Analyse de données et modèles stochastiques

15 novembre 2024

Durée : 1h30

> Supports et notes de cours autorisées. Calculatrices autorisées.
>
> Les exercices sont indépendants. Les questions au sein des exercices peuvent pour la plupart être considérées comme indépendantes. Le nombre de point par exercice est donné à titre indicatif et pourra être modifié a posteriori, dans les limites du raisonnable.
>
> Quand vous répondez aux questions, ne vous limitez pas à un simple résultat. Donnez le détail de vos calculs et précisez le raisonnement suivi.

Exercice 1 – Chaînes de Markov cachées (10 points)

> On utilisera pour cet exercice la table des logarithmes simplifiés donnée en annexe.

On s'intéresse à l'état de forme d'un·e athlete au cours du temps. Iel peut être en repos (R), en reprise d'activité (A), en forme (F), ou en super forme (S). Les mesures commencent toujours alors que l'athlète est en reprise d'activite (A) et les probabilitées de passer d'un état à un autre sont les suivantes :

	$vers\ R$	$vers\ A$	$vers\ F$	$vers\ S$
R	0.3	0.7	0.0	0.0
A	0.0	0.3	0.7	0.0
F	0.0	0.0	0.8	0.2
S	0.9	0.0	0.0	0.1

.

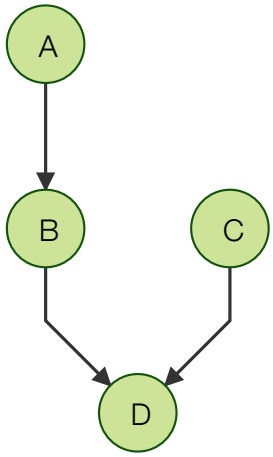
Lors des entrainement, les temps de l'athlète peuvent être de niveau a, b, ou c, et dépendent de l'état de forme come suit :

$$P(temps|forme) = \begin{array}{c|ccc} & vers\ a & vers\ b & vers\ c \\ R & 0.9 & 0.1 & 0.0 \\ A & 0.2 & 0.8 & 0.0 \\ F & 0.1 & 0.5 & 0.4 \\ S & 0.1 & 0.3 & 0.6 \end{array}.$$

Questions

1. Quelle est la probabilité de la séquence (A, a) (F, a) (S, a) (S, c) (S, a) ?
2. Étant donnée la séquence d'observations a b b c a, pour chaque pas de temps $t \in \{1, \ldots, 5\}$ et pour chaque état $i \in R, A, F, S$, calculez la probabilité $P(o_1, \ldots, o_t, S_t = i)$ d'après ce modèle (cette probabilité est un des calcul intermédiaire de l'algorithme EM d'estimation des paramètre d'un modèle de Markov caché).
3. Le modèle de Markov caché repose sur deux hypothèses : l'hypothèse markovienne pour le processus caché ; l'hypothèse d'indépendance conditionnelles des données. Expliquez cette dernière hypothèse (indépendance conditionnelle) : définition, intéret, limite. On pourra s'appuyer pour cela sur un exemple concret.

Exercice 2 – Réseaux bayésiens (6 points)



On considère le réseau bayésien ci-dessus. Les variables sont binaires et le modèle est défini par les paramètres suivants :

$P[A = 0] = 0.3$ $P[B = 0 \mid A] =$

A	
0	0.5
1	0.7

$P[C = 0] = 0.5$ $P[D = 0 \mid B, C] =$

B	C	
0	0	0.3
0	1	0.1
1	0	0.8
1	1	0.5

.

Questions

1. Donnez la forme factorisée de la probabilité conjointe $P[A, B, C, D]$.
2. Calculez $P[C = 1|A = 0, B = 0]$. L'information apportée par A est-elle utile ?
3. Dans le cas particulier de ce réseau, quelle serait un ordre possible d˝elimination des variables pour calculer la probabilité marginale P[C] ?
4. En vous appuyant sur la factorisation de la probabilité conjointe $P[A, B, C, D]$, montrez que les

variables B et C sont indépendantes.

Exercice 3 – Test d'hypothèse (4 points)

Vous souhaitez comparer votre algorithme (Our) à celui de l'état de l'art (SOTA) en ce qui concerne le temps de calcul. Vous les exécutez dans 9 contextes différents et obtenez les temps de calcul (en secondes) :

Setting	$S1$	$S2$	$S3$	$S4$	$S5$	$S6$	$S7$	$S8$	$S9$
`Our`	9.22	9.38	12.30	9.77	10.70	13.00	9.44	12.70	7.08
`SOTA`	8.10	10.74	11.456	8.50	10.98	11.5	10.85	11.19	6.13

Questions

1. Le temps de calcul de Our est-il différent du temps de calcul de SOTA avec un niveau de confiance de 5% ? Préciser les hypothèse de votre test.

Si nécessaire, vous avez en annexe la table de la loi de Student.

Annexe

x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$ln(x)$	-2.3	-1.6	-1.2	-0.9	-0.7	-0.5	-0.4	-0.2	-0.1

Quantile d'ordre γ de la loi de Student à k degrés de liberté.

	γ										
k	0.25	0.20	0.15	0.10	0.05	0.025	0.010	0.005	0.0025	0.0010	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291