# تمرین سوم درس یادگیری ماشین
## دکتر باباعلی

سیدعلیرضا مولوی

# فهرست مطالب

## ۱ LDA برای داده های چند کلاسه

### ۱.۱ مدل محاسباتی

فرض کنیم که داده های ما شامل $K \geq 2$ کلاس باشند.
معادله $f(x;\omega)$ را به صورت معادله ۱ تعریف می کنیم.

$$f(x;\omega) = \omega^T x \tag{۱}$$

در واقع تابع $f(\cdot;\omega)$ داده را بر روی ابرصفحه[۱] $\omega$ نظیر می کند.
**توجه** اگر هدف کاهش به بعد $d'$ باشد، کافی است که $\omega \in R^{(d,d')}$ در نظر بگیرم.

### ۲.۱ ماتریس پراکندگی درون کلاسی

برای محاسبه ماتریس پراکندگی درون کلاسی ابتدا طبق معادله ۲ میانگین تمام داده ها را محاسبه می کنیم، سپس بر اساس معادله ۳ میامگین هر کلاس را جداگانه بدست می آوریم ، و سپس طبق معادله ۴، که $N_k$ تعداد اعضای کلاس $k$ اند ، ماتریس پراکندگی درون کلاسی را محاسبه می کنیم.

$$\mu = \frac{1}{N} \sum_i x^{(i)} \tag{۲}$$

$$\mu_k = \frac{1}{N_k} \sum_{i;y_i=k} x^{(i)} \tag{۳}$$

$$S_B = \sum_k N_k(\mu - \mu_k)(\mu - \mu_k)^T \tag{۴}$$

### ۳.۱ ماتریس پراکندگی بین کلاسی

برای محاسبه ماتریس پراکندگی بین کلاسی ابتدا ماتریس پراکندگی هر کلاس را طبق معادله ۵ محاسبه میکنیم و سپس طبق معادله ۶ با هم جمع می کنیم.

$$s_k^2 = \sum_{i;y_i=k} (\mu_k - x^{(i)})(\mu_k - x^{(i)})^T \tag{۵}$$

$$S_w = \sum_{i=1}^{K} s_i^2 \tag{۶}$$

### ۴.۱ تابع معیار

تابع معیار $J(\omega)$ به صورت معادله ۱۰ تعریف می کنیم.

$$\mu\prime = \omega^T \mu \tag{۷}$$

$$\mu'_k = \omega^T \mu_k \tag{۸}$$

$$s'^2_k = \sum_{i;y_i=k} (\omega^T \mu_k - \omega^T x^{(i)})(\omega^T \mu_k - \omega^T x^{(i)})^T \tag{۹}$$

$$J(\omega) = \frac{\sum_k N_k(\mu'_k - \mu')(\mu'_k - \mu')^T}{\sum_k s'^2_k} \tag{۱۰}$$

[۱] HyperPlane

اکنون سعی می کنیم که معادلات را ساده تر کنیم. با فاکتورگیری به معادله ۱۲ می رسیم.

$$J(\omega) = \frac{\omega^T(\sum_k N_k(\mu_k - \mu)(\mu_k - \mu)^T)\omega}{\omega^T(\sum_k s_k^2)\omega} \tag{۱۱}$$

$$J(\omega) = \frac{\omega^T S_B \omega}{\omega^T S_w \omega} \tag{۱۲}$$

## ۵.۱ محاسبه $\omega$ بهینه

هدف ما بیشینه کردن تابع معیار $J(\omega)$ است.

$$\omega^* = argmax_\omega J(\omega) \tag{۱۳}$$

اکنون برای بدست آوردن $\omega$ بهینه کافی است، مشتق تابع $J(\omega)$ را محاسبه کرده و مقدار مشتق را برابر $0$ قرار داده و $\omega$ بهینه را استخراج کنیم. معادلات کاملا مشابه معادلات درس داده شده در بخش $\mathrm{LDA}$ است.

$$\frac{\partial J(\omega)}{\partial \omega} = 0 \implies S_B\omega = \lambda S_w \omega \tag{۱۴}$$

$$\implies S_w^\dagger S_B \omega = \lambda \omega \tag{۱۵}$$

$$\lambda = J(\omega) = \frac{\omega^T S_B \omega}{\omega^T S_w \omega} \tag{۱۶}$$

با مشتق گیری به معادله ۱۵ می رسیم که در آن $S_w^\dagger$ شبه معکوس است؛ معادله ۱۵ در واقع معادله بردار ویژه است و با حل آن می توانیم مقدار بهینه $\omega$ را بدست بیاوریم. هدف بیشینه کردن $\lambda = J(\omega)$ است، بنابراین $\omega$ بردار ویژه بزرگترین مقدار ویژه معادله ۱۵ است (اگر $\omega \in R^{d'}$ باشد، آنگاه $\omega$ برابر $d'$ بزرگترین بردار های ویژه است.) .

## ۲ سوال دوم

برای حل از روش چن کلاسه استفاده شده است.
**توجه** ماتریس پراکندگی بین کلاسی در روش چن کلاسه با روش دو کلاسه متفاوت است ولی در نهایت جواب های بدست آمده یکسان است.

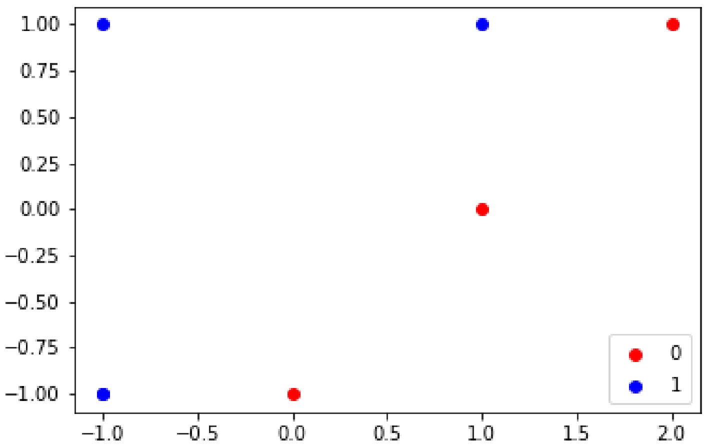
### ۱.۲ داده ها

$$X_0 = \left\{ \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\} \tag{۱۷}$$

$$X_1 = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\} \tag{۱۸}$$

(۱۹)

شکل ۱: داده های آموزشی در فضای ۲ بعدی




### ۲.۲ مدل محاسباتی

با توجه به روش ذکر شده مدل محاسباتی را به صورت معادله ۲۰ تعریف می کنیم. در واقع تابع $f(\cdot;\omega)$ نقطه $x$ را بر روی خط $\omega$ نظیر می کند.

$$f(x;\omega) = \omega^T x \tag{۲۰}$$

### ۳.۲ محاسبه ماتریس پراکندگی بین کلاسی

**توجه** ماتریس پراکندگی بین کلاسی در روش چن کلاسه با روش دو کلاسه متفاوت است ولی در نهایت جواب های بدست آمده یکسان است.

$$\mu = \begin{pmatrix} 0.14285714 \\ 0 \end{pmatrix} \tag{۲۱}$$

$$\mu_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{۲۲}$$

$$\mu_1 = \begin{pmatrix} 1 \\ -0.5 \end{pmatrix} \tag{۲۳}$$

اکنون مقادیر بالا را در معادله ۴ جایگذاری می کنیم.

$$S_B = 3(\mu - \mu_0)(\mu - \mu_0)^T + 4(\mu - \mu_1)(\mu - \mu_1)^T \tag{۲۴}$$

$$= \begin{pmatrix} 3.85714286 & 0 \\ 0 & 0 \end{pmatrix} \tag{۲۵}$$

### ۴.۲ محاسبه ماتریس پراکندگی بین کلاسی

با استفاده از معادلات ۵ و ۶ مقادیر زیر را حساب می کنیم.

$$s_0^2 = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \tag{۲۶}$$

$$s_1^2 = \begin{pmatrix} 3 & 2 \\ 2 & 4 \end{pmatrix} \tag{۲۷}$$

$$S_w = \begin{pmatrix} 5 & 4 \\ 4 & 6 \end{pmatrix} \tag{۲۸}$$

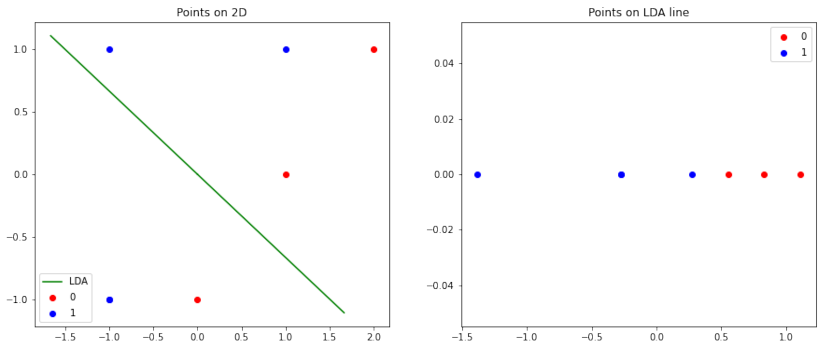
### ۵.۲ محاسبه $\omega$ بهینه

با جایگزاری مقادیر محاسبه شده $S_B$ و $S_w$ در معادلات ۲۵، ۲۸ در معادله ۱۵ و حل معادله مقدار ویژه، مقدار بهینه زیر بدست می آید. در شکل ۲ داده ها و خط LDA رسم شده است.

$$\omega = \begin{pmatrix} 0.83205029 \\ -0.5547002 \end{pmatrix} \tag{۲۹}$$

$$J(\omega) = 1.6530612244897962 \tag{۳۰}$$

شکل ۲: **شکل چپ:** داده ها در فضای دو بعدی و خط قرمز رنگ خط LDA است. **شکل راست:** داده ها نگاشت شده بر روی خط LDA

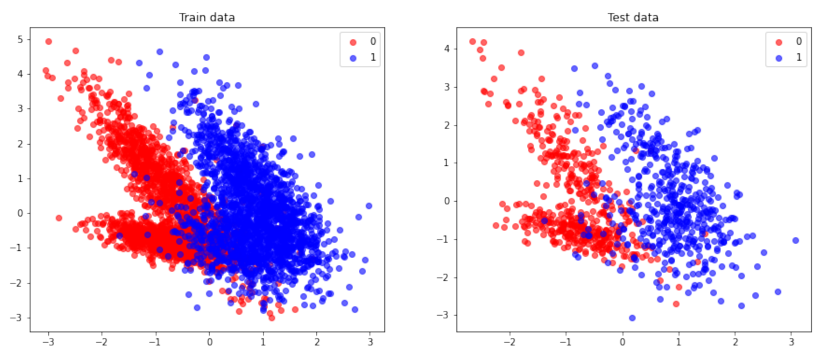



# ۳ سوال سوم

## ۱.۳ داده ها

داده ها دارای دو ویژگی اند و هر داده به کلاس $0$ و یا $1$ تخصیص داده شده است. بنابراین مسئله دسته بندی دوتایی[۲] است. داده ها به دو دسته، داده های آموزشی و داده های تست تقسیم شده است. تعداد داده های آموزشی ۴۰۰۰ و تعداد داده های تست ۲۰۰۰ است. در شکل ۳ داده ها را مشاهده می کنید.

شکل ۳: **شکل راست:** داده های تست. **شکل چپ:** داده های آموزشی




## ۲.۳ نظیر کردن داده ها به فضای ۱ بعدی

با استفاده از الگوریتم LDA ما داده ها به فضای یک بعدی کاهش داده ایم (شکل ۴). مقدار بهینه

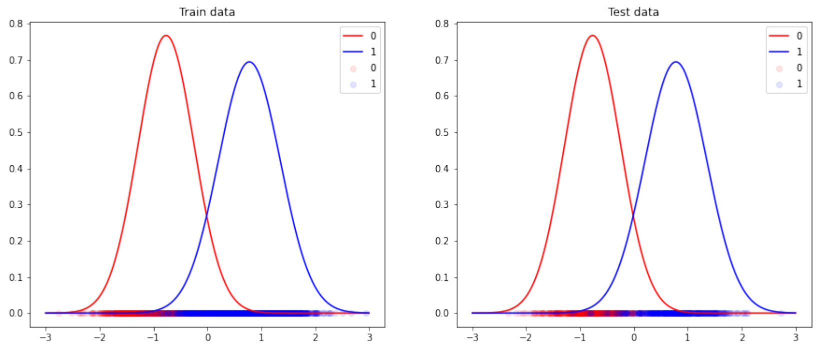
$$\omega = \begin{pmatrix}0.9708\\0.2310\end{pmatrix}$$

است. مقدار تابع معیار $J(\omega)$ برای داده ها به صورت زیر است:

$$J_{train}(\omega) = 1.9894 \quad J_{test}(\omega) = 2.1813$$

[۲] Binary Classifivation

شکل ۴: **شکل راست:** داده های تست. **شکل چپ:** داده های آموزشی




## ۳.۳ مسئله دسته بندی [۳]

فرض می کنیم که داده ها از یک توزیع احتمالاتی تولید می شوند $(x, y) \sim P$. مقدار بهینه $y$ را به صورت زیر تعریف می کنیم:

$$y^* = argmax_y P(y|x;\theta) \tag{۳۱}$$
$$\propto argmax_y P(x|y;\theta) P(y;\theta) \tag{۳۲}$$
$$\propto argmax_y logP(x|y;\theta) + logP(y;\theta) \tag{۳۳}$$

فرض می گیریم که هر داده از یک توزیع گاووسی چند متغیره [۴] تولید می شود و مقدار پیشین [۵] $P(y;\theta)$ را میانگین $y$ در نظر می گیرم.

$$P(x|y;\theta) \sim N(x;\mu_y, \Sigma_y) \tag{۳۴}$$
$$P(y;\theta) = \frac{N_y}{N} \tag{۳۵}$$
$$\mu_y = \frac{1}{N_y} \sum_{i, y_i = y} x^{(i)} \tag{۳۶}$$
$$\Sigma_y(i, j) = Cov(x_i, x_j) \tag{۳۷}$$

در شکل ۴ نمودار های گاووسی بر روی داده رسم شده اند.

## ۴.۳ بررسی نتایج

مقدار دقت پس از آموزش:

$$Train\ Accuracy = 0.922 \quad Test\ Accuracy = 0.932$$

با توجه به شکل ۴ بخشی از داده ها توسط هر دو توزیع گاووسی پوشانده شده اند و مدل معمولا به اشتباه این داده ها را دسته بندی می کند.در شکل ۵ داده هایی که اشتباه دسته بندی شده اند رو مشاهده می کنید.

---

[۳] Classification Problem
[۴] Multivariate normal distribution
[۵] Prior

شکل ۵: **شکل راست:** داده های تست. **شکل چپ:** داده های آموزشی
داده های که اشتباه دسته بندی شده اند به رنگ مشکی اند.