انتخاب نمونه
Instance Selection

عادل قاضی خانی
هیئت علمی گروه کامپیوتر دانشگاه امام رضا (ع)

تعریف

- یک **فرایند پیش پردازش برای کاهش تعداد نمونه ها** به منظور **کاهش سربار محاسباتی یادگیری یا حذف نویز** بدون اینکه قدرت پیش بینی تضعیف شود

- یک کاربرد مهم این فرایند برای الگوریتم **نزدیکترین همسایه**

انتخاب نمونه

- **ویژگی‌های الگوریتمها:**

- **جهت جستجو**
- **نوع انتخاب**
- **ارزیابی جستجو**

1. Instance selection
2. Prototype selection

S. Garcia, J. Derrac, J. Cano, F. Herrera, Prototype selection for nearest neighbor classification: Taxonomy and empirical study, Pattern Anal. Mach. Intell. IEEE Trans. 34 (3) (2012) 417–435,

انتخاب نمونه – ویژگی الگوریتم ها

- **جهت جستجو**

❖افزایشی
❖کاهشی
❖تکه ای
❖مخلوط
❖ثابت

1. Incremental
2. Decremental
3. Batch
4. Mixed
5. Fixed

انتخاب نمونه – ویژگی الگوریتم ها

- **جهت جستجو**

❖**افزایشی:** شروع از مجموعه تهی از نمونه ها و افزودن تا برآورده شدن معیار ارزیابی جستجو

❖**کاهشی :** شروع از مجموعه تمام نمونه ها و کاهش تا برآورده شدن معیار ارزیابی جستجو

❖**تکه ای :** شروع از مجموعه تمام نمونه ها و بررسی اینکه هر داده بر اساس معیار ارزیابی جستجو امکان حذف را دارد سپس حذف یکجا نمونه ها

انتخاب نمونه – ویژگی الگوریتم ها

- **جهت جستجو – ادامه**

❖**مخلوط** : شروع از زیر مجموعه ای نمونه ها و افزودن یا کاهش نمونه هایی برای برآورده شدن معیار خاص

❖**ثابت** : مشابه مخلوط با این تفاوت که تعداد افزایش و کاهش با هم برابر است

انتخاب نمونه – ویژگی الگوریتم ها

- **نوع انتخاب**

- تراکم
- نسخه
- ترکیبی

1. Condensation
2. Edition
3. Hybrid

انتخاب نمونه – ویژگی الگوریتم ها

- **نوع انتخاب**
- **تراکم** : در این روشها نقاط مرزی حفظ شده نقاط داخلی حذف می شوند
- **نسخه** : در این روشها سعی می شود نقاط مرزی نویزی حذف شوند
- **ترکیبی** : حذف می تواند از نقاط مرزی یا داخلی باشد

1. Condensation
2. Edition
3. Hybrid

انتخاب نمونه – ویژگی الگوریتم ها

- **ارزیابی جستجو**

- فیلتر
- پوشش

1. feature selection
2. feature extraction

انتخاب نمونه – ویژگی الگوریتم ها

- **ارزیابی جستجو**

- **فیلتر:** ارزیابی جستجو با زیر مجموعه ای از داده های آموزشی با استفاده از knn صورت می گیرد
- **پوشش:** ارزیابی جستجو با تمام داده های آموزشی و با استفاده از knn صورت می گیرد

1. feature selection
2. feature extraction

انتخاب نمونه – مثال

- **استفاد از انتخاب نمونه در knn برای رگرسیون**
- **knn برای رگرسیون؟**
- یافتن k نزدیکترین همسایه برای یک نمونه تست و میانگین گیری بین خروجی ها برای تعیین خروجی داده تست

Yunsheng Song, Jiye Liang, Jing Lu, Xingwang Zhao,"An efficient instance selection algorithm for k nearest neighbor regression", Neurocomputing, Volume 251,2017, Pages 26-34,

انتخاب نمونه – مثال

- **استفاد از انتخاب نمونه در knn برای رگرسیون**
- در این مقاله دو مرحله اصلی وجود دارد:
  - شناسایی نویز
  - حذف داده های غیر متمایز

Yunsheng Song, Jiye Liang, Jing Lu, Xingwang Zhao,"An efficient instance selection algorithm for k nearest neighbor regression", Neurocomputing, Volume 251,2017, Pages 26-34,

انتخاب نمونه – مثال

- **استفاد از انتخاب نمونه در knn برای رگرسیون**
- در این مقاله دو مرحله اصلی وجود دارد:
  - شناسایی نویز با رابطه زیر

$$PD(x_i) = |y_i - \widehat{y}_i| > (1-\theta)y_i$$

Yunsheng Song, Jiye Liang, Jing Lu, Xingwang Zhao,"An efficient instance selection algorithm for k nearest neighbor regression", Neurocomputing, Volume 251,2017, Pages 26-34,

انتخاب نمونه – مثال

- **استفاد از انتخاب نمونه در knn برای رگرسیون**
- در این مقاله دو مرحله اصلی وجود دارد:
  - **مرحله اول:** شناسایی نویز با رابطه زیر

$$PD(x_i) = |y_i - \widehat{y}_i| > (1-\theta)y_i$$

Yunsheng Song, Jiye Liang, Jing Lu, Xingwang Zhao,"An efficient instance selection algorithm for k nearest neighbor regression", Neurocomputing, Volume 251,2017, Pages 26-34,

انتخاب نمونه – مثال

- **استفاد از انتخاب نمونه در knn برای رگرسیون**
- در این مقاله دو مرحله اصلی وجود دارد:

❖**مرحله دوم:** شناسایی داده های غیر متمایز با توجه به اختلاف خطای رگرسیون بعد از حذف با قبل از حذف است که اگر کمتر مساوی صفر باشد خوب است.

Yunsheng Song, Jiye Liang, Jing Lu, Xingwang Zhao,"An efficient instance selection algorithm for k nearest neighbor regression", Neurocomputing, Volume 251,2017, Pages 26-34,

جمع بندی

- الگورتیمها انتخاب نمونه برای کاهش بار محاسباتی یا حذف نویز استفاده می شوند
- الگوریتمها بر اساس نوع جستجو دسته بندی می شوند