

7) مجموعه داده مورد نیاز برای FFM

این سبک ها به مقدار زیادی داده آموزشی نیاز دارند. مقدار زیاد به این معناست که اگر تعداد داده های که داریم خیلی زیاد است، باید به نوعی جن $test/train$ جدا کنیم که توزیع داده در هر بخش دیتاست داده به طور داده ای نوعی هر بخش باشد. اما اگر داده خیلی زیاد نداریم، می توان داده جدید ساخت (به کمک چرخش افزودن نویز، scale و...) همین طور می توان از روش $N-fold$ استفاده کرد. به این صورت که طبقه را با $n-1$ داده که بخش N امی است کنیم و این عملیات را N بار تکرار می کنیم. داده ها تکرار کنیم. یک صفت دیگر، تعداد سبک داده آموزشی داریم و این بود که اگر می خواهیم صفات N مورد استفاده N آوزن دارد. باید $\frac{N_{train}}{2}$ داده آموزشی داشته باشیم و تا جایی که $\frac{N}{2}$ برای داده آموزشی شش داریم.

مقدار از داده سبک، داده های نرسانه سازی شده است. زیرا اگر نرسانه سازی انجام ندهیم یعنی h از ویژگی ها ممکن است نادیده گرفته شوند. انواع مختلفی $normalization$ وجود دارد. هم چنین در خصوص داده های باستانی می دانیم که نمایش دو قطبی بهتر از نمایش $1/0$ یعنی باستانی می باشد.