



تمرین : سری ۹

مهلت تحویل ۱۶م خردادماه

مدرس: دکتر محمد صفدری

- پاسخ‌های خود را در قالب یک فایل فشرده در درس‌افزار شریف ارسال کنید.
- در پوشه‌ی پاسخ لازم است به ازای هر سوال یک پوشه ایجاد کنید و کدهای مربوط به آن سوال را در آن قرار دهید.
- در پوشه‌ی مربوط به هر سوال، پوشه‌ای حاوی نمودارهای و توضیحات تکمیلی (در صورت لزوم) بگنجانید.
- پیش از اجرای هر بخش seed را برابر ۱۰۰ قرار دهید.
- سوالات خود پیرامون تمرینات را با دستیاران درس مطرح کنید.

تمرینات نظری

یک تعمیم طبیعی

همانطور که در درس دیدید، رگرسیون ستیغی^۱ از تعدیل $\|\beta\|_1$ و رگرسیون لاسو از تعدیل $\|\beta\|_2$ استفاده می‌کند. رگرسیون “زیبا” را به طوری تعریف می‌کنیم که عبارت تعدیل در مساله‌ی بهینه‌سازی متناظر با آن به فرم ترکیب خطی این دو نُرم است. به عبارت دیگر، تابع هزینه

$$L(\lambda_1, \lambda_2, \beta) = \|y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2$$

است و بردار ضرایب بهینه β^* مقدار L را به ازای مقادیر مشخص و ثابت λ_1 و λ_2 کمینه می‌کند.

الف) نشان دهید مساله‌ی دوگان متناظر با این بهینه‌سازی معادل است با

$$\beta^* = \arg \min_{\beta} \|y - X\beta\|_2^2 \quad s.t. \quad (1 - \alpha) \|\beta\|_1 + \alpha \|\beta\|_2^2 \leq t$$

^۱Ridge Regression

^۲Regularization

به ازای ثوابت α, t . در ادامه رابطه‌ی λ_1 و λ_2 با t, α را بدست آورید.

ب) مشابه تعبیر بیزی برای رگرسیون‌های ستیغی و لاسو، رگرسیون زیبا را هم به صورت بیزی تعبیر کنید. به طور مشخص شما باید توزیع پیشین مناسبی برای بردار ضرایب معرفی کنید که بروز رسانی بیزی با استفاده از داده معادل با حل مساله‌ی بهینه‌سازی مذکور باشد. (راهنمایی: فرآیند حل مساله‌ی مشابه برای رگرسیون ستیغی و لاسو به طور کامل در جزوه‌ی استاد تشریح شده است.)

۱.۰ PCA روی کاغذ

در این بخش می‌خواهیم تحلیل اجزای اساسی^۳ را مرور کنیم برای سادگی فرض کنید که بجای مجموعه‌ای از داده‌ها، خود توزیع به ما داده شده است.

الف) با این فرض که تابع چگالی را در دست داریم، صورت‌بندی مناسبی برای پیدا کردن متوالی راستاهای اساسی معرفی کنید. (راهنمایی: صورت بندی مطلوب به این صورت است که یک مساله‌ی بهینه‌سازی برای بردار اساسی k ام معرفی می‌شود و در قیود این بهینه‌سازی وابستگی به $k - 1$ بردار قبلی وجود دارد.)

فرض کنید که تابع چگالی نرمال توأم با پارامترهای $\mu = 0, \Sigma \in \mathbb{R}^{n \times n}$ داده شده است.

ب) راستای اساسی اول را برحسب Σ معرفی کنید.

حاصل تصویر کردن تابع چگالی اولیه روی صفحه‌ی عمود بر راستای اساسی اول، یک نرمال توأم در یک بعد پایین‌تر خواهد بود. همچنین، ماتریس کواریانس توزیع متناظر با این تصویر کردن بر اساس Σ قابل محاسبه است (فرم بسته وجود دارد). بنابراین با پیدا کردن مکرر راستای اساسی اول و تصویر کردن روی صفحه‌ی عمود متناظر می‌توان تمامی راستاهای اساسی را پیدا کرد.

ج) به طور مستقیم تمامی راستاهای اساسی را براساس ماتریس Σ معرفی کنید. در این سوال مجاز به استفاده از حقایق نابديهی نیستید و در صورتی که ادعایی می‌کنید اثبات آن لازم است. (راهنمایی، ابتدا نشان دهید سطوح تراز تابع چگالی نرمال، بیضی‌گون^۴ هستند و ماتریس متناظر با این بیضی‌گون را بر حسب Σ معرفی کنید. سپس نشان دهید که قطرهای اصلی این بیضی‌گون‌ها همان راستاهای اساسی هستند. نهایتاً رابطه‌ی بین قطرهای اصلی این بیضی‌گون و ماتریس Σ را تحقیق کنید.)

^۳Principal Component Analysis

^۴Ellipsoid

تمرینات عملی

الف) روش PLS را پیاده‌سازی کنید. ورودی ماتریس حقیقی مقدار ویژگی‌هاست که هر ستون یک ویژگی و هر سطر یک مشاهده مستقل را نشان می‌دهد. همچنین، بردار متغیر هدف با طولی به اندازه‌ی تعداد سطرهای ماتریس ویژگی‌ها داده شده‌است. با استفاده از روش کراس ولیدیشن مقدار بهینه‌ی کاهش بعد را تعیین کنید. برای آزمایش کارایی روش خود از داده‌ی abalone استفاده کنید. (توضیح: در داده‌ی abalone متغیر هدف تعداد حلقه‌ها، متناظر با سن صدف است.)

ج) با استفاده از کتابخانه‌ی glmnet رگرسیون لاسو و ستیجی به داده‌ی abalone (در تمرینات قبل و کلاس‌های حل تمرین خیلی با آن سر و کار داشتیم) برازش کنید و پارامترهای تعدیل را با استفاده از بوت استرپ بهینه کنید. مراجعه به بخش Lab در کتاب توصیه می‌شود. (توضیح: در داده‌ی abalone متغیر هدف تعداد حلقه‌ها، متناظر با سن صدف است.)

ب) ابتدا متغیر جنسیت را از مجموعه داده‌ی abalone کنار بگذارید و مرکز ثقل داده را صفر کنید. با استفاده از تابع prcomp موجود در کتابخانه‌ی stats دو راستای اساسی اول را در مجموعه داده‌ی اصلاح شده پیدا کنید. نقاط داده را با برچسب جنسیت (نقاط متعلق هر گروه جنسیتی با یک رنگ) در این فضای دو بعدی متناظر با این دو بردار اساسی تصویر کنید.