



تمرین ۳-ب - آشنایی با یادگیری ماشین (DNF)

مسئله: در این مسئله یادگیری DNF مورد بررسی قرار می‌گیرد.

الف) یک DNF با ۱۰۰ مشخصه و ۲۰۰۰ نمونه (مثال) که توسط یک ماتریس ۲۰۰۰ در ۱۰۱ (فایل ارسالی DNFData.TXT) توصیف می‌شود در نظر بگیرید. در هر سطر ماتریس ۱۰۰ مقدار اول مقادیر مشخصه‌ها در آن مثال و مقدار آخر مقدار خروجی DNF به ازای این مشخصه‌ها است. الگوریتم یادگیری DNF را در حالت بدون نویز به ازای طول داده آموزشی از ۲۰ تا ۱۰۰۰ (با فاصله قدم ۲۰) اجرا کرده و خطای DNF بدست آمده را بر روی باقی مثال‌ها (که در اینجا آن را داده اعتبارسنج در نظر می‌گیریم) محاسبه نمایید. نمودار یادگیری را ترسیم نمایید. به ازای چه طولی از داده آموزشی خطا به صفر می‌رسد؟ در این حالت ضابطه DNF را به طور دقیق مشخص نمایید.

ب) یک DNF دیگر با ۵۰ مشخصه و ۵۰۰ نمونه (مثال) که توسط یک ماتریس ۵۰۰ در ۵۱ (فایل ارسالی DNFNoisy.TXT) توصیف می‌شود در نظر بگیرید. در این مثال حدود ۱۰٪ خروجی‌ها اشتباه هستند. الگوریتم یادگیری DNF را در حالت نویزی بر روی داده آموزشی به طول ۲۵۰ و ازای مقادیر مختلف epsilon از ۰.۰۰۱ تا ۱ (با فاصله قدم ۰.۰۵) اجرا کرده و خطای DNF بدست آمده را بر روی باقی مثال‌ها (که در اینجا آن را داده اعتبارسنج در نظر می‌گیریم) محاسبه نمایید. نمودار خطا بر حسب epsilon را ترسیم نمایید. به ازای چه مقدار یا مقادیری از epsilon کمترین میزان خطا را داریم؟ این خطا چقدر است؟ در این حالت (کمترین خطا) ضابطه DNF را به طور دقیق مشخص نمایید.