

بِسْمِ اللَّهِ

Expert Systems

سیستم های خبره

مدرس: ابوالقاسم حسن پور

مدرس: ابوالقاسم حسن پور

مطالعه‌ی بسیار و پی‌گیر در مسائل علمی، باعث شگفتی عقل و تقویت نیروی فکر و فهم است. امام صادق (ع)

مدرس: ابوالقاسم حسن پور

Expert Systems

فصل هشتم

جستجوی نزدیکترین همسایه

Nearest Neighbor Search

مدرس: ابوالقاسم حسن پور

یادگیری بر پایه نمونه

- * در روش یادگیری بر پایه نمونه به سادگی فقط مثالها را ذخیره می‌کنیم و هرگونه تعمیم تا مشاهده مثال جدید به تعویق می‌افتد. در سایر روش‌ها، سعی می‌شود با استفاده از مثال‌های آموزشی تابعی پیدا کنیم که بتواند توصیف‌کننده داده‌ها باشد.
- * با مشاهده مثال‌های جدید رابطه آن با نمونه‌های ذخیره شده بررسی شده و یک مقدار برای تابع هدف آن نسبت داده می‌شود.
- * دسته‌بندی هر نمونه جدید هنگام مشاهده آن و بر اساس نزدیکترین مثال‌های ذخیره شده، انجام خواهد شد.

مدرس: ابوالقاسم حسن پور

مشخصه های یادگیری بر پایه نمونه

- **تابع شباهت:**
مشخص می کند که دو نمونه چقدر نزدیک به هم هستند. مثلاً چگونه می توان شباهت رنگ موی ۲ نفر را بیان نمود؟
- **انتخاب نمونه ها برای ذخیره:**
در این الگوریتم سعی می شود نمونه هایی ذخیره شوند که عمومی تر باشند. تشخیص اینکه آیا یک نمونه عمومیت دارد یا خیر، می تواند کار مشکلی باشد.
- **تابع دسته بندی کننده:**
تابعی است که با مشاهده یک مثال دسته بندی آنرا تعیین می کند.

مدرس: ابوالقاسم حسن پور 5

جستجوی نزدیکترین همسایه

- جستجوی نزدیکترین همسایه یک مسئله بهینه سازی برای پیدا کردن نزدیکترین نقطه ها در فضاها می باشد.
- مجموعه S شامل تعدادی نقطه در یک فضای متریک مانند M و نیز یک نقطه پرس و جوی $q \in M$ داده شده است، هدف پیدا کردن نزدیک ترین نقطه در S به q است. در بسیاری از موارد، فضای M بصورت یک فضای اقلیدسی d بعدی و فاصله بین نقاط با معیار فاصله اقلیدسی، فاصله منتهن یا دیگر فاصله های متریک سنجیده می شود.

مدرس: ابوالقاسم حسن پور 6

KNN

- جستجوی k نزدیکترین همسایه روش یادگیری بر پایه نمونه است.
- جستجوی k نزدیکترین همسایه، K همسایه نزدیک تر به نقطه پرس و جو را برمی گرداند. این روش معمولاً در تجزیه و تحلیل پیش بینی، به منظور تخمین و یا دسته بندی یک نقطه بر اساس اجماع همسایگان آن استفاده می شود. گراف k نزدیکترین همسایه، گرافی است که در آن هر نقطه در گراف K نزدیک ترین همسایگان خود متصل است.
- در این روش تمام نمونه ها نقاطی در فضای n بعدی حقیقی هستند و همسایه ها بر مبنای فواصل اقلیدسی استاندارد تعیین می شوند.
- اگر k خیلی کوچک باشد، نسبت به نویز حساس خواهد بود و اگر K خیلی بزرگ باشد ممکن است یک همسایگی نقاطی از سایر کلاس ها را نیز در بر بگیرد.

مدرس: ابوالقاسم حسن پور 7

فاصله اقلیدسی

- اگر یک مثال دلخواه را بصورت یک بردار ویژگی نمایش دهیم:

$$\langle a_1(x), a_2(x), \dots, a_n(x) \rangle$$

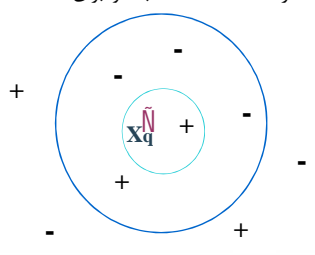
- فاصله بین دو مثال X_i و X_j بصورت زیر تعریف می شود:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

مدرس: ابوالقاسم حسن پور 8

• اگر $k=1$ انتخاب شود الگوریتم 1-NN مقدار نزدیکترین نمونه به x_q را انتخاب خواهد نمود. برای مقادیر بزرگتر k متداولترین مقدار بین k همسایه نزدیک انتخاب خواهد شد. در این مثال x_q در حالت 1-NN مثبت و برای 5-NN منفی خواهد بود.

مثال




مدرس: ابوالقاسم حسن پور

روش کار KNN

- ابتدا توسط تابعی فاصله بین نقاط تعلیم و تست محاسبه می شود و در ماتریسی به نام d قرار می گیرد.
- ماتریس d مرتب می شود. یعنی برای هر نقطه تست نقاط تعلیم به صورت صعودی بر حسب فاصله آن نقطه تعلیم از نقطه تست در سطر مربوط به آن قرار می گیرد.
- K نقطه تعلیم اول هر یک از سطرها را به عنوان K نزدیکترین همسایه جدا می شود.
- برای هر نقطه تست محاسبه می شود که هر یک از K نزدیکترین همسایه در چه کلاسی قرار دارد.
- کلاسی که ماکزیموم تعداد همسایه ها در آن قرار دارند را به عنوان کلاس خروجی انتخاب می شود.

مدرس: ابوالقاسم حسن پور



مدرس: ابوالقاسم حسن پور