



دانشگاه صنعتی خواجه نصیرالدین طوسی

به نام خدا

دکتر علیاری

شناسایی سیستم‌ها ۱۴۰۲

موضوع تمرین: شناسایی سیستم‌های خطی ایستا

تمرین سری اول - مهلت تحویل ۱۲ آبان ماه ۱۴۰۲



گروه پژوهشی ایک

بخش ۱: سؤالات تحلیلی

۱. الف) در روش Regularization بایاس و دقت تخمین پارامتر θ را محاسبه کنید. در مقایسه با حالت unregularized تغییرات را چگونه توجیه می‌کنید؟
- ب) (امتیازی) - انتظار دارید کدام یک از پارامترها بعد از اعمال Regularization به صفر میل کند؟
۲. اگر ماتریس هسیان مربوط به تابع هزینه حداقل مربعات به صورت $H = \text{diag}(80, 0, 0.1, 6)$ باشد:
 - الف) آیا پاسخ یکتا است؟
 - ب) کدام پارامتر از دقت تخمین بهتری برخوردار است؟
 - ج) حساسیت خطا نسبت به کدام پارامتر کم‌تر است؟
۳. اگر در یک مسئله LS، پارامترهای به یکدیگر وابسته باشند (برای مثال $\theta_1 = \theta_2 + \theta_3$)، چه راهی پیشنهاد می‌کنید؟
۴. معادله دیفرانسیلی رو به رو را در نظر بگیرید:

$$\ddot{x} - 3\dot{x} + 2x = 0$$
 اگر مقادیر $x(t)$ در دست باشد، چگونه می‌توان $x(0)$ و $\dot{x}(0)$ را به دست آورد؟
۵. (امتیازی) در مورد الگوریتم QR-decomposition-based RLS تحقیق کنید.

بخش ۲: سؤالات شبیه‌سازی

سؤال اول:

برای تخمین سیستم ورودی-خروجی معادله (۱) را که به صورت چندجمله‌ای است، در نظر بگیرید:

$$y = -1.5 - 0.8u + 0.01u^3 - 0.65u^5 + 2.25u^6 - 1.7u^8 + n \quad (1)$$

فرض بر این است که سیستم فوق برای شما ناشناخته است. بنابراین تنها برای تولید داده خروجی واقعی از سیستم فوق استفاده کنید! مدل رگرسیون خطی که در معادله (۲) آمده، تنها اطلاعات شما از سیستم است.

$$\hat{y} = \sum_{i=0}^n \theta_i u^i \quad (2)$$

۱۰۰۰ داده با توزیع نرمال بین $[0,1]$ تولید نموده و آن را به نسبت ۰,۷۵ و ۰,۲۵ به دو دسته آموزش و تست تقسیم کنید. نویز را با توزیع نرمال با سه نوع واریانس کم، متوسط و زیاد تولید نمایید. (این سه مقدار واریانس باید قابل توجه باشند و از هر مقدار دلخواه استفاده نکنید)

۱. برای تخمین LS داده‌های تولیدشده را به ده قسمت صدتایی تقسیم کرده و تخمین خود را انجام دهید ($K = 10, N = 100$). تعداد رگرسورها را برابر با ۱۰ در نظر بگیرید.

ا. خروجی تخمین و واقعی را برای داده‌های تست در یک نمودار رسم نموده و به تحلیل آن با توجه به سه مقدار واریانس نویز بپردازید. آیا تخمین انجام‌شده درست است؟ معیار یا دلیل خود را ذکر کنید. پارامترهای تخمین زده‌شده از نویز با سه مقدار واریانس چه تأثیری می‌گیرند؟
ب. نمودار Error Bar را ترسیم و تحلیل نمایید. (آیا برای محاسبه Error Bar به واریانس نویز نیاز است؟ در صورت مثبت بودن پاسخ آن را چگونه به دست می‌آورید؟) تحلیل خود را برای ابتدا و انتهای نمودار عنوان نمایید.

ج. تأثیر تعداد مشاهدات را با انتخاب‌های مختلف بررسی کنید.
د. فرض کنید تعداد رگرسورها نامشخص است ($n \neq 10$). چه مکانیزمی را برای انتخاب مقدار بهینه n اتخاذ می‌کنید.

ه. آیا می‌توان بدون تخمین پارامترها، مهم‌ترین رگرسور را انتخاب کرد؟
۲. داده‌ها را برای آموزش از بازه $[0,0.8]$ انتخاب کنید، اما داده‌های تست را از همان بازه اصلی انتخاب نمایید. نمودارهای مربوط به سؤال (۱) بخش دوم را ترسیم نموده و نتایج را تحلیل کنید. آیا تغییری در نمودار Error Bar ایجاد می‌شود؟ تحلیل خود را بنویسید.

۳. سؤال (۱) بخش دوم را به کمک پنجره لغزشی انجام دهید. آیا تغییری در نتایج حاصل می‌شود؟
۴. سؤال (۱) بخش دوم را اعمال تخمین Weighted LS انجام دهید. در انتخاب وزن‌ها چه معیاری را در نظر گرفته‌اید؟ دلیل خود را ذکر کنید و معیار خود را حداقل با یک معیار دیگر مقایسه کنید.
۵. فرض کنید داده‌های انتخاب‌شده برای آموزش به صورت بهنگام^۱ تولید شوند و ساختار مدل شناخته‌شده باشد. با کمک روش RLS پارامترهای مدل را تخمین بزنید. نمودارهای همگرایی پارامترها و نحوه همگرایی آن‌ها و همگرایی مقادیر ویژه ماتریس کوواریانس را ترسیم نموده و تحلیل نمایید. مقدار اولیه ماتریس کوواریانس چه تأثیری دارد؟

۶. برای داده‌های بدون نویز روش B.E و F.S را برای انتخاب رگرسورهای مهم به کار بگیرید. آیا نتایج حاصل از دو روش یکسان است؟ نتایج دلیل خود را بیان نمایید.

¹ Online

۷. در ادامه سؤال (۶) روش OLS را توضیح داده و با استفاده از آن، B.E و F.S را برای انتخاب رگرسورهای مهم به کار بگیرید. آیا نتایج با سؤال (۶) یکسان است؟ دلیل خود را عنوان نمایید.
۸. روش Ridge Regression را پیاده‌سازی کرده و رگرسورهای مهم را از غیر مهم تمییز دهید.

سؤال دوم:

در سیستم معرفی شده در معادله (۱) فرض کنید پارامتر θ_3 به جای 0.1 به صورت معادله (۳) باشد:

$$\theta_3 = 0.5 + \tanh(g \times (t - 400)) \quad (3)$$

که در آن $g = 1$ است.

۱. با استفاده از روش RLS نمودار همگرایی پارامترها و نحوه همگرایی و مقادیر ویژه ماتریس کوواریانس را رسم نموده و سپس نتایج خود را تحلیل نمایید.
۲. پارامتر g را تغییر دهید. آیا از روش RLS می‌توان استفاده نمود؟ دلیل خود را ذکر کنید. ($g = 1, 0.01$)
۳. روش FFRLS را بکار گرفته و پارامتر مناسبی برای فاکتور فراموشی در نظر بگیرید. چه رابطه‌ای بین پارامتر فاکتور فراموشی و دقت شناسایی وجود دارد؟ دلیل و نتایج خود را تحلیل نمایید.
۴. با استفاده از روش فیلتر کالمن مدل به دست آمده را گزارش کنید. همگرایی پارامترهای را تحلیل نموده و خروجی تخمین را با خروجی مدل را مقایسه کنید. برای ارزیابی درست بودن مدل از چه ابزاری استفاده می‌کنید آن را توضیح داده و مدل خود را ارزیابی نمایید.

سؤال سوم (اختیاری):

مدل رگرسیون خطی با سه رگرسور را به صورت معادله (۴) در نظر بگیرید:

$$\hat{y} = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \quad (4)$$

- فرض کنید $x_3 = x_1 + x_2 + n$ که در آن n نویز سفید گاوسی با میانگین صفر و واریانس σ^2 است. برای مقادیر کوچک σ وابستگی خطی شدید بین x_1, x_2 و x_3 وجود خواهد داشت.
۱. با انتخاب مقادیر θ به صورت دلخواه (منطقی) داده‌های ورودی-خروجی را به نحو مناسب ایجاد کنید.
 ۲. تخمین LS را برای سه مقدار کم و زیاد و متوسط σ انجام دهید. خروجی تخمین و واقعی را برای داده‌های تست در یک نمودار رسم نموده و تحلیل کنید.
 ۳. آیا مشکلی در تخمین وجود دارد؟ در صورت مثبت بودن پاسخ، چگونه می‌توان مشکل را حل کرد؟

به موارد زیر توجه کنید:

- تمرینات همراه با فایل گزارش و کلیه فایل‌های شبیه‌سازی در قالب یک فایل zip باید در سایت آپلود شود. نام فایل حتما باید شامل نام و شماره دانشجویی شما باشد.
- انجام شبیه‌سازی با هر زبان برنامه‌نویسی‌ای مجاز است.
- توضیحات مربوط به شبیه‌سازی و همچنین حل تحلیلی تمرین‌ها را به صورت PDF تبدیل نمایید و همراه با فایل‌های شبیه‌سازی در قالب یک فایل zip با نام و شماره دانشجویی ارسال کنید.
- به موعدهای تحویل تمرین دقت نمایید؛ به تمرین‌هایی که بعد از موعدهای تحویل داد شوند، به ازای هر روز ۲ⁿ نمره کسر می‌شود. (بدلیل حجم بالای تمرینات امکان تمدید موعدهای تحویل وجود ندارد).
- هر تمرین بر اساس گزارش ارسالی و ارائه شما ارزیابی می‌شود.
- استفاده از تمرینات و ایده‌گرفتن از تمرینات سال‌های گذشته مانعی ندارد. اما لازم است تا به تمام آنچه ارسال می‌کنید تسلط داشته باشید. در صورت کپی کردن و عدم تسلط به فایل‌های ارسالی نمره‌ای به شما تعلق نخواهد گرفت.
- تمرینات به صورت انفرادی است. به حل‌های مشابه شامل گزارش‌ها و فایل‌های شبیه‌سازی نمره‌ای تعلق نخواهد گرفت.

موفق باشید

مؤذنی - محمدزاده