

دانشگاه صنعتی امیرکبیر

دانشکده مهندسی برق

یادگیری ماشین

تمرین عملی بیزین

فهرست گزارش سوالات (لطفاً پس از تکمیل گزارش، این فهرست را به روز کنید.)

سوال 1 - تشخیص دیابت.....2

سوال ۲ - تشخیص سرطان سینه.....3

سوال 1 – تشخیص دیابت

در این سوال قرار است که افرادی که دیابت دارند را تشخیص دهید. به این لینک برای دریافت دیتاست بروید. این دیتاست 8 ویژگی به همراه 2 کلاس دارد. آنالیز بیزین بر روی این دیتاست باید انجام شود.

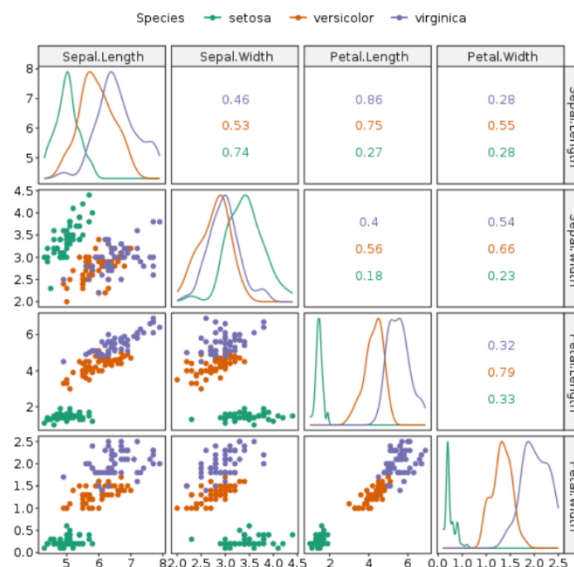
1- با استفاده از اسکترپلات رابطه بین فیچر های مختلف را نشان دهید. (همانند شکل 1 اپندیکس)

2- دیتاست را به صورت 70 به 30 برای آموزش و تست جدا کنید.

3- مدل سازی : با استفاده از Gaussian Naïve Bayes دسته بندی را انجام دهید. (با استفاده از کتابخانه)

4- ماتریس آشفتگی را برای کلاسیفایر های روی آموزش و تست نشان دهید.

Appendix :



سوال ۲ – تشخیص سرطان سینه

مجموعه داده ای در فایل پیوست قرار دارد، شامل مجموعه ای از ویژگی های اندازه گیری شده پزشکی برای تشخیص سرطان سینه. داده ها در فایل BreastCancer.csv قرار داده شده اند. ابتدا داده ها را به صورت یکنواخت از هر دو کلاس بین داده های آموزش، ارزیابی و آزمون تقسیم بندی کنید.

الف) ابتدا با استفاده از مدل Naïve Bayes و با روش Cross validation عملکرد این الگوریتم را بر روی داده ها مشاهده نمایید و Accuracy و ماتریس درهم ریختگی را گزارش کنید.

ب) این بار با استفاده از مدل Gaussian Naïve Bayes فرایند فوق را تکرار کنید.