دانشگاه تهران
دانشکده علوم ریاضی، آمار و علوم کامپیوتر

روش‌های خطی طبقه‌بندی

پروژه کارشناسی آمار

پریسا علی محمدی

استاد راهنما

دکتر مرتضی امینی

خرداد ۱۳۹۶

با سپاس از ...

خرداد ۱۳۹۶

کلیه حقوق مادی مترتّب بر نتایج مطالعات، ابتکارات و نوآوری‌های ناشی از تحقیق موضوع این پایان‌نامه متعلّق به دانشگاه تهران است.

فهرست مطالب

فهرست تصاویر ..... پنج

فصل ۱ آنالیز تفکیک کننده‌ی خطی و درجه دوم ..... شش

۱.۱ آنالیز تفکیک کننده‌ی خطی ..... شش

۲.۱ آنالیز تفکیک کننده درجه دوم ..... نه

۳.۱ آنالیز تفکیک کننده منظم ..... نه

۴.۱ محاسبه‌ی LDA و QDA ..... ده

۵.۱ کاهش بعد در انالیز تفکیک کننده خطی ..... دوازده

۶.۱ ارتباط LDA و رگرسیون لوژستیک ..... دوازده

فصل ۲ معرفی بسته های نرم افزاری موجود در نرم افزار R ..... شانزده

۱.۲ مقدمه ..... شانزده

۲.۲ تحلیل مشخصه‌ی خطی (LDA) ..... شانزده

۳.۲ تحلیل مشخصه‌ی درجه‌ی دوم (QDA) ..... هجده

۴.۲ پیش‌بینی ..... هجده

فهرست تصاویر

۱.۱ شکل ۱ . . . . . هشت
۲.۱ شکل ۲ . . . . . ده
۳.۱ شکل ۳ . . . . . یازده
۴.۱ شکل ۴ . . . . . سیزده
۵.۱ شکل ۵ . . . . . سیزده
۶.۱ شکل ۶ . . . . . پانزده

فصل ۱

آنالیز تفکیک کننده‌ی خطی و درجه دوم

۱,۱ آنالیز تفکیک کننده‌ی خطی

برای طبقه‌بندی بهینه احتیاج به دانستن احتمالات پسین $Pr(G|X)$ داریم، که در آن G طبقه و X مشاهده است. فرض کنید $f_k(x)$ چگالی شرطی نقطه‌ی x در کلاس $G = k$ و π_k احتمال پیشین کلاس k باشد. به طوری که $\sum_{k=1}^{K} \pi_k = 1$.
در این صورت بنابر فرمول بیز داریم

$$Pr(G = k|X = x) = \frac{f_k(x)\pi_k}{\sum_{K=1}^{K} f_k(x)\pi_k} \tag{۱.۱}$$

در آنالیز تفکیک‌کننده‌ی خطی و درجه دو از چگالی گاوسی برای (f_k) استفاده می‌شود. با این حال ترکیب‌های انعطاف‌پذیرتری از چگالی‌های گاوسی، امکان مرزهای تصمیم غیرخطی را فراهم می‌آورد. هم چنین براورد چگالی ناپارامتری کلی برای چگالی هر کلاس بیشترین انعطاف‌پذیری را نتیجه می‌دهد. یک حالت خاص دیگر مدل‌های بیز ساده انگارانه است که در ان چگالی هر کلاس را حاصل‌ضرب چگالی‌های حاشیه‌ای در نظر می‌گیرند، به این معنا که فرض کنیم مشاهده ها در هر کلاس مستقل شرطی هستند.
حال چگالی گاوسی چند متغیره به صورت زیر ررا برای f_k در نظر بگیرید.

$$f_k(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \tag{۲.۱}$$

$$exp\frac{۱}{۲}(x-\mu_k)^T\Sigma^{-۱}(x-\mu_k) \tag{۳.۱}$$

در آنالیز تفکیک کننده خطی (LDA) فرض می کنیم کلاس ها ماتریس های واریانس کواریانس برابر دارند یعنی، $\Sigma_k = \Sigma$ برای $k = ۱, \cdots, K$.
به عنوان ملاکی از میزان درستی گروه بندی $G = k$ نسبت به گروه بندی $G = l$ برای مشاهده $X = x$، از لگاریتم نسبت درستنمایی به صورت زیر استفاده می کنیم

$$\log\frac{Pr(G=k|X=x)}{Pr(G=l|X=x)} = \log\frac{f_k(x)}{f_l(x)} + \log\frac{\pi_k}{\pi_l} \tag{۴.۱}$$

$$۲(\mu_l^T - \mu_k^T)\Sigma^{-۱}x - (\mu_l - \mu_k)^T\Sigma^{-۱}(\mu_l - \mu_k) + \log\frac{\pi_l}{\pi_k} \tag{۵.۱}$$

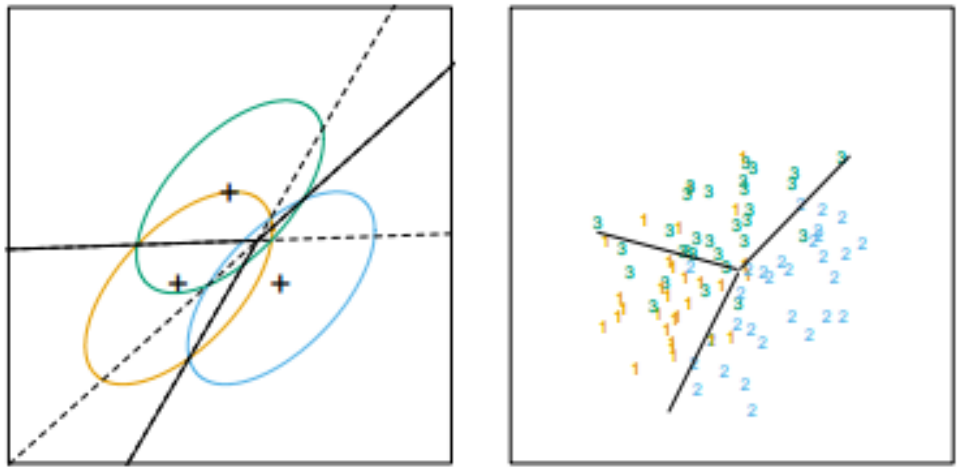
همانطور که می بینید، برابری ماتریس کواریانس گروه ها باعث حذف جملات درجه دوم از x میشود. بنابراین اگر از ملاک (۳/۱) برای طبقه بندی استفاده شود ، مرز های تصمیم ابر صفحه های P بعدی (بعد X)و به عبارت بهتر خطی خواهند بود. با توجه به (۳/۱) قرار می دهیم.

$$\sigma_k(x) = X^T\Sigma^{-۱}\mu_k - \frac{۱}{۲}\mu_k^T\Sigma^{-۱}\mu_k + \log\pi_k \tag{۶.۱}$$

و گروه بهینه برای x برابراست با $G(x) = argmax_k\sigma_k(x)$. در عمل پارامترهای π_k،σ μ_k، مجهول هستند. بنابر این این پارامتر ها با مقادیر برآورد های زیر جایگزین می‌شوند.

$$\widehat{\pi_k} = \frac{N_k}{N}, \tag{۷.۱}$$

$$\widehat{\mu_k} = \sum\frac{x_i}{N_k}, \tag{۸.۱}$$



شکل ۱.۱

پانل سمت چپ سه توزیع گوسین را نشان می دهد با کواریانس یکسان و میانگین متفاوت شامل خطوطی با چگالی ثابت و در هر مورد دربردارنده احتمال ۹۵ است. مرز تصمیم بیز بین هر جفت از کلاس ها نشان داده شده است و مرز تصمیم بیز جدا کننده هر ۳ کلاس با خطوط پررنگ نشان داده شده است و در شکل سمت راست نمونه ای ۳۰ تایی را می بینیم که همه از توزیع گاوسین امده اند و مرز تصمیم خطی LDA برازش داده شده است.

$$\hat{\Sigma} = \sum_{k=1}^{K} \sum \frac{(x_i - \mu_k)(x_i - \mu_k)^T}{N - K},. \tag{۹.۱}$$

با استفاده از ملاک (۴/۱) و قاعده بیان شده ، بین دو طبقه ۱و ۲ ، مشاهده x در طبقه ۲ گروه بندی می شود، اگر

$$X^T \Sigma^{-1}(\widehat{\mu_2} - \widehat{\mu_1}) > \frac{1}{2}\widehat{\mu_2}^T \Sigma^{-1}\widehat{\mu_2} - \frac{1}{2}\widehat{\mu_1}^T \Sigma^{-1}\widehat{\mu_1} + \log\frac{N_1}{N} - \log\frac{N_2}{N} \tag{۱۰.۱}$$

و در طبقه ۱ گروه بندی می شود، اگر

$$X^T \Sigma^{-1}(\widehat{\mu_2} - \widehat{\mu_1}) < \frac{1}{2}\widehat{\mu_2}^T \Sigma^{-1}\widehat{\mu_2} - \frac{1}{2}\widehat{\mu_1}^T \Sigma^{-1}\widehat{\mu_1} + \log\frac{N_1}{N} - \log\frac{N_2}{N} \tag{۱۱.۱}$$

بنابراین ابر صفحه جدا کننده این دوطبقه به صورت زیر است

$$\frac{1}{2}\widehat{\mu_2^T} \Sigma^{-1}\widehat{\mu_2} = \frac{1}{2}\widehat{\mu_1^T} \Sigma^{-1}\widehat{\mu_1} + \log\frac{N_1}{N} - \log\frac{N_2}{N} \tag{۱۲.۱}$$

۲,۱ آنالیز تفکیک کننده درجه دوم

در انالیز تفکیک درجه دوم (QDA) ماتریس واریانس کواریانس گروه ها یکسان در نظر گرفته نمی شود و بنابراین ملاک تفکیک درجه دوم در این حالت برابر است با

$$\sigma_k(x) = -\frac{۱}{۲}\log|\Sigma_k| - \frac{۱}{۲}(x-\mu_k)^T\Sigma_k^{-۱}(x-\mu_k) + \log\pi_k \qquad (۱۳.۱)$$

با استفاده از مرز تصمیم بین هر جفت مشاهده از کلاس k و l یک معادله درجه۲ است. شکل ۲.۱ (؟؟) مرز های تفکیک درجه دوم و خطی سه کلاس گوسی را نشان می دهد. در این حالت ابر منحنی جدا کننده درجه دوم دو طبقه مثلا ۱و۲ به صورت زیر است.

$$x^T(\widehat{\Sigma_۱}^{-۱}-\widehat{\Sigma_۲}^{-۱})x-۲x^T(\Sigma۱^{-۱}\widehat{\mu_۱}-\widehat{\Sigma_۲}^{-۱}\widehat{\mu_۲}) = \widehat{\mu_۲}\widehat{\Sigma_۲}^{-۱}\widehat{\mu_۲}-\widehat{\mu_۱}\Sigma_۱^{-۱}\widehat{\mu_۱}+\widehat{\log\left|\widehat{\Sigma_۲}\right|}-\log\left|\widehat{\Sigma_۱}\right|+\log\frac{N_۱}{N}-\log\frac{N_p}{N} \qquad (۱۴.۱)$$

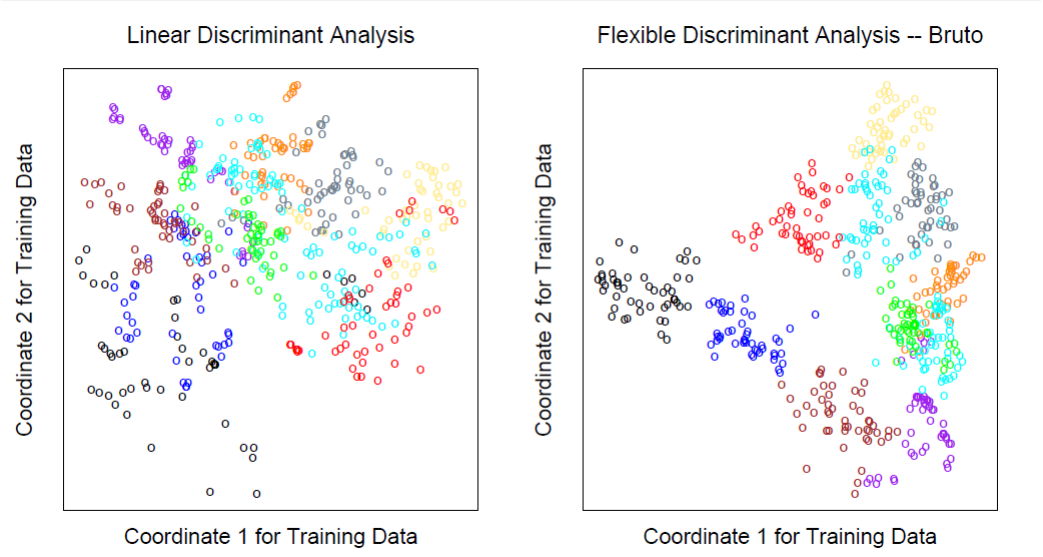
هر چه قدر اختلاف پراکندگی دو کلاس بیشتر باشد، ضریب قسمت درجه دوم و در نتیجه انحنای مرز تفکیک کننده به سمت کلاس با پراکندگی کمتر، بیشتر خواهد شد (تقعر به سمت کلاس با پراکندگی کمتر و تحدب به سمت کلاس با پراکندگی بیشتر).

۳,۱ آنالیز تفکیک کننده منظم

فریدمن(۱۹۸۹) یک توافق بین LDA و QDA مطرح کرد، که تلفیقی از کوواریانس های جداگانه در روش QDA کواریانس های مشترک در روش LDA را در نظر میگیرد. ماتریس کواریانس منظم در این روش به صورت زیر است.

$$\Sigma_k(\alpha) = \alpha\Sigma_k + (۱-\alpha)\Sigma, \qquad (۱۵.۱)$$

پارامتر تنظیم α می‌تواند بر پایه کارایی مدل داده‌های اعتبارسنجی و یا به وسیله ملاک های اعتبار سنجی متقابل انتخاب شود، که در ادامه توضیح داده می‌شود.
برای یک اصلاح مشابه اجازه می دهد که Σ به صورت زیر تفکیک شود.



شکل ۲.۱

شکل سمت چپ دسته‌ی اول متغیرهای کانونی LDA را برای داده‌های مصوت نشان می‌دهد. شکل سمت راست نگاشت حاصل از استفاده از $FDA/BRUTO$ برای برازش مدل را نشان می‌دهد که در آن مقادیر برازیده‌ی توابع رگرسیونی نمایش داده شده است. به وضوح جداسازی داده‌های مصوته در قالب ۱۱گروه بهبود یافته است.

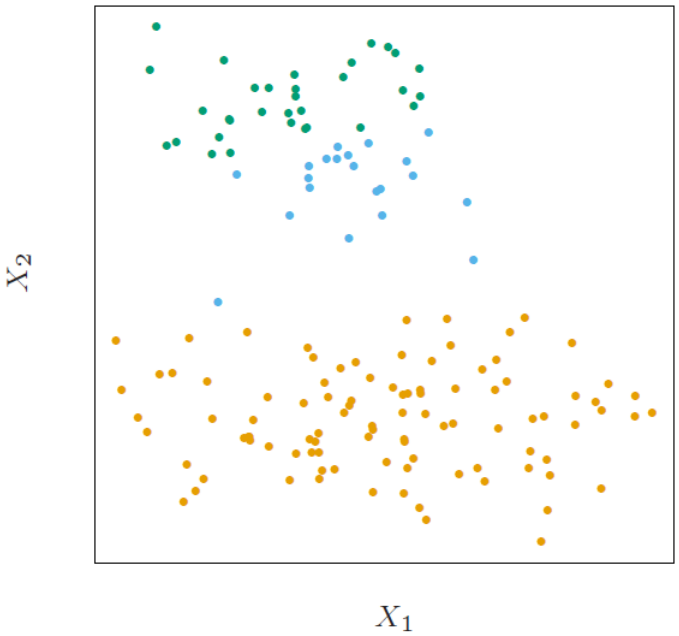
$$\Sigma(\gamma) = \gamma\Sigma + (1-\gamma)\,\sigma^2 I, \tag{۱۶.۱}$$

که در آن $\alpha \in [0,1]$.

جایگزینی Σ در (۱۰.۱) با $\Sigma(\gamma)$ منجر به یک خانواده کلی‌تر از ماتریس های کواریانس مانند $\Sigma(\gamma,\alpha)$ می شود که با جفت پارامتر γ و α اندیس گذاری شده است.

۴,۱ محاسبه‌ی LDA و QDA

محاسبه ی ابر صفحه ها و یا ابر منحنی های تفکیک کننده در روش LDA و QDA نیاز به محاسبات پیچیده مکرر دارد. به خصوص زمانی که بعد P بالا باشد، این محاسبات نیز سنگین تر خواهند شد. با این حال تجزیه منعامد ماتریس های کواریانس به صورت زیر به ما امکان درک ومحاسبه ساده ترباز طبقه بندی در این روش می دهد. فرض کنید ماتریس کواریانس Σ_k به صورت $\Sigma_k = U_k D_k U_k^T$تجزیه شده باشد، که در ان U_k ماتریس بردار های ویژه D_k و σ_k ماتریس قطری با اعضای قطر اصلی مقاذیر ویژهσ_k باشد. در این صورت با قرار دادن $x_k^* = D_k^{-\frac{1}{2}} U_K^T x_k$ در روشQDA و $x^\alpha = D^{-\frac{1}{2}} U^T x$ در روش LDA (با تجزیه $\Sigma = UDU^T$) داریم



شکل ۳.۱

داده‌های شبیه‌سازی شده در یک صفحه که توسط روش $K-Means$ به سه خوشه (آبی، نارنجی و سبز) تقسیم شده اند.

$$Cov(x_k^{\alpha}) = D_k^{\frac{-1}{2}} U_K^T D_k^{\frac{-1}{2}} = I \tag{۱۷.۱}$$

وبه طور مشابه $cov(x^*) = I$. با توجه به متعامد یکه بودن بردار های ویژه

$$U^T U = U_k^p U = I. \tag{۱۸.۱}$$

بنابراین برای داده های تبدیل یافته X^* کافی است، هر مشاهده با توجه به تنها فاصله اقلیدسیش از مرکز هرطبقه $\widehat{\mu_k}$ و احتمال پیشین ان طبقه $\widehat{\pi_k}$ طبقه بندی می شود، یعنی بر اساس

$$\sigma_k(x) = \log \widehat{\pi_k} - \frac{1}{2}\|x - \widehat{\mu_k}\|^2 \tag{۱۹.۱}$$

۵,۱ کاهش بعد در انالیز تفکیک کننده خطی

روش های معمول کاهش بعد به دنبال یافتن ترکیب های خطی مانند $y = \alpha x$ با بعد کمتر از x هستند، به طوری که دارای بیشترین پراکندگی باشند.با این حال در روش های طبقه بندی مانند LDA و QDA بهتر است این کاهش بعد به گونه ای باشد که پراکندگی بین طبقات حداکثر شود. با در نظر گرفتن یک طبقه بندی مشخص می توان پراکندگی کل Σ را به صورت زیر تفکیک کرد

$$T = \sum_{k=1}^{K}(x_i - \mu)^T(x_i - \mu) = \sum_{k=1}^{K}\Sigma(x_i - \mu_k)^T + \sum_{k=1}^{K}N_k(\mu_k - \mu)^T(\mu_k - \mu) = W + B, \quad (۲۰.۱)$$

که در ان W تغییرات داخل طبقات و B پراکندگی بین طبقات نامیده می شود. فیشر مساله کاهش بعد در LDA را به این صورت مطرح می کند که ضرایب خطی $y = \alpha x$.با پراکندگی داخل طبقات $\alpha^T W \alpha$ و پراکندگی بین طبقات $\alpha^T B \alpha$ را طوری بیابیم که $\alpha^T W \alpha$ حد اکثر و در نتیجه $\alpha^T W \alpha$ مینیمم شود. بنابراین مساله فیشر ماکزیمم کردن مقدار خارج قسمت زیر است،

$$\max_{\alpha} \frac{\alpha^T B \alpha}{\alpha^T W \alpha} \quad (۲۱.۱)$$

یا به طور معادل

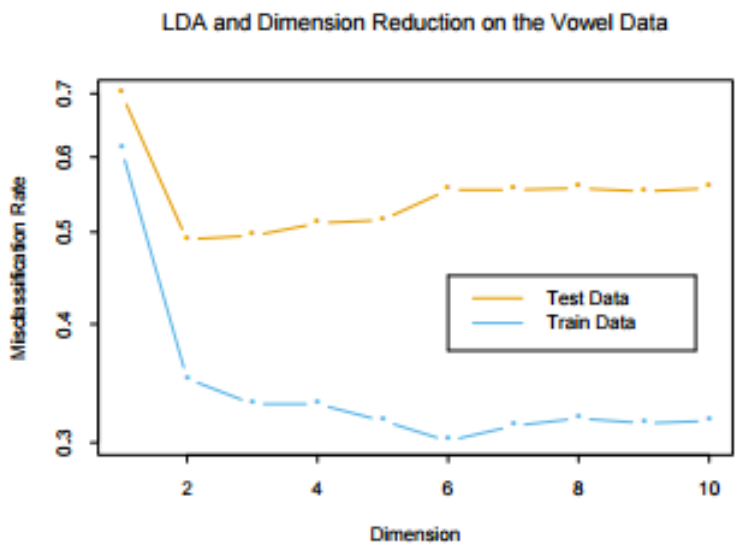
$$\max_{\alpha} \alpha^T B \alpha \qquad \alpha^T W \alpha = 1 \quad (۲۲.۱)$$

با یک تبدیل $B^T = \alpha^T W^{\frac{1}{2}}$ می توان دریافت که α در رابطه (۱۶.۱) صدق کند بیشترین مقدار ویژه $W^{-1}B$ است.

۶,۱ ارتباط LDA و رگرسیون لوژستیک

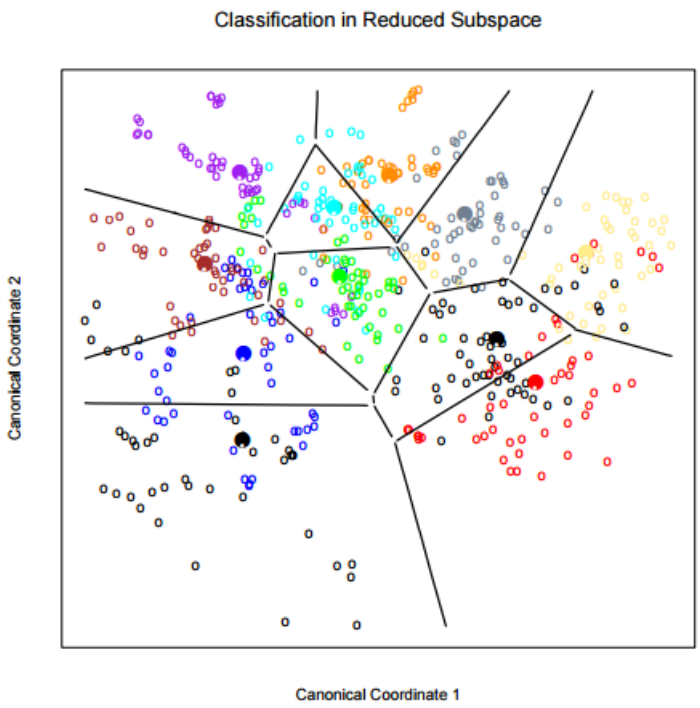
مدل رگرسیون لوجستیک به صورت زیر برای طبقه بندی مشاهدات به K کلاس یک روش طبقه بندی است که ارتباط نزدیکی با روش LDA دارد.

$$\log \frac{Pr(G = k|X = x)}{Pr(G = K|X = x)} = \beta_{k0} + \beta_k^T x, \quad k = 1, \cdots, K-1. \quad (۲۳.۱)$$



شکل ۴.۱

نرخ های خطای اموزشی و ازمون داده های مصوته، مانند تابعی از بعد فضای فرعی تفکیک کننده، در این مورد بهترین نرخ خطا برای دو بعد است.شکل ۴.۱۱ مرز تصمیم در این فضا را نشان می دهد.



شکل ۵.۱

مرزهای تصمیم برای داده های اموزشی مصوت را نشان می دهد. در فضای فرعی دو بعدی با اولین دو متغیر استاندارد اندازه گرفته شد.

در این روش ضرایب β_k و $\beta - k_0$ براساس روش درستنمایی ماکزیمم و در نظر گرفتن توزیع جند جمله ای برای G به شرط X با احتمالت زیر براورد می شوند.

$$Pr(G = k|X = x) = \frac{\exp \beta_{k0} + \beta_k^T x}{1 + \sum_{l=1}^{K-1} \exp\{\beta_{l0} + \beta_l^T x\}}. \tag{۲۴.۱}$$

همانطور که در بخش اول این فصل گفتیم در روش LDA داریم:

$$\log \frac{Pr(G = k|X = x)}{Pr(G = K|X = x)} = \log \frac{\pi_k}{\pi_K} - \frac{1}{2}(\mu_k - \mu_K)^T \Sigma^{-1}(\mu_k - \mu_K) + x^T \Sigma^{-1}(\mu_k - \mu_K) = \alpha_{k0} + \alpha_K^T x. \tag{۲۵.۱}$$

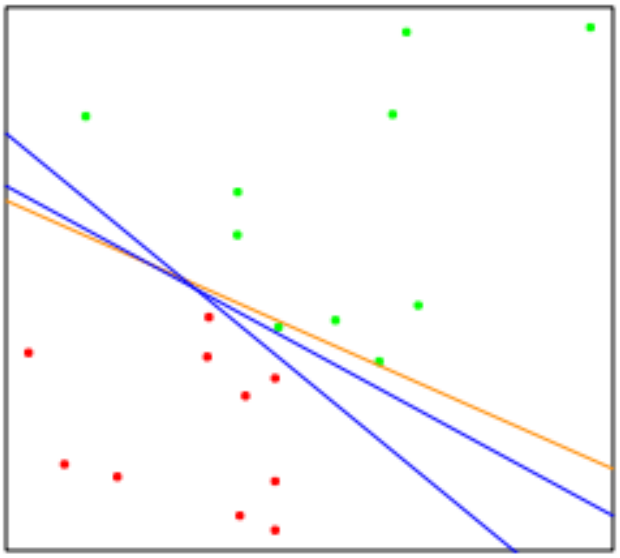
بنابراین به نظر می رسد که روش LDA و رگرسیون لوجستیک یکی هستند، با این حال تفاوت نهفته در روش براورد ضرایب است.یک تفاوت عمده این دوروش این است که رگرسیون لوجستیک تنها از توزیع نرمال G به شرط X استفاده می کند در حالی که روش LDA توزیع توام G و X را به صورت زیر در نظر می گیرد.

$$Pr(X,\ G = k) = Pr(X) \cdot Pr(G = k|X) \tag{۲۶.۱}$$

که در ان برای محاسبه احتمال حاشیهای $Pr(X)$ فرض می‌شود $X|G = k \ \sim N(\mu, \Sigma)$ تفاوت دوم این است که در روش LDA احتمالات پیشین $Pr(G = k) = \pi_k$ نیز در نظر گرفته می شوند. بنابراین تابع چگالی حاشیه‌ای X برابر است با

$$f_x(X) = \sum_{k=1}^{K} \pi_k \cdot \Phi(x;\ \mu_k,\ \Sigma), \tag{۲۷.۱}$$

که در ان Φ تابع چگالی نرمال با میانگین μ_k و ماتریس واریانس کواریانس Σ است.
در نظر گرفتن توزیع حاشیه‌ای X در روش LDA باعث می شود مشاهدات دور از مرز تصمیم نقش برابری در براورد ماتریس کواریانس مشترک با مشاهدات نزدیک به مرز تصمیم داشته باشند.بنابراین روش LDA نسبت به مشاهدات دور است و استوار نیست.با این حال و با تمام این تفاوت ها، اغلب، طبقه بندی حاصل از هر روش بسیار شبیه به هم هستند.



شکل ۶.۱

دو کلاس که توسط ابر صفحه ها مجزا شده اند. خط نارنجی حداقل مربعات را که کدام یکی از داده های اموزشی نادرست طبقه بندی شده اند را نشان می دهد. هم چنین دو ابر صفحه مجزا کننده ابی بر پایه الگوریتم اموزشی پرسپترون با مبدا های تصادفی متفاوت ساخته شده اند.

فصل ۲

معرفی بسته های نرم افزاری موجود در نرم افزار R

۱,۲ مقدمه

از مهم‌ترین و معروف‌ترین روش‌های یادگیری خطی که برای حل مسائل دسته‌بندی استفاده می‌شوند، روش‌های تحلیل مشخصه هستند. از بین بسته‌های متعدد موجود برای نرم‌افزار R، بسته‌ی معروف $MASS$ که توسط دانشگاه آکسفورد تهیه شده و آخرین نسخه‌ی آن در ایپریل ۲۰۱۶ عرضه شده است، شامل دستورهایی برای برازش مدل‌های تحلیل مشخصه‌ی خطی و درجه‌ی دوم است که در ادامه‌ی متن به ترتیب آن‌ها را شرح خواهیم داد. لازم به ذکر است که بنا به تشابه اغلب گزینه‌های این دو دستور، با ذکر وجود گزینه تنها یک بار شرح آن داده خواهد شد.

۲,۲ تحلیل مشخصه‌ی خطی (LDA)

تحلیل مشخصه‌ی خطی توسط بسته‌ی مذکور از طریق دستور $lda()$ انجام می‌شود. با توجه به شکل داده، این دستور به دو طریق قابل اجراست:

۱. در صورتی که مشخص (یا مطلوب) باشد که متغیر پاسخ (کلاس) با کدامین متغیرهای مشخصه در ارتباط است، این متغیرها از طریق گزینه‌ی $formula = groups \sim x_1 + x_2 + \cdots$ مشخص می‌شوند، به طریقی که $groups$ نشان‌دهنده‌ی متغیر پاسخ و از جنس فاکتور $(factor)$ است و x_iها متغیرهای مشخصه هستند. در این صورت گزینه‌ی $data$ داده‌ی حاوی این متغیرها را معین می‌کند. در صورتی که تمامی متغیرهای درون $data$ (به جز پاسخ) به عنوان متغیر مشخصه در نظر گرفته شده باشد، می‌توان از نوشتار $formula = groups \sim .$ نیز استفاده کرد. هم‌چنین در صورت استفاده از گزینه‌ی $formula$

می‌توان با ذخیره‌ی مدل و استفاده از دستور $update$ رابطه و متغیرهای مورد استفاده را بروز رسانی کرد.

۲. در صورتی که ارتباط بین متغیرهای مشخصه با متغیر کلاس از پیش تعیین شده نباشد، داده که در قالب $data\ frame$ یا ماتریس $(matrix/Matrix)$ است توسط گزینه‌ی x مشخص می‌شود. در این حالت گزینه‌ی $grouping$ برداری از جنس فاکتور است که مشاهدات پاسخ را مشخص می‌کند.

در واقع استفاده از دستور $lda(x = discriminants,\ grouping = result)$ از حالت دوم که به آن $lda.default$ گفته می‌شود، معادل است با استفاده از دستور $lda(result \sim discriminants)$. نکته‌ی لازم به ذکر دیگر آن است که روش دوم نسبت به روش اول—که به آن $lda.formula$ گفته می‌شود—سریع‌تر است.

گزینه‌ی دیگر دستور lda، $prior$ است که بردار احتمالات پیشین عضویت در کلاس‌ها را مشخص می‌کند. به طور مثال در صورتی که احتمالات پیشین برای یک مسأله‌ی سه‌کلاسه برابر با ۰/۱, ۰/۳, ۰/۶ باشد، بایستی از نوشتار $prior = c(۰/۱, ۰/۳, ۰/۶)$ استفاده شود. در صورتی که احتمالات پیشین تعیین نشود این مقادیر با استفاده از مشاهدات کل مجموعه داده برآورد می‌شوند.

در واقع یک متغیر وقتی به عنوان یک متغیر مشخصه کارساز و قابل استفاده است که حاوی اطلاعات باشد، بر این اساس به طور مثال متغیرهای ثابت ناکارامد هستند. چنین متغیرهای کم‌اطلاع یا فاقد اطلاعی موجب معکوس ناپذیر شدن ماتریس بین دسته‌ای می‌شوند که فرآیند حل مسأله را با مشکل روبرو می‌کند. از این رو بایستی حتی‌الامکان از چنین رویدادی جلوگیری کرد. در این راستا گزینه‌ی tol حدی را تعیین می‌کند و در صورتی که انحراف معیار یک بردار از این مقدار کم‌تر باشد آن متغیر و ترکیب‌های آن را در فرآیند یادگیری نمی‌پذیرد. مقدار پیش‌فرض این گزینه برابر با $۱۰^{-۴}$ است.

در شرایطی که نیاز باشد تنها بخشی از داده‌ها در فرایند آموزش به کار گرفته شوند، نوشتار $subset$ برای تعیین این زیرمجموعه از داده‌ها قابل استفاده است. این زیرمجموعه ممکن است از قبل تعیین شده باشد که در این صورت تنها نام آن مقابل $subset =$ قرار می‌گیرد یا به هر طریق دیگری مانند نمونه‌گیری یا برشی از داده‌ها تعیین شود.

برای تعیین نحوه‌ی برآورد میانگین و واریانس طی فرآیند یادگیری از گزینه‌ی $method$ استفاده می‌کنیم. مقادیر ممکن این گزینه برابرند با: $moment$ (برآورد گشتاوری/میانگین و واریانس نمونه‌ای)، mle (برآوردهای حداکثر درستنمایی)، t (برآورد استوار بر اساس توزیع $t-student$) و mve (برآورد مکان و مقیاس چندمتغیره که در واقع به تعبیری میانگین و واریانس قسمت مناسبی از داده‌ها را محاسبه می‌کند). در صورتی که از برآوردهای استوار استفاده شود، گزینه‌ی nu درجه‌ی آزادی توزیع t مدنظر را تعیین می‌کند.

در نهایت گزینه‌ی $na.action$ که نحوه‌ی برخورد با داده‌های گمشده را تعیین می‌کند و گزینه‌ی CV که انجام شدن یا نشدن اعتبارسنجی متقابل را مشخص می‌کند، نیز از گزینه‌های ممکن دستور $lda()$ هستند.

در صورتی که $CV = T$ یعنی اعتبارسنجی متقابل انجام شود، خروجی دستور شامل کلاس‌های برآورد

شده برای مشاهدات و برآورد توزیع پسین هر کلاس برای هر مشاهده است که توسط روش بیشینه‌ی پسین (MAP) برآورد می‌شوند.
اما در صورتی که اعتبارسنجی متقابل انجام نگیرد، خروجی دستور lda شامل این موارد می‌شود: توزیع پیشین ($prior$)، میانگین‌ها ($means$)، ماتریس نرمال‌ساز ($scaling$)، مقدار svd که مشخص کننده‌ی نسبت انحراف معیار بین گروهی به درون گروهی را مشخص می‌کند و توان دوم آن مقدار آماره‌ی F مرکزی است، تعداد مشاهدات (N) و در نهایت تابع فراخوانی شده ($call$).

۳,۲ تحلیل مشخصه‌ی درجه‌ی دوم (QDA)

در مورد دستور $qda()$ برای انجام تحلیل مشخصه‌ی درجه‌ی دوم، تمامی گزینه‌ها و خروجی‌ها تشریحی مشابه با دستور $lda()$ دارند و مازاد بر آن در خروجی qda، گزینه‌ی $ldet$ برداری از دترمینان ماتریس اغتشاش است و lev سطوح پاسخ را نشان می‌دهد.

فرم کلی دستورهای lda و qda:

$$lda(formula, data, method, na.action, prior = proportions, CV = F, nu, ...) \qquad (۱.۲)$$

$$qda(formula, data, method, na.action, prior = proportions, CV = F, nu, ...) \qquad (۲.۲)$$

۴,۲ پیش‌بینی

برای پیش‌بینی از مدل‌های تحلیل مشخصه می‌توان مانند اغلب مدل‌هایی که در محیط R ساخته می‌شوند به این صورت عمل کرد که مدل در یک مؤلفه مانند $model$ ذخیره شود و سپس با استفاده از دستور $predict()$ و معرفی داده‌ی جدید، پیش‌بینی مقادیر نظیر این داده‌ی جدید مانند کلاس برآورد شده و یا توزیع پسین انجام گیرد.