



دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

هوش مصنوعی

بهار ۱۴۰۰

استاد: محمدحسین رهبان

گردآورندگان: سپهر اشرف زاده

پاسخ تمرین پنجم، بخش دوم

نمونه گیری و HMM

مهلت ارسال: -

HMM (۵۰ نمره)

۱. (۵۰ نمره)

(آ) داریم:

$$\forall S_k \in \{S_1, S_2\}, \alpha_1^k = P(x_1 | \pi_1 = S_k) P(\pi_1 = S_k)$$

$$\forall S_k \in \{S_1, S_2\}, t \in \{2, 3, 4, 5, 6\}$$

$$\alpha_t^k = P(x_t | \pi_t = S_k) \sum_{i \in \{1, 2\}} \alpha_{t-1}^i a_{i,k}$$

α	$S_k = S_1$	$S_k = S_2$
α_1^k	0.05	0.2
α_2^k	0.032	0.017
α_3^k	0.0029	0.008
α_4^k	0.000392	0.0028
α_5^k	0.0003488	0.0002312
α_6^k	0.00013011	0.000025472

(ب) ابتدا از الگوریتم *backward* استفاده می کنیم.

β	$S_k = S_1$	$S_k = S_2$
β_1^k	0.00057856	0.00079744
β_2^k	0.0051	0.0022
β_3^k	0.0150	0.122
β_4^k	0.04	0.112
β_5^k	0.34	0.16
β_6^k	1	1

سپس داریم:

$$P(\pi_i = S_1 | x, M) = \alpha_i^1 \beta_i^1 / \sum_{k \in \{1, 2\}} \alpha_i^k \beta_i^k$$

	$S_k = S_1$
$P(\pi_1 = S_1 x, M)$	0.2563
$P(\pi_2 = S_1 x, M)$	0.4476
$P(\pi_3 = S_1 x, M)$	0.4476
$P(\pi_4 = S_1 x, M)$	0.2822
$P(\pi_5 = S_1 x, M)$	0.7622
$P(\pi_6 = S_1 x, M)$	0.8363

(ج) با استفاده از این الگوریتم V_t^k را محاسبه می کنیم و حالاتی که احتمال را بیشینه می کنند را نگهداری می کنیم.

V	$S_k = S_1$	$S_k = S_2$
V_1^k	0.05	0.2
V_2^k	0.016	0.016
V_3^k	0.0013	0.0013
V_4^k	0.0001024	0.0016
V_5^k	0.00013107	0.00013107
V_6^k	0.000041943	0.000010486

$Ptr(k, t)$	$S_k = S_1$	$S_k = S_2$
$Ptr(k, 1)$	S_1	S_1
$Ptr(k, 2)$	S_1	S_2
$Ptr(k, 3)$	S_1	S_2
$Ptr(k, 4)$	S_1	S_2
$Ptr(k, 5)$	S_2	S_2
$Ptr(k, 6)$	S_1	S_2

محتمل ترین مسیر پنهان به شکل زیر خواهد بود :

$S_2, S_2, S_2, S_2, S_1, S_1$

Intro to ML (۵۰ نمره)

۱. (۵۰ نمره)

- فرض می کنیم که دسته ی A را داریم. این دسته مجموعه ای از کلمات را داراست که با بردارهایی به شکل $[a_1, a_2, \dots, a_n] : \alpha$ نشان می دهیم. می دانیم که در متن ها با توجه به مفاهیم موجود در متن کلمات مختلف در یک محیط با توجه به یک ارتباط موجود کنار هم قرار می گیرند و رخداد ها از هم مستقل نیستند. اما در قاعده ی بیز ساده انگارانه فرض را بر این می گذاریم که کلمات مستقل از یکدیگر کنار هم می آیند.
- کلاس های موجود را با داده های آموزشی بررسی می کنیم. فرض کنید کلاس B یک کلاس آموزشی است. به مقایسه ی بردار به دست آمده از متن A و داده آموزشی B می پردازیم به این شکل که تعداد تکرار کلمات مشابه را مدنظر قرار داده تا به معیاری برای نزدیکی متن مورد نظر به متن های آموزشی برسیم سپس نزدیک ترین کلاس آموزشی را برمی گزینیم.
- طبق فرض بیز ساده انگارانه ، فرض بر استقلال کلمات خواهد بود که مشخصا در این مورد فرض غلطی خواهد بود. البته برای انجام خوشه بندی محاسبات دقیق وقت گیر و پر هزینه خواهند بود و باعث کند شدن خوشه بندی خواهند شد. در خوشه بندی هدف اختصاص دادن متن به یک کلاس است که نیاز به محاسبات ساده تر دارد. بیز ساده انگارانه برای مواردی که نیاز به محاسبات دقیق دارند مناسب نیست اما در خوشه بندی نسبتا خوب عمل می کند.