



قبل از اینکه شروع به حل تمرین‌ها کنید، حتما یک بار فایل **Policies.pdf** (موجود در وبگاه) که حاوی نکات مهم در تحویل تکالیف هست را مطالعه فرمایید.

۱. فرض کنید داده های X_1, \dots, X_n داده هایی iid از یک توزیع یکنواخت حول دایره ای به شعاع θ در فضای R^2 باشند به طوری که $X \in R^2$ و

$$P(x|\theta) = \begin{cases} \frac{1}{\pi\theta^2} & \text{if } |x| \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

با استفاده از MLE، θ را تخمین بزنید.

۲. تابع چگالی احتمال توزیع پواسون با پارامتر $\lambda > 0$ به صورت زیر است:

$$f(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots$$

آیا پارامتر تخمین زده با استفاده از بیشینه درست‌نمایی دارای سوگیری است؟

۳. در این سوال با برنامه نویسی در محیط Python آشنا خواهید شد. برای شروع کدهای زیر را در محیط برنامه کپی نمایید سپس گام به گام مراحل خواسته شده در زیر را طی کنید. داده‌هایی که مورد استفاده قرار می‌گیرد داده‌های "Fisher iris" می‌باشد این داده‌ها شامل ۴ اندازه حقیقی به عنوان ویژگی‌های سه نوع گل زنبق (iris) می‌باشد.

```
import numpy as np
import matplotlib.pyplot as plt

iris = np.genfromtxt("data/iris.txt", delimiter=None) # load the text file
Y = iris[:, -1] # target value is the last column
X = iris[:, 0:-1] # features are the other columns
```

الف) با استفاده از برنامه‌نویسی تعداد ویژگی‌ها و تعداد داده‌ها را به دست آورید.

ب) برای هر ویژگی هیستوگرام مقادیر داده‌ها را رسم کنید.

ج) برای هر ویژگی میانگین آن‌ها را محاسبه کنید

د) برای داده‌های هر ویژگی مقادیر واریانس و انحراف معیار را به دست آورید.

ه) داده‌ها را با محاسبه تفاضل میانگین از داده‌ها بخش بر انحراف معیار نرمال نمایید.

۴. روش Bayes یکی از راه‌های به دست آوردن مرز تصمیم در دسته‌بندی‌هاست^۱. در این حالت ما مرز تصمیم بهینه را برای دو کلاس C_1 و C_2 با استفاده از نسبت likelihood به دست می‌آوریم.

^۱ Bayesian Decision Boundary

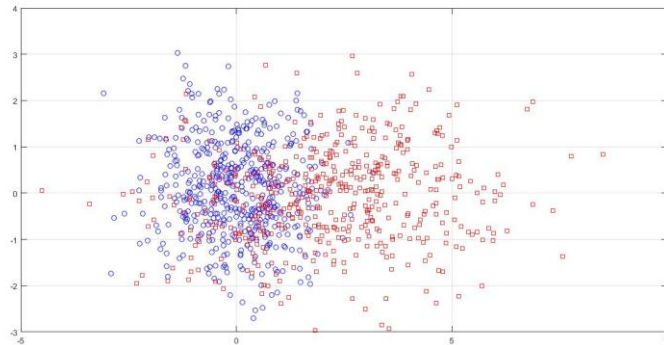
داده های Class1 و Class2 مربوط به شکل زیر را در نظر بگیرید هر یک شامل ۵۰۰ داده دویبعدی است. قصد داریم این داده‌ها را به وسیله یک مرز از هم جدا سازیم:

فرض کنیم می‌دانیم، توزیع داده‌ها در کلاس C_1 بصورت گوسی با بردار میانگین $[0,0]$ و واریانس ۱ و برای کلاس C_2 با بردار میانگین $[0,2]$ و واریانس ۴ می‌باشد. و threshold ما ۱ باشد.

الف) ابتدا داده‌ها را وارد کنید و شکل هر کدام را ترسیم کنید.

ب) با استفاده از تکنیک تخمین با استفاده از بیشینه‌ی درست‌نمایی سعی کنید مرزی برای جداسازی این دو کلاس را به صورت دستی و با کد متلب بیابید.

ج) به نظر تان مرز بدست آمده بهینه است.



شکل ۱- پراکندگی داده‌ها با هر دو کلاس

۵. در این سوال به تخمین یک تابع سینوسی با استفاده از Polynomial Regression پرداخته می‌شود، سپس اثر regularization را بر آن بررسی خواهد شد.

الف) تابع \sin با ورودی زاویه که بر حسب درجه بین ۰ تا ۳۶۰ درجه با فاصله ۵ درجه‌ای را در نظر بگیرید (۵، ۱۰، ۱۵، ...) و مقداری نویز به آن بیافزایید. (تابعی تصادفی) سپس شکل داده‌ها را رسم نمایید.

ب) با استفاده از Polynomial regression تابع تخمینی با درجات ۱، ۳، ۵، ۷ و ۱۰ را به دست آورده و رسم نمایید. سپس پاسخ دهید ضرایب درجات مختلف چند جمله‌ای با زیاد شدن درجه چند جمله‌ای چه تغییری می‌کند.

ج) در هر مورد میزان بایاس و واریانس را حساب کنید.