بسمه تعالی

بازشناسی آماری الگو

دانشگاه صنعتی خواجه نصیرالدین طوسی — تمرین سری هفتم — تاریخ تحویل: ۱۴۰۴/۹/۸ — آزمایشگاه بینایی ماشین و پردازش تصاویر پزشکی

۱- مساله امتحان پایان ترم ۱۳۹۵

تابع هزینه پیشنهادی فیشر برای یک دسته‌بند خطی بصورت زیر تعریف گردید:

$$J_F = \frac{|\mathbf{w}^t(\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_2)|^2}{\mathbf{w}^t \hat{\Sigma}_w \mathbf{w}}$$

در رابطه فوق $\mathbf{w}$ بردار وزنهای کلاسیفایر خطی فیشر، $\hat{\mathbf{m}}_1$ و $\hat{\mathbf{m}}_2$ مراکز ثقل دو کلاس و $\hat{\Sigma}_w$ تخمین ماتریس پراکندگی داخل کلاسی هستند.

الف- با بهینه‌سازی تابع هزینه فیشر نشان دهید که بردار وزن $\mathbf{w}$ بصورت زیر بدست می‌آید:

$$\mathbf{w} \propto \hat{\Sigma}_w^{-1}(\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_2)$$

یعنی $\mathbf{w}$ با تقریب یک ضریب با حاصلضرب $\hat{\Sigma}_w^{-1}$ در بردار واصل مراکز ثقل دو کلاس $(\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_2)$ برابر است.

ب- می‌دانیم تخمین ماتریس پراکندگی داخل کلاسی مجموع وزن‌دار تخمین‌های ماتریس پراکندگی هر یک از کلاسها بصورت زیر است:

$$\hat{\Sigma}_w = \frac{1}{N}(N_1\hat{\Sigma}_1 + N_2\hat{\Sigma}_2)$$

که در آن N تعداد کل نمونه‌ها، N_1 و N_2 تعداد نمونه‌های هر یک از دو کلاس (N=$N_1 + N_2$) و $\hat{\Sigma}_1$ و $\hat{\Sigma}_2$ تخمین ماتریسهای پراکندگی هر یک از کلاسها هستند. فرض کنید احتمال کلاسها برابر بوده ($N_1 = N_2$) و تخمینهای زیر با استفاده از نمونه‌های موجود از هر کلاس بدست آمده است:

$$\hat{\mathbf{m}}_1 = \begin{bmatrix}0\\0\end{bmatrix}, \hat{\mathbf{m}}_2 = \begin{bmatrix}3\\1\end{bmatrix}, \hat{\Sigma}_1 = \hat{\Sigma}_2 = \begin{bmatrix}3 & 0\\0 & 1\end{bmatrix}$$

بردار $\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_2$ را در فضای دوبعدی نشان دهید و با فرض اینکه کلاسهای ω_1 و ω_2 دارای توزیع گوسی باشند، با رسم یک منحنی تقریبی مکان نقاط همچگال را در هر یک از دو کلاس نشان دهید.

ج- با کمک شکل رسم شده در بالا توضیح دهید چرا بردار واصل مراکز ثقل دو کلاس (یعنی بردار $\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_2$) انتخاب مناسبی به عنوان بردار وزن یک کلاسیفایر خطی برای دسته‌بندی این دو کلاس نیست.

د- از روابط زیر کمک بگیرید و توضیح دهید چرا ضرب $\hat{\Sigma}_w^{-1}$ در بردار واصل بین مراکز ثقل دو کلاس، یعنی بردار $\hat{\Sigma}_w^{-1}(\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_2)$ می تواند انتخاب بهتری نسبت به بردار $\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_2$ به عنوان بردار وزن یک کلاسیفایر خطی باشد. (راهنمایی: باید نشان دهید که ضرب $\hat{\Sigma}_w^{-1}$ در بردار $(\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_2)$ راستای آن را به سمت راستای کمترین پراکندگی معرفی شده توسط ماتریس پراکندگی داخل کلاسی تغییر می‌دهد.)

$$\hat{\Sigma}_w = \lambda_1\boldsymbol{\varphi}_1\boldsymbol{\varphi}_1^t + \lambda_2\boldsymbol{\varphi}_2\boldsymbol{\varphi}_2^t$$
$$\hat{\Sigma}_w^{-1} = (1/\lambda_1)\boldsymbol{\varphi}_1\boldsymbol{\varphi}_1^t + (1/\lambda_2)\boldsymbol{\varphi}_2\boldsymbol{\varphi}_2^t$$

در روابط فوق λ_1 و λ_2 مقادیر ویژه و $\boldsymbol{\varphi}_1$ و $\boldsymbol{\varphi}_2$ بردارهای ویژه ماتریس $\hat{\Sigma}_w$ هستند.

۲- (اختیاری) دیدیم که روش حداقل مربعات (Minimum Squared Error (MSE)) در حالت خاص، تابع تمایز خطی فیشر (Fisher's linear discriminant) را نتیجه می‌دهد. با استفاده از تحلیل انجام شده در کلاس، نشان دهید مقدار h در کلاسیفایر فیشر از رابطه زیر بدست می‌آید:

$$h = \left[1 + \frac{N_1N_2}{N}(\mathbf{m}_1 - \mathbf{m}_2)^t S_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)\right]^{-1}$$

۳- (اختیاری) برای بدست آوردن بردار وزنهای یک فوق صفحه جداکننده دو کلاس و با در اختیار داشتن یک مجموعه از نمونه‌های آموزشی، استفاده از روش گرادیان نزولی متداول است. در این روش با تعریف یک تابع هزینه مناسب $J(\mathbf{v})$ و شروع از یک بردار وزن اولیه دلخواه $\mathbf{v}(0)$ می‌توان با استفاده از الگوریتم مرحله‌ای زیر و پس از تعداد محدودی تکرار (البته به شرط تفکیک‌پذیری خطی کلاس‌ها) بردار وزن مطلوب را بدست آورد:

$$\mathbf{v}(k+1) = \mathbf{v}(k) - a_k \nabla J(\mathbf{v}(k))$$

که در آن a_k نرخ آموزش نامیده می‌شود. انتخاب یک مقدار کوچک برای a_k موجب کندی همگرایی الگوریتم و انتخاب مقدار بزرگ موجب نوسانی شدن پاسخ و یا عبور آن از نقطه بهینه می‌شود. با استفاده از بسط تیلور تابع هزینه حول $\mathbf{v}(k)$ و تقریب مرتبه دوم نشان دهید مقدار بهینه a_k در هر مرحله بصورت زیر قابل محاسبه است:

$$a(k) = \frac{\|\nabla J\|^2}{\nabla J^t H \nabla J}$$

که در آن H ماتریس هسیان (Hessian Matrix) می‌باشد.

۴- (اختیاری) روابط مورد نیاز برای پیاده‌سازی برخط (online) الگوریتم پرسپترون در حالت جداسازی خطی چندکلاسه را بر اساس تعمیم بردار وزن و نمونه‌های آموزشی مطابق روش Kessler بنویسید.

۵- (اختیاری) مساله امتحان میان ترم بازشناسی آماری الگو سال ۱۳۸۷

در این مساله می خواهیم کلاسیفایر فیشر را به طریق دیگری بدست آوریم. فرض کنیم $p(\mathbf{x}|\omega_i)$ معرف چگالی احتمال دلخواه کلاس ω_i با متوسط $\mathbf{m}_i$ و ماتریس کواریانس Σ_i باشد (i=1,2). حال اگر $y=\mathbf{v}^t\mathbf{x}$ یک تبدیل خطی از نوع افکنشی (projection) باشد، بطوریکه چگالی احتمال شرطی $p(y|\omega_i)$ دارای متوسط m_i و واریانس σ_i^2 گردد، آنگاه:

الف- نشان دهید ماکزیمم کردن تابع هزینه زیر:

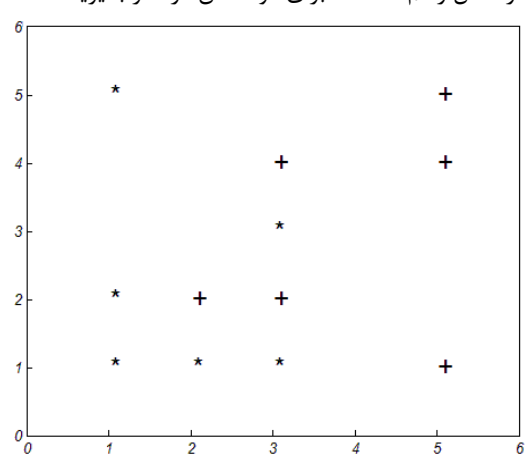
$$J(\mathbf{v}) = \frac{(m_1 - m_2)^2}{\sigma_1^2 + \sigma_2^2}$$

بردار مطلوب $\mathbf{v}$ را بصورت زیر نتیجه خواهد داد که همان کلاسیفایر فیشر می‌باشد:

$$\mathbf{v} = (\Sigma_1 + \Sigma_2)^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$$

ب- فرآیند فوق را با ترسیم یک شکل ساده در فضای دوبعدی تشریح و تفسیر کنید (توزیع‌های $p(\mathbf{x}|\omega_i)$، بردار $\mathbf{v}$، تبدیل افکنشی و کمیت‌های مقدار متوسط m_i و واریانس σ_i^2 را درشکل خود نشان دهید).

۶. (اختیاری) داده‌های دو بعدی زیر را که در شکل رسم شده‌اند، برای دو کلاس در نظر بگیرید:



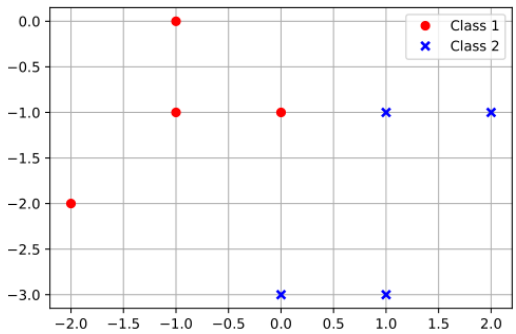
$\omega_1 = [(1,1), (1,2), (1,5), (2,1), (3,1), (3,3)]$
$\omega_2 = [(2,2), (3,2), (3,4), (5,1), (5,4), (5,5)]$

الف) معادله بردار بهینه تبدیل افکنشی $(\mathbf{V})$ بر اساس روش حداقل مربعات خطا (کلاسیفایر فیشر) را محاسبه کنید.

ب) تصویر نقاط مربوط به کلاس‌های ω_1 و ω_1 روی این بردار را نشان دهید و سپس با فرض یک توزیع مناسب برای کلاس‌ها در فضای یک بعدی جدید، آنها را با روش بیز دسته‌بندی کنید.

۷- (اختیاری) سوال امتحانی ۱۳۹۷

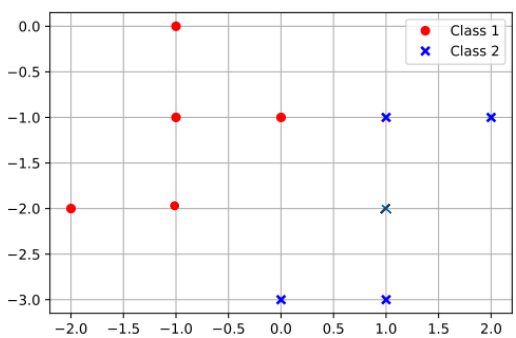
برای مساله دوکلاسه شکل زیر که در آن هر کلاس شامل ۴ نمونه می‌باشد یک دسته‌بند LDA بدست آورید و با انتخاب آستانه مناسب قاعده تصمیم‌گیری را بیان نمایید.



۸- سوال امتحانی ۱۳۹۸

الف- برای مساله دوکلاسه شکل زیر که در آن هر کلاس شامل ۵ نمونه می‌باشد راستای تقریبی دسته‌بند LDA را بدست آورید.

راهنمایی: برای صرفه‌جویی در وقت راستای دسته‌بند را بصورت تقریبی حدس بزنید (نیازی به محاسبه دقیق راستای دسته‌بند نیست)



ب- پاره‌خط واصل مراکز ثقل دو کلاس را در شکل بالا رسم و محل تقریبی تلاقی دسته‌بند LDA با آن را مشخص نمایید.

۹- (اختیاری) سوال امتحانی ۱۳۹۹

جدول زیر فهرستی از نقاط آموزشی در فضای دوبعدی را نشان میدهد. فرض کنید ما الگوریتم پرسپترون را با یک بعد اضافی روی این نقاط اجرا کنیم. تعداد دفعاتی که هرنقطه به دلیل دسته بندی اشتباه در الگوریتم گرادیان نزولی شرکت کرده است را در آخرین ستون سمت راست ثبت کرده‌ایم.

x_1	x_2	y	times misclassified
−3	2	+1	0
−1	1	+1	0
−1	−1	−1	2
2	2	−1	1
1	−1	−1	0

الف- فرض کنید نرخ آموزش برابر واحد و بردار وزن اولیه $\mathbf{v}_0 = (-3 \quad 2 \quad 1)^t$ باشد، که در آن عنصر آخر همان بایاس است. معادله مرز جداکننده دو کلاس پس از اجرای الگوریتم را برحسب x_1 و x_2 بنویسید.

ب- در برخی موارد، حذف حتی یک نقطه آموزشی می تواند موجب تغییر مرز بدست آمده توسط الگوریتم پرسپترون بشود. حذف کدامیک از نقاط مجموعه داده آموزشی جدول بالا می تواند موجب تغییر مرز بشود؟ پاسخ خود را توضیح دهید.

ج- اگر نقطه جدید آموزشی (2,-2) با برچسب $+1$ را به مجموعه داده های آموزشی اضافه کنیم نتیجه الگوریتم چه تغییری خواهد داشت.

۱۰- الگوریتم پرسپترون را در نظر بگیرید. مشاهدات زیر با وزن‌های اولیه $w = [1, 1]$ و $b = 0$ به شبکه داده می‌شوند. با انجام محاسبات نشان دهید که خط جداسازی که شبکه پس از یک بار دیدن همه نمونه‌ها به ترتیب داده شده در جدول پیشنهاد می‌کند چیست؟ (نرخ آموزش را ۱ در نظر بگیرید)

راهنمایی: تنها در صورتی که $y_i(w^t X_i + b) \leq 0$ به‌روز رسانی وزن‌ها به شکل $w^j = w^{j-1} + \mu y_i X$ و به روزرسانی بایاس به شکل $b^j = b^{j-1} + \mu y_i$ اتفاق می‌افتد که در آن $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.)

Instance	1	2	3	4
Label y	+1	-1	+1	-1
Data (x_1, x_2)	(10,10)	(0,0)	(8,4)	(3,3)

۱۱- (میان ترم ۱۴۰۲) الگوریتم پرسپترون را در نظر بگیرید. مشاهدات زیر با وزن‌های اولیه $w = [\frac{3}{2}, 1]$ و بایاس اولیه $b = \frac{-1}{2}$ به شبکه داده می‌شوند. با انجام محاسبات نشان دهید که خط جداسازی که شبکه پس از چند بار دیدن همه نمونه‌ها، به ترتیب داده شده در جدول، پیشنهاد میکند چیست؟ (نرخ آموزش را 0.75 در نظر بگیرید). در انتهای هر مرحله (دیدن همه نمونه‌ها)، خط جداسازی را به همراه نمونه‌ها رسم کنید.

Instance	1	2	3	4
Label	-1	-1	1	1
Data (x_1, x_2)	(0, 0)	(1, 0)	(0, 2)	(2, 0)

موفق باشید