

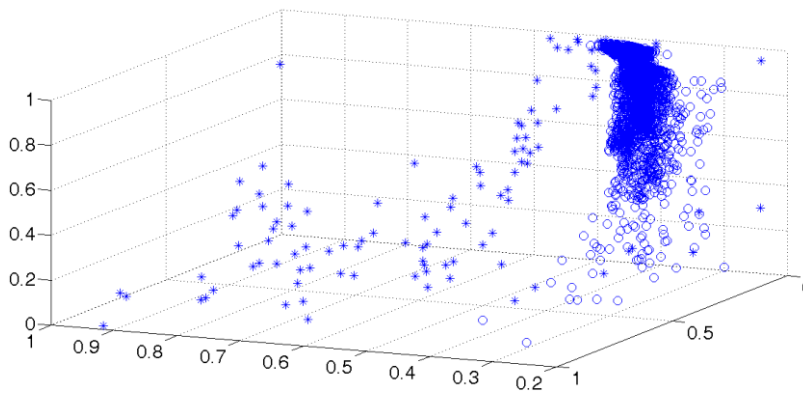
درس بازشناسایی آماری الگو (PRML)

تمرین (سری هشتم)

مهلت تحویل:

هدف: می‌خواهیم مباحث طرح شده در Lecture 7-11 را در این سری تمرین مرور کنیم.

پیاده‌سازی: نمونه‌هایی سه‌بعدی از دو کلاس S و NS را به صورت سطریهای ماتریس‌هایی به همین نام در فایل‌های پیوست در اختیار داریم. از هر کلاس، نیمی از نمونه‌ها را بصورت تصادفی به عنوان مجموعه آموزش (Training Set) و نیم باقیمانده را به عنوان مجموعه آزمون (Test Set) برگزینید (بهتر است Scatter Plot متناظر با مجموعه‌های آموزش را نیز نشان دهید).



شکل ۱- نمایش Scatter Plot کل داده‌های S (دایره) و NS (ستاره)

مسئله یک: الف- از روی مجموعه‌های آموزشی، برای هر یک از این دو کلاس، بهترین توزیع نرمالی که پراکندگی درون‌کلاسی هر یک را توصیف می‌کند، را بیابید.

ب- مبتنی بر توزیع‌های نرمال به دست آمده، دسته‌بندی نمونه‌های مجموعه آزمون را انجام دهید و تعداد Mis-Classification ها را در هر یک از کلاس‌ها بیابید.

ج- (اختیاری) اگر از هریک از سه ویژگی نمونه‌ها به تنهایی برای تخمین بهترین توابع توزیع نرمال درون‌کلاسی در مجموعه آموزش و سپس برای دسته‌بندی نمونه‌های آزمون بهره گیریم، خطای حاصل را با خطای قسمت ب مقایسه کنید.

مسئله دو: مبتنی بر روش KNN از روی نمونه‌های آموزش، دسته‌بندی نمونه‌های آزمون را انجام داده و به ازای K های گوناگون (مقادیرهایی برابر ۱، ۳، ۵ و ۷) خطای دسته‌بندی نمونه‌های مجموعه آزمون را تعیین و تحلیل کنید.

مسئله سه: به کمک SVM (در MATLAB یا Python) از روی نمونه‌های آموزش، دسته‌بندی نمونه‌های آزمون را انجام داده و به ازای اندازه پارامترهای گوناگون، خطای دسته‌بندی نمونه‌های مجموعه آزمون را تعیین و تحلیل کنید.