



به نام خدا

دانشگاه تهران



دانشکده مهندسی برق و کامپیوتر

درس تخمین و شناسایی سیستم‌ها

پروژه اول

دستیاران طراح	
علیرضا گلابوند	نام دستیار طراح
alirezag213@gmail.com	رایانامه
حمزه کاشی	نام دستیار طراح
hamzehkashi@gmail.com	رایانامه
فائزه فتحی	نام دستیار طراح
faezehfat97@gmail.com	رایانامه
۱۴۰۲/۰۲/۱۵	مهلت ارسال پاسخ

فهرست

۲	قوانین
۳	صورت پروژه
۳	مجموعه داده
۴	توضیحات
۵	سوال (۱)
۶	سوال (۲)
۷	سوال (۳)
۸	سوال (۴)
۹	سوال (۵)

قبل از پاسخ دادن به پرسش‌ها، موارد زیر را با دقت مطالعه نمایید:

- از پاسخ‌های خود یک گزارش در قالبی که در صفحه‌ی درس در سامانه‌ی Elearn با نام **REPORTS_TEMPLATE.docx** قرار داده شده تهیه نمایید.
- **کیفیت گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است؛** بنابراین، لطفا تمامی نکات و فرض‌هایی را که در پیاده‌سازی‌ها و محاسبات خود در نظر می‌گیرید در گزارش ذکر کنید.
- در گزارش خود، برای شکل‌ها زیرنویس و برای جدول‌ها بالانویس در نظر بگیرید. پاسخ به سوالات مطرح شده، شامل **شکل‌ها، خروجی شبیه‌سازی، تحلیل نتایج و نتیجه‌گیری** می‌باشد.
- الزامی به ارائه توضیح جزئیات کد در گزارش نیست، اما باید نتایج بدست آمده از آن را گزارش و تحلیل کنید.
- **تحلیل نتایج الزامی می‌باشد.**
- **در صورت مشاهده تقلب امتیاز تمامی افراد شرکت‌کننده در آن، ۱۰۰- لحاظ می‌شود.**
- تنها زبان برنامه نویسی مجاز **Matlab** است. کد شبیه‌سازی باید بدون خطا و صرفا با اعمال Run در Matlab اجرا شود و **مطابق** با متن گزارش باشد.
- برای پاسخ به سوالات مطرح شده، مجاز به استفاده از کتاب، جزوه، فایل ارائه دروس و مراجع آنلاین و آنلاین هستید؛ **اما مطلقا مجاز نیستید از هیچ فردی به هر نحوی کمک بگیرید (در صورت مشاهده تقلب امتیاز تمامی افراد شرکت‌کننده در آن، ۱۰۰- لحاظ می‌شود).**
- لطفا گزارش، کدها و سایر ضمیمه را به در یک پوشه با نام زیر قرار داده و آن را فشرده سازید، سپس در سامانه‌ی Elearn بارگذاری نمایید:

HW[Number]_[Lastname]_[StudentNumber].zip

(مثال: HW1_Ahmadi_810199101.zip)

- در صورت تمایل، استفاده از ابزار **CVX** برای مسائل شامل بهینه‌سازی، مجاز است (برای اطلاعات بیشتر و دانلود پکیج CVX به این [لینک](#) مراجعه کنید).

صورت پروژه

فرض سیستم خطی با ساختار زیر در نظر گرفته شده است. چند جمله‌ای زیر را به عنوان معادله توصیفی یک سیستم فیزیکی غیرخطی استاتیک در نظر بگیرید.

$$y = \sum_{i=0}^n \theta_i u_i + v$$

سیستم غیرخطی فوق می‌تواند به عنوان یک رگرسیون خطی ارائه شود و v بیانگر نویز است.

$$\begin{cases} x = (1, u_1, \dots, u_n) \\ \theta = (\theta_0, \theta_1, \dots, \theta_n) \end{cases} \rightarrow y = x^T \theta + v$$

مجموعه داده^۱

دادگان مورد استفاده در فایل Datasets موجود در فایل پروژه قرار دارند. در این فایل، داده‌های مربوط به پنج سیستم مختلف قرار دارند.

۱. مجموعه داده‌ها شامل ۵ دسته هستند که هر دسته شامل دو سری A و B است.

۲. مجموعه داده انتخاب شده برای هر سؤال، با توجه به قاعده زیر انتخاب می‌گردد.

a. ابتدا رقم یکان شماره دانشجویی خود را با شماره سوال جمع کنید؛

b. حاصل را بر ۵ تقسیم کنید؛

c. عدد باقی‌مانده، نشان‌دهنده شماره سیستمی است که از دادگان پوشه آن برای سوال مربوطه باید استفاده کنید.

به عنوان مثال برای دانشجویی که رقم آخر شماره دانشجویی برابر ۴ دارد و سؤال شماره ۳ را می‌خواهد حل کند داریم:

$$\begin{cases} a = 4 \\ q = 3 \end{cases} \rightarrow m = a + q = 7 \rightarrow f = 7 \% 5 = 2 \rightarrow \text{for } Q3 \text{ use Sys2 folder}$$

۳. برای سؤالات ۱ تا ۴ از سری A و برای سؤال ۵ از سری B در هر مجموعه داده استفاده گردد.

در هر کدام از دسته داده A و B، بردارهای ورودی x و بردار خروجی y وجود دارند. ماتریس X متشکل از بردارهای ورودی، به عنوان ماتریس ورودی در نظر بگیرید. در هر مرحله، می‌بایست مجموعه داده را به صورت کاملاً تصادفی و با نسبت ۷۰ به ۳۰، به دو دسته آموزش^۲ (برای تخمین رگرسورها θ) و آزمون^۳ (برای ارزیابی مدل) دسته بندی کنید.

^۱ Dataset

^۲ Train

^۳ Test

توضیحات

در تمامی بخش‌ها

- دقت تخمین
- دقت مدل
- قدرت نویز (واریانس نویز)

را به دست آورده و نمودارهای

- Box Plot
- Error Bar

را به همراه Confidence Interval نیز ترسیم نمایید. همچنین خروجی واقعی و تخمین زده شده را نیز در یک نمودار ترسیم نمایید.

سوال (۱)

بر اساس مقادیر مختلف N (تعداد مشاهدات) و n (درجه چند جمله‌ای)، موارد زیر را بدست آورده و نتایج را تحلیل، تفسیر و مقایسه نمایید. (حداقل ۳ حالت برای N و ۳ حالت برای n در نظر بگیرید که مجموعاً حداقل ۹ حالت بررسی شود)

الف) رگرسیونهای مدل را با استفاده از روش LS تخمین بزنید.

ب) حال در این قسمت، به جای تقسیم تصادفی داده‌ها به دو دسته آموزش و آزمون، این بار هفتصد داده اول را به عنوان داده‌های آموزش و بقیه را به عنوان داده‌های آزمون در نظر گرفته و مجدداً بخش قبلی را تکرار نمایید.

ج) نتایج را مقایسه کنید و نتیجه خود را تحلیل کنید.

سوال ۲)

فرض کنید ساختار سیستم، چند جمله‌ای با درجه ۴ (تعداد رگر سور برابر با ۴) است. برای هر کدام از توابع جریمه خطای مدل I_1, I_2, I_3 به صورت زیر، موارد زیر را انجام دهید.

الف) پارامترهای مدل‌های مختلف را بازیابی کرده و نتایج را تحلیل و مقایسه نمایید. (برای روش Regularization (I_3) $\alpha = 0.1$ در نظر بگیرید)

ب) برای تخمین پارامترها، در روش Regularization (I_3) اثر تغییرات پارامتر تنظیم را بررسی کرده و پارامتر تنظیم بهینه را با استدلال انتخاب کنید.

$$\begin{cases} I_1 = \sum |e_i| \\ I_2 = \text{Max}(e) \\ I_3 = \sum e_i^2 + \alpha(\theta_1^2 + \theta_2^2 + \theta_3^2) \end{cases}$$

سوال ۳)

- الف) با استفاده از Ridge Regression، رگرسورهای مهم را انتخاب کنید.
- ب) با استفاده از روش OLS رگرسورهای مهم را انتخاب کنید.
- ج) با استفاده از روش انتخاب رو به جلو^۱ و روش حذف عقب^۲ رگرسورها را انتخاب کنید.
- د) آیا ساختارهای به دست آمده از بخش‌های مختلف یکسان است؟ نتایج به دست آمده را تحلیل کنید.

^۱ Forward Selection
^۲ Backward Elimination

سوال ۴

از تخمین RLS با شرایط اولیه مختلف استفاده کنید. (۳ تکرار کافی است) نرخ همگرایی تدریجی $\theta(t)$ و $P(t)$ را رعایت کنید. در مورد تأثیر شرایط اولیه مختلف $\begin{Bmatrix} \theta(0) \\ P(0) \end{Bmatrix}$ بحث کنید.

سوال ۵)

مجموعه داده جدید را در نظر بگیرید اکنون سعی کنید سیستم را با استفاده روش‌های زیر شناسایی کنید (قابل توجه است در این سیستم جدید برخی از رگرورها متغیر با زمان در نظر گرفته شده‌اند).

الف RLS

ب روش Reset کواریانس

ج RLS با Forgetting Factor

د Kalman Filter

(برای فاکتور فراموشی و ماتریس Q ، چندین حالت مختلف را در نظر گرفته و نهایتاً آن‌ها را به طور مناسب انتخاب کرده و مسئله را حل کنید تا بتوانند مؤثرترین نتایج خود را نشان دهند).

و نتایج به دست آمده را تحلیل و مقایسه کنید.