

تمارین درس هوش محاسباتی

استاد: دکتر ابراهیم پور

دانشجو: الهه شفیع پور ۹۴۹۹۱۲۶۵

Eshb195@gmail.com

چکیده: در تمارین هدف مقایسه دو الگوریتم پیشبینی برچسب داده ها است. در این تمرین سه دیتاست که دو تا از دیتاست ها توسط خود ما برحسب اطلاعات داده شده ساخته میشود و سومین دیتاست iris هست. این تمارین بدون استفاده از توابع آماده انجام شده است.

1- مقدمه

2- مطلب اصلی

در مقایسه پارامتر هایی که لحاظ شده است شامل: زمان اجرا و تغییر درصد میزان داده ی تست و آموزش است. به علاوه جز دیتاست iris افزایش تعداد داده ها هم جز پارامتر هاست. برای مقایسه دو الگوریتم از معیار precision استفاده شده است برای سنجش کارایی و دقت الگوریتم که برابر است با:

$$P = \frac{T_p}{T_p + F_p}$$

صورت کسر تعداد حدس های درست است. جداول زیر هر کدام شامل این پارامترها و نتیجه ی این مقایسه هاست.

• معرفی الگوریتم ها:

الگوریتم اول الگوریتم بیز هست که براساس تابع احتمال داده ها را به کلاس موردنظر نگاشت میکند هر کدام احتمال بیشتری داشت برای بیز از فرمول زیر استفاده شده است:

$$g_i(x) = -\frac{d}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_i| - \frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) + \log P(\omega_i)$$

که در کد ما هم این فرمول پیاده سازی شده است.

احتمال عضویت هر داده تست به ازای کلاسها سنجیده شده و برای هر کلاس که بیشتر بود برچسب داده ما آن کلاس پیش بینی میگردد.

الگوریتم INN یا نزدیکترین همسایه : فاصله ی داده ی تست را با کل نقاط سنجیده و برچسب نزدیک ترین داده را می گیرد. که همین روش بدون توابع آماده پیاده سازی شده است. توسعه یافته این الگوریتم KNN نیز آمده است.

در کدها کامنت برای توضیحات بیشتر آمده است.

در ادامه مقایسه ی این دو روش را می بینیم.

دیتاست اول دو کلاسه با میانگین و واریانس داده شده:

دیتاست سوم iris سه کلاسه: نتایج اجرا و مقایسه ی دو الگوریتم

	precision	time	Split size
1nn	0.8809	0.0521	0.3
Bayes	0.8267	0.0219	0.3
1nn	0.9540	0.1065	0.5
Bayes	0.9220	0.0180	0.5
1nn	0.9433	0.0651	0.7
Bayes	0.8663	0.0180	0.7

Bayes and 1NN on dataset3 iris. 3

نتایج:

زمان اجرای کمتر بیز کاملاً مشهود است. در نسبت داده تست به آموزش در ۵۰ درصد بهترین عملکرد را دیدم. در دیتاست سوم تفاوت دقت دو روش خیلی نزدیک و گاهی هر دو حدود ۹۰ درصد بودند که در دیتاست اول این طور نبود. داده های یکسان گویا نتیجه ها را بهتر میکنند.

3- نتایج کلی

از نظر زمانی 1NN در مقیاس داده حجیم زمان بسیار زیادی نیاز دارد اما بیز در زما کمتر انجام می پذیرد.

عملکرد بیز وقتی تعداد داده بیشتر شود به دلیل بهبود تابع احتمال کارایی بیشتری دارد.

	precision	time	Size dataset	Split size
1nn	0.9687	0.0728	100	0.3
Bayes	0.7083	0.0080	100	0.3
1nn	0.9968	1.7488	1000	0.3
Bayes	0.7341	0.0574	1000	0.3
1nn	0.9779	2.0818	1000	0.5
Bayes	0.7253	0.0306	1000	0.5
1nn	0.9749	1.3720	1000	0.8
Bayes	0.7327	0.0674	1000	0.8
1nn		Very long	10000	0.5
Bayes	0.7215	0.2922	10000	0.5

مقایسه 1. Bayes and 1NN on dataset1 2 classes

نتایج:

- نتیجه ۱: برای تعداد داده زیاد زمان اجرای 1nn بسیار زیاد میشود و بیز مناسبتر است.
- نتیجه ۲: با افزایش داده ها عملکرد بیز بهتر میشود چون با افزایش نرخ مشاهده تابع احتمال بیز بهبود می یابد.
- نتیجه ۳: در بیز هرچه داده آموزش بیشتری در اختیار باشد به دلیل نتیجه ۲ بهتر عمل میکند البته تاثیر این عمل را در 1nn به وضوح ندیدم.

دیتاست دوم سه کلاسه با میانگین و کواریانس داده شده:

	Precision	time	Size dataset	Split size
1nn	0.9233	0.0572	150	0.5
Bayes	0.8166	0.0563	150	0.5
1nn	0.9493	4.12	1500	0.5
1nn	0.9444	3.7	1500	0.3

Bayes and 1NN on dataset2 3classes. 2