

قسمت ۱۱ سوال ۴:

شبکه عصبی و svm نمیتوانند در برابر داده های میس شده درست عمل کنند چون این الگوریتم ها باید به فیچر وزن بدهند. پس باید مقادیر میس شده شان را درست کرده و با مقداری پر کنیم یا در غیر اینصورت رکورد دارای مقدار میس شده را باید حذف کرد.

از طرف دیگر knn در برابر داده های میس شده مقاوم است. چون این الگوریتم باید فاصله اقلیدوسی میان فیچرها را محاسبه کند پس برای محاسبه فاصله لازم است تمامی ابعاد داده وجود داشته باشند پس اگر با داده ی میس شده ای برخورد کند باید آن فیچر یا بعد را حذف کند. اما الگوریتم درنمانده و با حذف بعد پاسخگو است.

درخت تصمیم و رندوم فارست نیز در برابر داده های میس شده مقاوم هستند. برخی از پیاده سازی های مدل های مبتنی بر درخت، رویه های هوشمندانه ای برای هندل کردن داده های ناقص دارند. مثلاً روش CART از ایده انشعابات جایگزین استفاده می کند. هنگام ایجاد یک درخت مجموعه ای از نقاط اسپلیت برای شاخه کردن درخت در نظر گرفته میشود. از پیش بینی کننده های جایگزین بجای پیش بینی کننده فعلی میتواند استفاده شود و اگر مقدار پیش بینی کننده فعلی وجود نداشته باشد میتواند با توجه به مقادیر دیگری که وجود دارد آن مقدار را تخمین بزند و جلو برود.

روش naivebayes در برابر داده های میس شده خیلی مقاوم است. و موقع حساب کردن احتمالات مورد نیاز مقادیر میس شده را درگیر نکرده و اصلاً در نظر نمیگیرد و روند خود را پیش میبرد.