q_1 = S_{i_1}) \cdot p(q_2 = S_{i_2} | p(q_1 = S_{i_1})) \cdot \dots \cdot p(q_n = S_{i_n} | p(q_{n-1} = S_{i_{n-1}})) \\ &= b_{i_1} \cdot \prod_{j=1}^{n-1} a_{i_j i_{j+1}} \end{aligned}$$

waarbij S_{i_j} de state van de letter x_j is.

De talen zijn weer Nederlands (NL), Engels (EN), Duits (DU) en Fins (FI) en we berekenen voor elke taal de waarde

$$2^H \text{ voor } H = -\frac{1}{n} \log(p(x_1 x_2 \dots x_n))$$

op de volgende testteksten:

T_1 : SINTERKLAAS KOMT NAAR ONS HUIS

T_2 : SANTA CLAUS COMES TO OUR HOUSE

T_3 : NIKOLAUS KOMMT IN UNSER HAUS

T_4 : HANNU MANNINEN

Als resultaat krijgen we de volgende tabel met de waarden van 2^H :

	NL	EN	DU	FI
T_1	14.1	28.3	16.2	19.0
T_2	18.2	12.4	28.9	18.0
T_3	14.4	23.5	9.8	16.5
T_4	19.2	25.0	16.8	14.0

Het is duidelijk dat we in elk geval de juiste taal kunnen achterhalen. Hoe typisch de testteksten voor de enkele talen zijn, kunnen we zien als we de boven gevonden waarden met de waarden op de teksten vergelijken waarop de Markov modellen getraind zijn, dus met de entropieën van de Markov modellen zelfs. De waarden van $2^{H(X)}$ voor de verschillende talen zijn:

NL: 9.2 EN: 9.6 DU: 9.3 FI: 9.7.

De classificatie met behulp van Markov modellen voor letter strings in de verschillende talen is de manier hoe in tekstverwerkingsprogramma's als WORD (OFFICE) automatisch de spellchecker naar een andere taal omgeschakeld wordt, als er bijvoorbeeld in een Nederlandstalige tekst een citaat in het Engels ingebouwd wordt.

11.5 Hidden Markov modellen

Tot nu toe hebben we steeds naar systemen gekeken, waarvoor we de states direct konden waarnemen. We hebben daarom ook geen onderscheiding gemaakt tussen states, uitkomsten en waarnemingen. We krijgen echter een grotere vrijheid in de Markov modellen, als we de states los van de gebeurtenissen en waarnemingen bekijken. Het idee is, dat de states de mogelijke uitkomsten wel veroorzaken, maar dat verschillende states dezelfde uitkomst kunnen produceren en dat niet (noodzakelijk) bekend is, welke state een bepaalde uitkomst heeft veroorzaakt. Om deze reden noemen we de states ook *verstopt* en een Markov model met verstopte states heet een *Hidden Markov model*, of in het kort een *HMM*.

We geven twee opzetten die het idee van de Hidden Markov modellen illustreren:

(1) Het munt model

Achter een gordijn zit iemand die met een of meer munten een muntworpen experiment uitvoert, maar alleen maar de rij uitkomsten (kop/munt) aan de waarnemer doorgeeft. De keuze van de munten voor de enkele worpen volgt een stochastisch proces die door overgangskansen tussen de munten bepaald wordt.

(2) Het vaas model

Er zijn N vazen met telkens ballen van M verschillende kleuren. Iemand trekt (met terugleggen) een bal uit een van de vazen en geeft de kleur van de bal aan de waarnemer door. Vervolgens wordt volgens een toevalskeuze, die afhankelijk van de laatst gekozen vaas is, een nieuwe vaas gekozen.

De algemene ingrediënten van een HMM (van orde 1) zijn als volgt:

- (1) Mogelijke uitkomsten x_1, \dots, x_M . De waargenomen uitkomst op tijdstip t wordt met o_t aangegeven (de letter o staat voor het Engelse *observation*).
- (2) Een aantal states S_1, \dots, S_N , waarbij de state op tijdstip t met q_t aangegeven wordt.
- (3) De overgangskansen $a_{ij} := p(q_t = S_j \mid q_{t-1} = S_i)$ voor de overgang van state S_i naar state S_j .
- (3) Voor elke state S_i een kansverdeling b_i voor de *emissiekansen*, d.w.z. $b_i(x_k) = p(o_t = x_k \mid q_t = S_i)$ is de kans dat in state S_i de uitkomst x_k geproduceerd wordt. Er wordt verondersteld dat deze kansen onafhankelijk van het tijdstip t zijn.

- (4) Een beginverdeling π die de kansen $\pi(i) := p(q_0 = S_i)$ aangeeft dat het systeem op tijdstip $t = 0$ in state S_i is.

We kunnen ook een gewoon Markov model (op een iets kunstmatige manier) als HMM opvatten: Hiervoor identificeren we de uitkomsten x_i met de states S_i en kiezen de emissiekansen b_i zo dat $b_i(x_i) = 1$ en $b_i(x_k) = 0$ als $k \neq i$.

Voorbeeld van een HMM

We bekijken een munt model met drie munten als states, waarvan de eerste eerlijk is, dus kansen $\frac{1}{2}$ voor *kop* en *munten* heeft, de tweede oneerlijk met kans $\frac{3}{4}$ voor *kop* en de derde oneerlijk met kans $\frac{1}{4}$ voor *kop*. Als we K voor de uitkomst *kop* en M voor de uitkomst *munten* schrijven, hebben we de emissiekansen $b_1(K) = b_1(M) = \frac{1}{2}$, $b_2(K) = b_3(M) = \frac{3}{4}$, $b_3(K) = b_2(M) = \frac{1}{4}$, die door de volgende tabel weergegeven worden:

	$b_i(K)$	$b_i(M)$
S_1	0.5	0.5
S_2	0.75	0.25
S_3	0.25	0.75

We veronderstellen verder dat de beginverdeling uniform is, d.w.z. de kans dat het systeem in het begin in state S_i is, is voor elke state $\pi(i) = \frac{1}{3}$.

In een eerste opzet nemen we nu aan dat alle overgangskansen hetzelfde zijn, dus alle $a_{ij} = \frac{1}{3}$.

Stel nu we nemen de rij $O = \text{KMKMK}$ waar.

Omdat de hoogste kans op de uitkomst K in state S_2 zit, de hoogste kans op de uitkomst M in S_3 en de overgangskansen alle hetzelfde zijn, kunnen we makkelijk zien dat de rij $q = S_2S_3S_2S_3S_2$ de rij van states is, waarvoor de kans op de waarneming O maximaal is. In dit geval is deze kans namelijk $p(O, q) = (\frac{1}{3})^5 \cdot (\frac{3}{4})^5 = (\frac{1}{4})^5 \approx 9.77 \cdot 10^{-4}$.

In tegenstelling hiermee is de kans dat deze waarneming door de rij $q' = S_1S_1S_1S_1S_1$ voortgebracht is, slechts $p(O, q') = (\frac{1}{3})^5 \cdot (\frac{1}{2})^5 = (\frac{1}{6})^5 \approx 1.29 \cdot 10^{-4}$. Deze kans is om een factor $(\frac{3}{2})^5 \approx 7.6$ kleiner dan voor de eerdere rij q van states.

Het probleem wordt iets ingewikkelder als de overgangskansen niet meer alle hetzelfde zijn. Stel we hebben de volgende matrix $A = (a_{ij})$ van overgangskansen a_{ij} tussen de states:

$$A = (a_{ij}) := \begin{pmatrix} 0.9 & 0.05 & 0.05 \\ 0.45 & 0.1 & 0.45 \\ 0.45 & 0.45 & 0.1 \end{pmatrix}$$

dan is de kans $p(O, q \mid A)$ (we geven hier voor de duidelijkheid de matrix van overgangskansen mee aan) voor dezelfde rijen waarnemingen en states als boven $p(O, q \mid A) = \frac{1}{3} \cdot 0.45^4 \cdot (\frac{3}{4})^5 \approx 3.24 \cdot 10^{-3}$ en de kans $p(O, q' \mid A)$ is $p(O, q' \mid A) = \frac{1}{3} \cdot 0.9^4 \cdot (\frac{1}{2})^5 \approx 6.83 \cdot 10^{-3}$, dus deze keer is $p(O, q' \mid A)$ om een factor $2^4 (\frac{2}{3})^5 \approx 2.1$ groter dan $p(O, q \mid A)$.

We zien dus dat in het tweede geval de hypothese dat het systeem door de rij q' van states gelopen is, een hogere kans voor de waarneming geeft dan de rij q van states. Het is nu natuurlijk een voor de hand liggende vraag, of er een verdere rij van states is, die een nog hogere kans voor de rij O van waarnemingen geeft. Voor korte rijen kunnen we dit met brute kracht nog wel achterhalen (voor het voorbeeld met 5 waarnemingen en 3 states zijn er $3^5 = 243$ mogelijkheden voor de rij q van states), maar voor langere rijen is dit ondoenlijk. (In het speciaal geval van het voorbeeld is de rij q' inderdaad optimaal, omdat de overgangskans $a_{11} = 0.9$ minstens twee keer groter is dan alle andere overgangskansen en de emissiekansen $b_1(K) = b_1(M) = \frac{1}{2}$ zijn.)

We zitten dus met de vraag hoe we bij een rij waarnemingen de rij states vinden, die de hoogste kans aan de waarnemingen geeft. Dit is één van drie fundamentele problemen in het kader van Hidden Markov modellen die we in de volgende les gaan bespreken.

BELANGRIJKE BEGRIPPEN IN DEZE LES

- Markov processen
- overgangsmatrix
- state diagram
- stochastische automaat
- Markov model
- Hidden Markov model (HMM)

OPGAVEN

52. We bekijken de emotionele robot uit sectie 11.2 en bepalen de kansverdeling voor zijn toestand na twee inputs.
- (i) Veronderstel dat de robot in het begin gelukkig is en bereken de kansverdeling voor elk van de vier mogelijke inputs XX , XY , YX en YY .
 - (ii) Bereken de kansverdelingen voor de verschillende inputs ook voor de gevallen dat de robot in het begin bedroefd of mal was.
53. De states S_1, S_2, S_3 van een Hidden Markov model zijn (net als in het voorbeeld) drie munten die de emissiekansen $\frac{1}{2}, \frac{3}{4}, \frac{1}{4}$ op kop (K) en de emissiekansen $\frac{1}{2}, \frac{1}{4}, \frac{3}{4}$ op munt (M) hebben. De beginverdeling van de states is uniform, dus $\pi(1) = \pi(2) = \pi(3) = \frac{1}{3}$. We bekijken de drie rijen waarnemingen $O_1 = KKKK$, $O_2 = KKKM$, $O_3 = KKMM$.
- (i) Veronderstel dat alle overgangskansen hetzelfde zijn, dus gelijk aan $\frac{1}{3}$. Bepaal de rijen q_1, q_2, q_3 van states, waarvoor de kans dat zij de waarnemingen O_1, O_2, O_3 geproduceerd hebben maximaal is. Bereken voor de gevonden rijen van states de kansen $p(O_1, q_1)$, $p(O_2, q_2)$, $p(O_3, q_3)$.

- (ii) Vergelijk de kansen uit (i) met de kansen $p(O_i, q)$ die men krijgt, als men aanneemt dat altijd de eerlijke munt geworpen wordt, dus als $q = S_1 S_1 S_1 S_1$ is.
- (iii) Veronderstel nu dat de overgangskansen niet uniform zijn, maar gegeven door de matrix

$$A = (a_{ij}) := \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.4 & 0.2 & 0.4 \\ 0.4 & 0.4 & 0.2 \end{pmatrix}.$$

Bereken de kansen $p(O_i, q_i | A)$ voor de rijen van states uit deel (i) en de kansen $p(O_i, q | A)$ voor de rij q van states uit deel (ii) met betrekking tot deze overgangskansen.

- (iv) Probeer in deel (iii) de rijen q'_1, q'_2, q'_3 van states te vinden, zo dat $p(O_i, q'_i | A)$ maximaal wordt.