

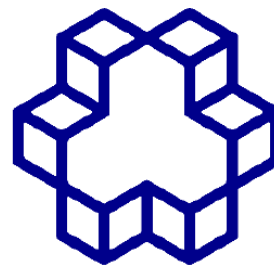


آزمایشگاه تشخیص و شناسایی خطا

به نام خدا

درس تشخیص و شناسایی عیب

تمرین سری اول



دانشگاه صنعتی خواجه نصیرالدین طوسی

موعد تحویل: ۱۴۰۱/۱۲/۲۸

سوال شماره یک - ۱۵ نمره

۱) یک مسأله تشخیص دوکلاسه با مجموعه‌های سالم ($\omega = 1$) و معیوب ($\omega = 2$) را در نظر بگیرید.

۱-الف) ابتدا فرض کنید که یک ویژگی (بعد) داریم و توابع چگالی احتمال این دو کلاس برابر با $N(0, \sigma^2)$ و $N(1, \sigma^2)$ باشند. نشان دهید آستانه τ ، که میانگین ریسک را کمینه می‌کند به صورت زیر محاسبه می‌شود:

$$\tau = \frac{1}{2} - \sigma^2 \ln \frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}$$

توضیحات و راهنمایی: فرض می‌کنیم که $\lambda_{11} = \lambda_{22} = 0$ باشد. اگر تغییری را در رابطه نیاز می‌بینید اعمال کنید.

۱-ب) فرض کنید که دو ویژگی (بعد) به صورت $\mathbf{x} = (x_1, x_2)$ داشته باشیم و چگالی شرطی دو کلاس، $p(x | \omega = 1)$ و $p(x | \omega = 2)$ ، به صورت توزیع‌های گوسی دوبعدی با نقاط مرکزی $(4, 11)$ و $(10, 3)$ و ماتریس کواریانس $\Sigma = 3\mathbf{I}$ (ماتریس همانی است) باشد. با در نظر گرفتن احتمالات پیشین به صورت $P(\omega = 1) = 0.6$ و $P(\omega = 2) = 0.4$:

۱-ب-۱) با استفاده از قانون تصمیم بیز، دو تابع جداکننده $g_1(x)$ و $g_2(x)$ را تشکیل دهید و یک معادله برای مرز تصمیم‌گیری $g_1(x) = g_2(x)$ بنویسید.

۱-ب-۲) مرز تصمیم با تغییر $P(\omega = 1)$ و $P(\omega = 2)$ چگونه تغییر خواهد کرد؟ با تغییر کواریانس چطور؟ تحلیل کنید.

۱-ج) برنامه‌ای بنویسید که ۱۰۰ نقطه از هر یک از دو توزیع را نمونه‌برداری کرده و سپس آن‌ها را به همراه مرز تصمیم روی فضای ویژگی دوبعدی رسم کند.

سوال شماره دو - ۱۰ نمره

۲) داده برداری از سامانه‌ای نشان می‌دهد که سامانه در حالت عادی (سالم) از توزیع گاما با پارامترهای $\alpha = 16$ و $\beta = 8$ و در حالت معیوب از توزیع گاما با پارامترهای $\alpha = 3$ و $\beta = 4$ پیروی می‌کند.

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-\beta x} \beta^\alpha}{\Gamma(\alpha)} \quad \text{for } x > 0 \quad \alpha, \beta > 0$$

۲-الف) هر دو تابع چگالی احتمال را در یک نمودار رسم و میزان خطای تصمیم‌گیری را روی شکل مشخص کنید.

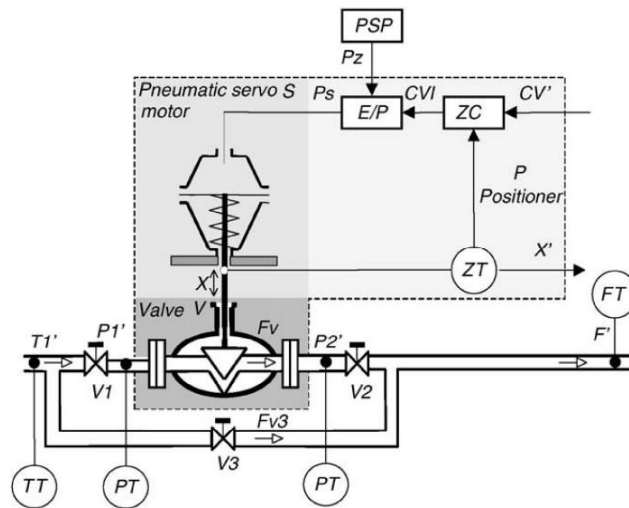
۲-ب) فرض کنید احتمال پیشین دسته‌ها برابر هستند. تابع جداساز را برای کمینه‌کردن احتمال خطای بیز به‌دست آورید.

۲-ج) میزان ریسک را چنان تعیین کنید که عدد ۱.۲ عادی (سالم) تشخیص داده شود.

سوال شماره سه - ۸۵ نمره

۳- مجموعه داده DAMADICS که مربوط به سیستمی به شکل زیر است را در نظر بگیرید.

CV	process control external signal
CVI	internal current acting on E/P unit
E/P	electro-pneumatic transducer
F	main pipeline flow rate
Fv	control valve flow rate
Fv3	actuator by-pass pipeline flow rate
FT	flow rate transmitter
P	positioner
P1, P2	pressures on valve: inlet and outlet
Ps	E/P transducer output pressure
PSP	positioner supply pressure unit
PT	pressure transmitter
Pz	positioner air supply pressure
S	pneumatic servo-motor
Tl	liquid temperature
TT	temperature transmitter



هر فایل با پسوند txt. که در مجموعه داده DAMADICS موجود است، از یک روز مشخص به‌دست آمده که نام‌گذاری عنوان فایل‌های موجود در این مجموعه داده نیز بر اساس همین موضوع صورت پذیرفته است. برای مثال فایل 20112001.txt حاوی داده‌های روز ۲۰ نوامبر سال ۲۰۰۱ است. این داده‌ها در قالب یک ماتریس با ۸۶۴۰۰ ردیف و ۳۳ ستون در دسترس هستند. ردیف‌ها بیان‌گر گام‌های زمانی و ستون‌ها مربوط به متغیرهای صنعتی‌ای هستند که در [این لینک](#) شرح داده شده‌اند.

۳-الف) در صنعت معمولاً داده‌ها بدون پیش‌پردازش خاصی در اختیار شما قرار می‌گیرند. مجموعه داده مربوط به این سوال که DAMADICS نام دارد و از طریق [این لینک](#) در دسترس است هم از این قاعده مستثنی نیست. بنابراین، شما نیاز دارید تا با مهارت برنامه‌نویسی خود داده‌ها را به صورتی که برای کارهای پردازشی مناسب باشد، آماده کنید. با انجام این قسمت از سوال، به صورت گام به گام و کاربردی با یکی از انواع آماده‌سازی و پیش‌پردازش مجموعه داده آشنا می‌شوید.

فایل‌های این مجموعه داده را از طریق [این لینک](#) یا [این لینک](#) دانلود کنید و نام فایل فشرده را به Data.zip تغییر دهید. سپس، برنامه‌ای بنویسید که تمامی فایل‌های txt. موجود در پوشه مجموعه داده Data.zip را به فایل‌هایی با فرمت‌های csv. و mat. تبدیل کند و نتایج را در پوشه‌های جدیدی ذخیره کند. برنامه را به صورتی بنویسید که به صورت خودکار ستون اول که حاوی اطلاعات گام‌های زمانی است را حذف کند.

برای آماده‌سازی داده‌ها می‌توانید از دستورات موجود در [این لینک](#) استفاده کنید؛ اما انتظار می‌رود که روند کلی کدها را فهمیده باشید. اگر به زبان پایتون مسلط نیستید به سادگی و با جستجو در اینترنت با دستورات به کار برده شده آشنا شوید.

۳-ب) داده‌های تاریخ ۱۷ نوامبر ۲۰۰۱ را در نظر بگیرید. عیوبی در گام‌های زمانی مشخص و مطابق جدول زیر به سیستم وارد می‌شوند:

Item	Fault	Sample	Date
1	f18	54600 - 54700	November 17, 2001
2	f16	56670 - 56770	November 17, 2001
3	f17	53780 - 53794	November 17, 2001
4	f17	54193 - 54215	November 17, 2001
5	f19	55482 - 55517	November 17, 2001
6	f19	55977 - 56015	November 17, 2001
7	f19	57030 - 57072	November 17, 2001
8	f16	57475 - 57530	November 17, 2001
9	f16	57675 - 57800	November 17, 2001
10	f19	58150 - 58325	November 17, 2001

داده‌ها را در پنج دسته عادی (سالم)، عیب ۱ (f16)، عیب ۲ (f17)، عیب ۳ (f18) و عیب ۴ (f19) در نظر بگیرید (ترتیب و شماره دسته‌ها به انتخاب شماست). شاخص‌های آماری که در درس فراگرفته‌اید (مانند میانگین، کواریانس و غیره) را برای هریک از این دسته‌ها محاسبه کنید و برای هر دسته یک توزیع گوسی تخمین بزنید.

۳-ج) برنامه‌ای بنویسید که طبقه‌بند بیزی را روی این مجموعه داده اعمال کند. برای ساختار خود آستانه گذار مناسبی طراحی کنید و نتایج را به صورت ماتریس درهم‌ریختگی و حداقل سه شاخص ارزیابی فراگیری شده در درس نشان دهید. همچنین با در نظر گرفتن شاخص‌های آماری مربوط به هر دسته، درایه‌های ماتریس درهم‌ریختگی را تحلیل کنید.

۳-د) سوالات «۳-ب» و «۳-ج» را برای حالتی که فقط ۲۰۲ داده از دسته عادی (سالم) را در نظر گرفته‌اید، تکرار کنید. نتیجه را با حالت قبلی مقایسه کنید. تحلیل کنید که چه چیزی باعث این تغییرات شده است.

۳-ه) آیا می‌توانید با استفاده از تحلیل ریسک در سوال «۳-ج»، خطای دسته‌ها را کاملاً صفر کنید؟ اگر اطمینان ما به دانش خبره (فرد داده‌گیرنده) ۸۵ درصد باشد، ماتریس وزنی ریسک چگونه تغییر می‌یابد؟ می‌توانید تعداد دسته‌ها را هم کاهش دهید.

۳-و) با استفاده از داده‌های روز ۹ نوامبر ۲۰۰۱ به عنوان داده تست، عمل کرد ساختار طراحی شده خود را آزمایش و نتیجه را گزارش و تحلیل کنید.

Item	Fault	Sample	Date
1	f16	57275 - 57550	November 9, 2001
2	f18	58830 - 58930	November 9, 2001
3	f18	58520 - 58625	November 9, 2001
4	f16	60650 - 60700	November 9, 2001
5	f16	60870 - 60960	November 9, 2001

۳-ز) روش‌های مختلفی برای برچسب‌گذاری داده‌ها وجود دارند. از جمله این روش‌ها می‌توان به one-hot، باینری و یا عددی اشاره کرد. ضمن اشاره به این که از چه روشی برای کد کردن دسته داده‌ها استفاده کرده‌اید، تحقیق کنید که شیوه برچسب‌گذاری one-hot در کارهایی که از روش‌های هوشمند استفاده می‌کنند چه مزیتی ایجاد می‌کند.

۳-ح- امتیازی) سوال «۳-ج» را با طبقه‌بندهای اقلیدسی و ماهالانوبیس تکرار کنید و با مقایسه نتایج و جستجو در اینترنت، تحقیق کنید که مزایا و معایب هریک از این سه طبقه‌بند چیست، و هر کدام تحت چه شرایطی (در داده‌ها یا مسأله) بهتر عمل می‌کنند.

راهنمایی: مفهوم این دو طبقه‌بندی در حوالی صفحه ۳۰ کتاب مرجع (Theodoridis) معرفی شده و توابع برنامه‌نویسی شده آن‌ها در صفحه ۸۲ کتاب آورده شده است.

۳-ط- امتیازی) فرض کنید یکی از متغیرهای فرآیند این سامانه، دو وضعیت عادی (سالم) و معیوب را با توزیع گوسی ای تجربه می کند که عیب، تنها واریانس را افزایش خواهد داد و تغییری در میانگین ایجاد نخواهد کرد. برای طراحی آستانه تمایز دو وضعیت عادی و معیوب، فیلتر واریانس متحرکی با طول بهینه m به صورت زیر را به کار می برند ($T - L$ به اندازه کافی بزرگ است):

$$y_t = \frac{1}{m} \sum_{i=t}^{L+t} (x_i - \mu)^2, t = 1, \dots, T - L$$

۳-ط-۱) برای متغیر فیلترشده y_t ، آستانه تمایز را به نحوی تعیین کنید که خطای طبقه بند بیز کمینه گردد.

۳-ط-۲) مکانیزمی طراحی کنید تا به کمک تنها یک آستانه هر پنج وضعیت از هم تفکیک شوند.

ضمن عرض سلام و خداقوت، لطفاً برای ارسال تمرین ها به نکات زیر توجه فرمایید:

نوشتن تمرین به صورت مرتب و در قالب LaTeX مدنظر، درصدی متغیر به نمره شما اضافه خواهد کرد.

هر مطلبی در خصوص تمرین ها و نمرات را از طریق رایانامه faultcourse@gmail.com، در میان بگذارید.

در سوالاتی که از شما تحلیل خواسته شده، عمق تحلیل ها و بحث روی حالات و علل مختلف اهمیت بالایی دارد.

اگر فرآیند آماده سازی داده ها برای سوال سوم طولانی شد، ایمیل دهید تا برای شما ارسال شود؛ هر چند توصیه نمی شود!

هر تمرین ۱۰۰ نمره دارد و در صورت تأخیر، 2^n نمره (n تعداد روز تأخیر) از مجموع نمره تمرین شما کاسته خواهد شد.

«در مسیر حق موفق و پیروز باشید.»