



به نام خدا

1



# پردازش سیگنال‌های مغزی به منظور طبقه‌بندی

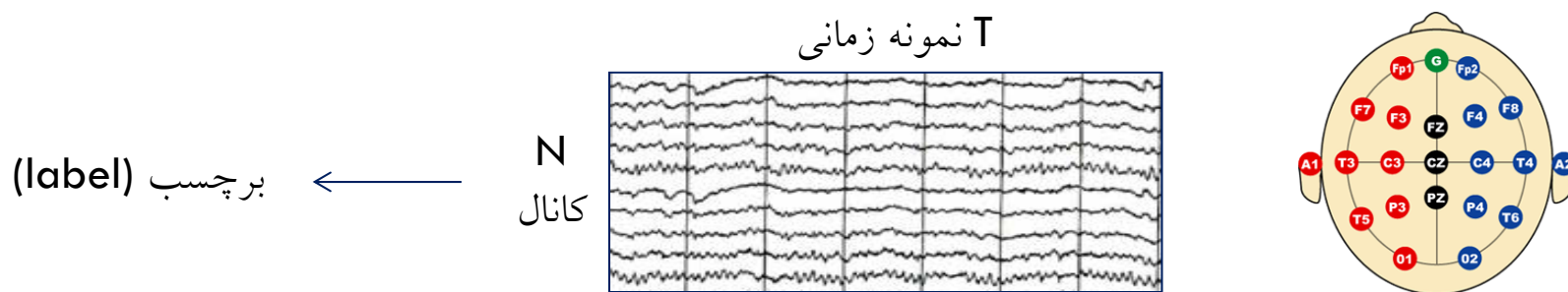
استخراج و انتخاب ویژگی‌های مؤثر، طبقه‌بندی، ارزیابی

# داده‌ها در مسئله طبقه‌بندی



2

□ داده در یک ثبت یا یک آزمایش (trial)



□ داده‌ها:

□ داده‌های آموزشی (Training)

■  $M$  آزمایش

■ برای هر آزمایش،  $N$  کانال و  $T$  نمونه زمانی

■ برای هر آزمایش یک برچسب

← هدف: تعیین برچسب برای داده‌های آزمون

□ داده‌های آزمون (Test)

■  $L$  آزمایش

■ برای هر آزمایش،  $N$  کانال و  $T$  نمونه زمانی

■ بدون برچسب

# داده‌ها در مسئله طبقه‌بندی



3

□ ممکن است داده‌های آموزشی و آزمون به صراحت تعریف و جدا نشده باشند یا اینکه برچسب‌ها دقیقاً مشخص نشده باشند:

□ ← پردازش داده‌ها به منظور تعریف دقیق داده‌های آموزشی و آزمون

□ ممکن است داده‌های برچسب‌دار نداشته باشیم:

□ ← مسئله خوشه‌بندی (clustering)



## داده‌ها در مسئله طبقه‌بندی: یک نمونه

4

- دسته چهارم مسابقات BCI2003
- داده‌ها از یک شخص عادی در حالت بدون فیدبک ثبت شده است.
- شخص مورد آزمایش بر روی یک صندلی عادی قرار گرفته است. دست‌ها بر روی میز تکیه داده شده و انگشت‌های او در موقعیت استاندارد برای تایپ بر روی کیبورد کامپیوتر قرار گرفته‌اند.
- وظیفه شخص فشردن ۴ کلید مشخص شده بر روی صفحه کلید با انگشتان کوچک یا اشاره دست راست یا چپ به خواست خود و با زمان بندی دلخواه است.
- داده‌ها به صورت ۴۱۶ دوره به طول ۵۰۰ میلی ثانیه هستند که ۱۳۰ میلی ثانیه قبل از فشردن کلید تمام می شوند.
- فرکانس نمونه برداری ۱۰۰۰ هرتز بوده و داده‌ها از فیلتر میان گذر ۰/۰۵ تا ۲۰۰ هرتز گذرانده شده‌اند.
- ۲۸ کانال اندازه‌گیری EEG در موقعیت استاندارد جهانی سیستم ۲۰-۱۰



# داده‌ها در مسئله طبقه‌بندی: یک نمونه

□ ۳۱۶ آزمایش برچسب‌دار:

□ ۰: حرکت دست چپ

□ ۱: حرکت دست راست

□ ۱۰۰ آزمایش دیگر بدون برچسب:

□ داده‌های آزمون

□ پیش‌پردازش:

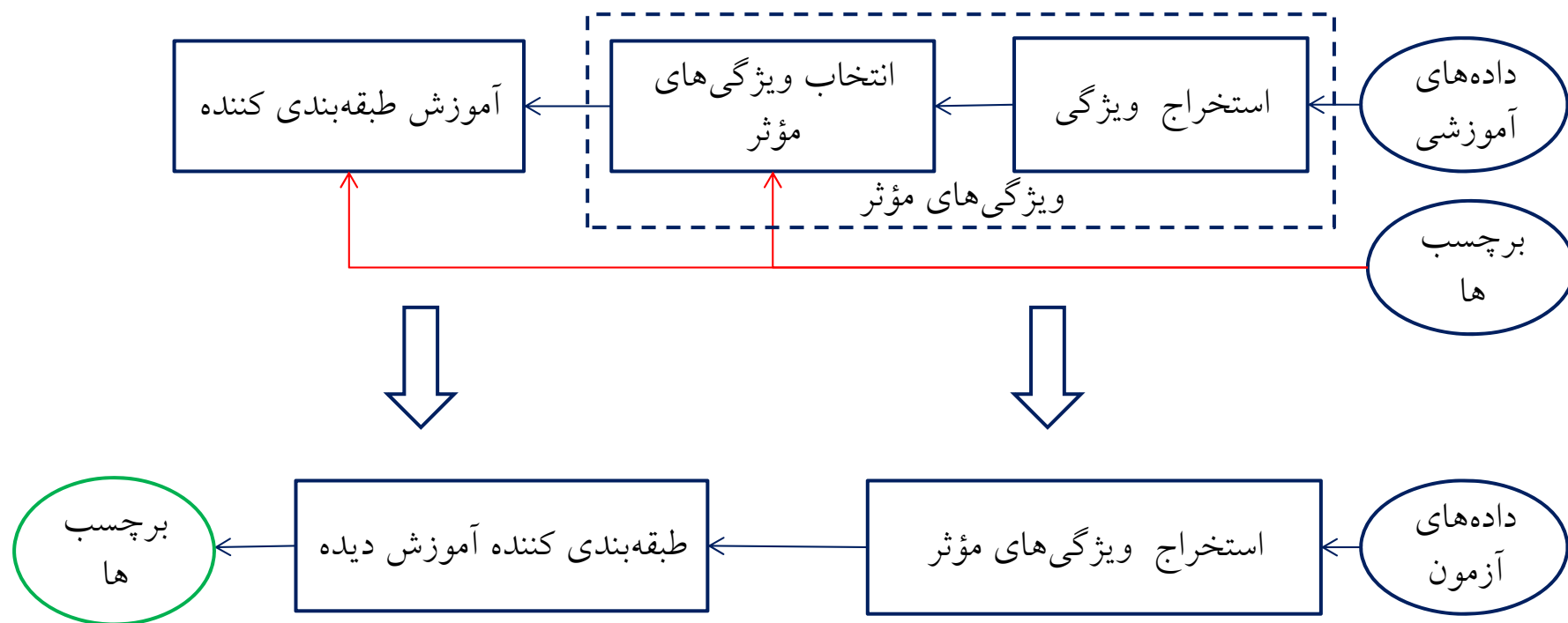
□ داده‌ها از یک فیلتر پایین گذر ۴۵ هرتز گذرانده شده‌اند.

□ برای افزایش سرعت پردازش، نرخ نمونه برداری به ۱۰۰ هرتز کاهش داده شده است.

# بلوک دیاگرام مسئله طبقه‌بندی



6



# ویژگی ها



7

□ ویژگی های آماری

□ میانگین، واریانس، همبستگی میان کانال ها، مومان، کومولان، فرم فاکتور و ...

□ ویژگی های ریخت شناسی (شکلی-زمانی) (مورفولوژیک)

□ زمان قله ها و دره ها، دامنه سیگنال در زمان قله ها و دره ها، شیب صعود یا سقوط و ...

□ ویژگی های وابسته به تبدیل های فرکانسی

□ تبدیل فوریه، تبدیل سینوسی، تبدیل کسینوسی

□ ویژگی های وابسته به انرژی باندهای فرکانسی

□ انرژی نرمالیزه در باندهای  $\delta$ ،  $\theta$ ،  $\alpha$  و  $\beta$

# ویژگی ها



8

□ ویژگی های وابسته به تبدیل های زمان-فرکانس  
□ تبدیل ویولت پیوسته و گسسته و ...

□ ویژگی های وابسته به مدل های پارامتری  
□ ضرایب مدل های  $AR$ ،  $MA$ ،  $ARMA$

□ ویژگی های مربوط به گراف ارتباطات (کارکردی و مؤثر) مغزی

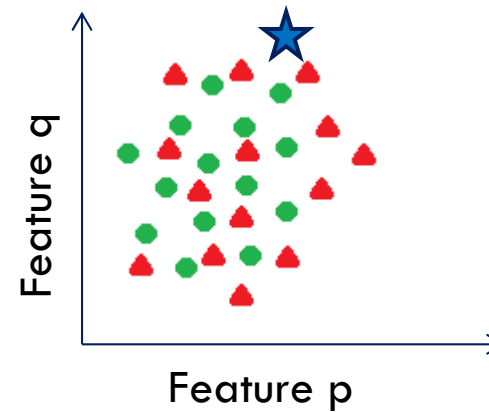
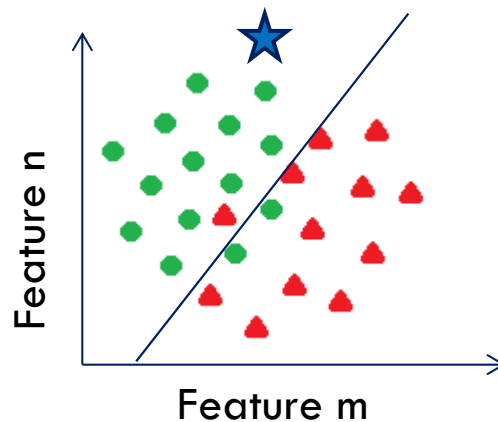
□ ویژگی های حوزه آشوب (ویژگی های غیرخطی)  
□ بعد همبستگی، بعد فرکتال، نمای لیپانوف، آنتروپی



# ویژگی ها و طبقه بندی



9



- انتخاب بردار ویژگی های مؤثر مهم است.
- در اینجا فرض می کنیم ویژگی های مناسبی را استخراج کرده ایم و در مورد روش های طبقه بندی صحبت می کنیم.
- نحوه انتخاب ویژگی های بهینه پس از مبحث طبقه بندی توضیح داده می شود.

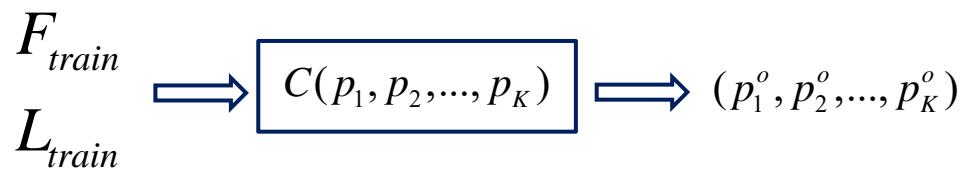
# طبقه‌بندی



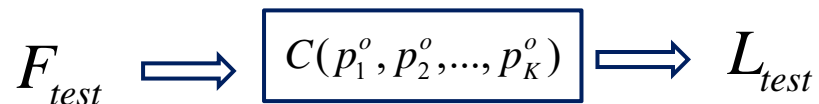
10

□ طبقه‌بندی کننده  $C$  با پارامترهای  $(p_1, p_2, \dots, p_K)$

□ آموزش



□ آزمون



□ طبقه‌بندی کننده‌ها:

□ کمترین فاصله اقلیدسی

□  $K$  نزدیکترین همسایگی (k-NN)

□ ماشین بردار پشتیبان (SVM)

□ LDA

□ شبکه‌های عصبی

□ ...



# طبقه‌بندی کننده کمترین فاصله اقلیدسی

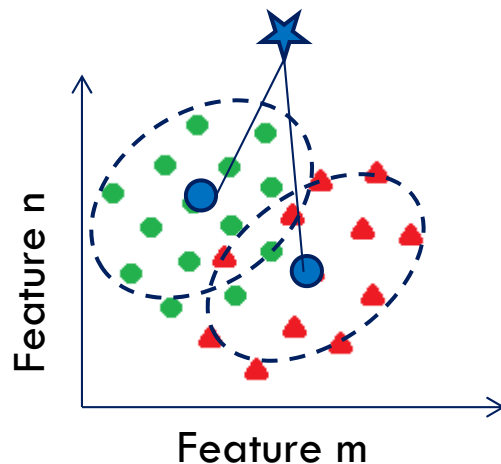
□ آموزش:

□ میانگین داده‌های هر کلاس را محاسبه می‌کنیم:  $\mu_1, \mu_2, \dots, \mu_C$

□ آزمون:

□ برای هر داده تست، فاصله داده (بردار ویژگی) را از میانگین هر کلاس محاسبه می‌کنیم.

□ کلاسی انتخاب می‌شود که کمترین فاصله را داشته باشد.





# طبقه‌بندی کننده $k$ -نزدیکترین همسایگی

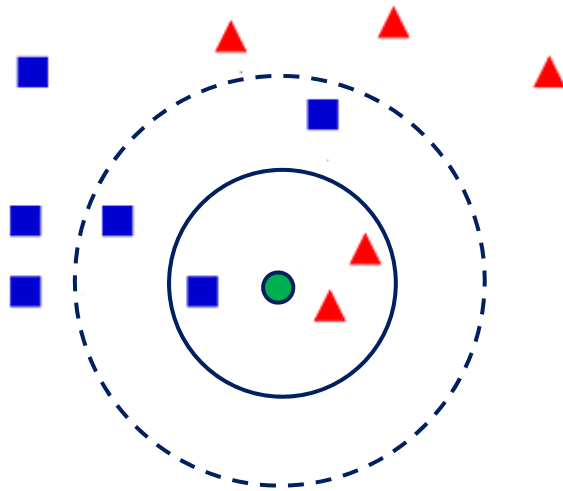
□ آموزش:

□ تعیین  $k$  مناسب با استفاده از داده‌های آموزشی.

□ آزمون:

□ برای هر داده تست،  $k$  نزدیکترین همسایه‌های آن را از بین داده‌های آموزش تعیین می‌کنیم.

□ کلاسی انتخاب می‌شود که بیشترین تکرار را داشته باشد.





# طبقه‌بندی‌کننده ماشین بردار پشتیبان خطی

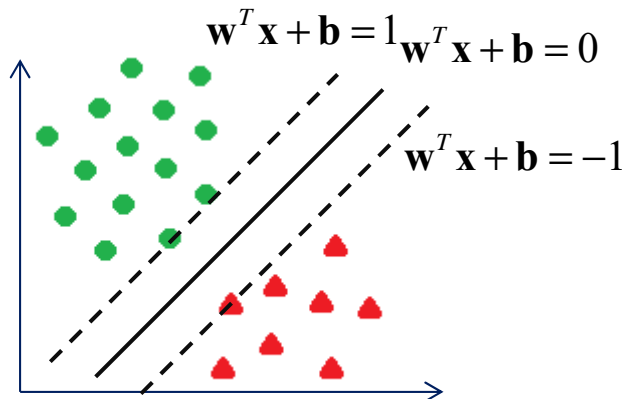
13

□ آموزش: (حالت دو کلاسه)

□ داده‌های آموزشی را با توجه به کلاس آنها با یک ابرصفحه جدا می‌نماییم.  $\mathbf{w}^T \mathbf{x} + \mathbf{b} = 0$

$\{(\mathbf{x}_1, c_1), (\mathbf{x}_2, c_2), \dots, (\mathbf{x}_M, c_M)\}$  داده‌های آموزش  $c_m \in \{-1, 1\}$

□ ابرصفحه انتخاب شده باید به گونه‌ای باشد که ماکزیمم فاصله با نزدیکترین داده‌ها از دو کلاس را داشته باشد.



□ هم‌چنین با استفاده از داده‌های آموزش داریم:

$$\mathbf{w}^T \mathbf{x}_i + \mathbf{b} \geq 1 \quad \mathbf{x}_i \in C_1$$

$$\mathbf{w}^T \mathbf{x}_i + \mathbf{b} \leq -1 \quad \mathbf{x}_i \in C_2$$

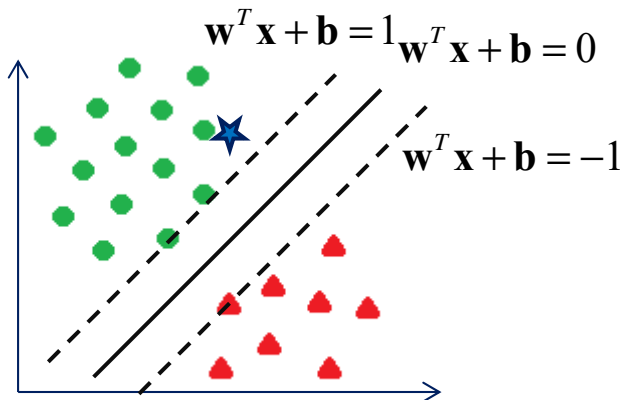


# طبقه‌بندی‌کننده بردار پشتیبان خطی

14

□ آزمون:

□ بردار ویژگی برای داده آزمون محاسبه کرده و در معادله ابرصفحه قرار می‌دهیم.



if  $\mathbf{w}^T \mathbf{x}_{Test} + \mathbf{b} > 0$  then  $\mathbf{x}_{Test} \in C_1$

if  $\mathbf{w}^T \mathbf{x}_{Test} + \mathbf{b} < 0$  then  $\mathbf{x}_{Test} \in C_2$



# معیارهای ارزیابی

- True Positive (TP) → Hit
- True Negative (TN) → Correct rejection
- False Positive (FP) → False alarm
- False Negative (FN) → Miss
- Sensitivity or true positive rate (TPR)

$$\text{Sensitivity} = \text{TPR} = \frac{TP}{P} = \frac{TP}{TP + FN}$$

- False positive rate (FPR) false alarm rate

$$\text{FPR} = \frac{FP}{N} = \frac{FP}{TN + FP}$$

- Accuracy

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN}$$

		Actual Value (as confirmed by experiment)	
		positives	negatives
Predicted Value (predicted by the test)	positives	<b>TP</b> True Positive	<b>FP</b> False Positive
	negatives	<b>FN</b> False Negative	<b>TN</b> True Negative

Confusion Matrix

# معیارهای ارزیابی



16

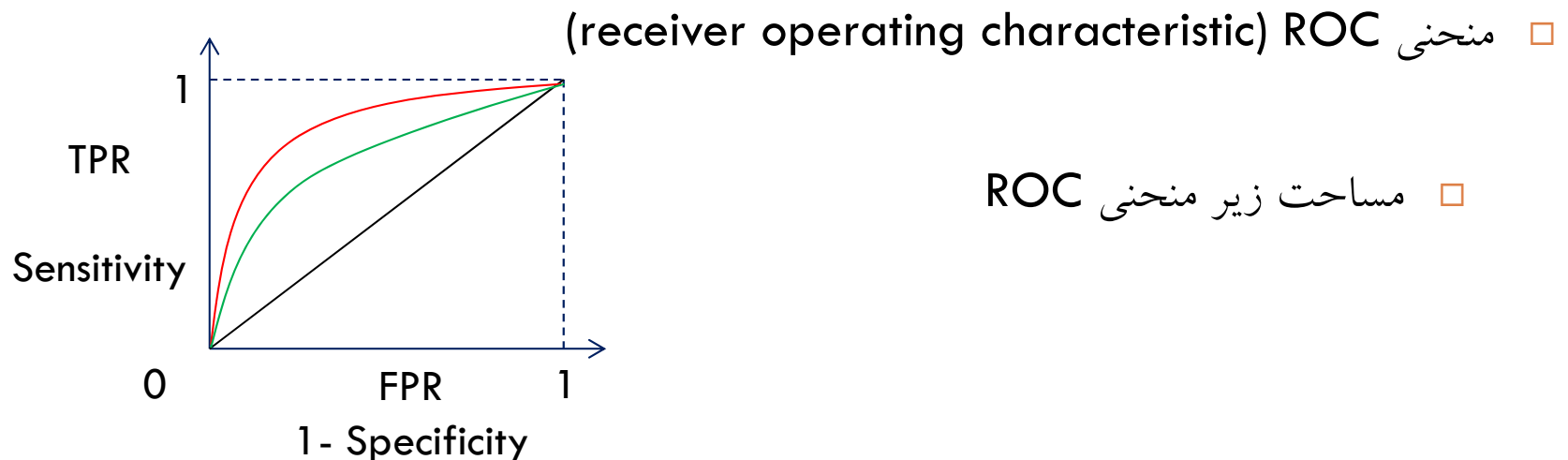
- Specificity, True negative rate

$$Specificity = TNR = \frac{TN}{N} = \frac{TN}{TN + FP} = 1 - FPR$$

- Positive predictive value, Precision

$$PPV = \frac{TP}{TP + FP}$$

		Actual Value (as confirmed by experiment)	
		positives	negatives
Predicted Value (predicted by the test)	positives	<b>TP</b> True Positive	<b>FP</b> False Positive
	negatives	<b>FN</b> False Negative	<b>TN</b> True Negative







# کاهش تعداد ویژگی‌ها

□ تعداد داده‌های آموزشی:  $N$

□ بعد ویژگی‌ها:  $L$

□ تعداد پارامترهای طبقه‌بندی کننده:  $f(L)$

$$\frac{N}{L} > 10 \quad \Rightarrow \quad L < \frac{N}{10}$$

□ مزیت‌های کاهش بعد ویژگی‌ها:

□ کاهش پیچیدگی محاسباتی/کاهش تعداد پارامترهای طبقه‌بندی کننده

□ حذف ویژگی‌های زائد/همبستگی بین ویژگی‌ها

□ قدرت تعمیم‌پذیری نسبت مستقیم با  $N/f(L)$  دارد.

# کاهش تعداد ویژگی‌ها



18

□ دو دسته روش برای کاهش تعداد ویژگی‌ها وجود دارد:

□ کاهش بعد به روش ترکیبی

PCA ■

FLD یا LDA ■

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_P \end{bmatrix} \Rightarrow \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_Q \end{bmatrix} = f \left( \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_P \end{bmatrix} \right)$$

□ کاهش بعد به روش گزینشی

■ معیار مناسب برای انتخاب ویژگی‌ها

■ روش مناسب برای جستجوی ویژگی‌ها

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_P \end{bmatrix} \Rightarrow \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iQ} \end{bmatrix}$$

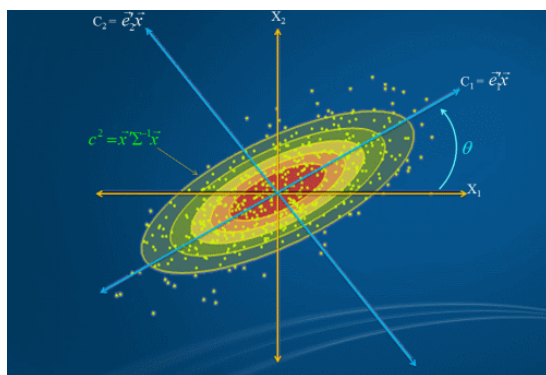
# کاهش بعد به روش ترکیبی



19

## □ PCA: (Principal Component Analysis)

□ آنالیز مولفه‌های اصلی یک تبدیل خطی متعامد است که داده را به دستگاه مختصات جدید می‌برد به طوری که بزرگترین واریانس داده بر روی اولین محور مختصات، دومین بزرگترین واریانس بر روی دومین محور مختصات قرار می‌گیرد و همین طور برای بقیه.



## □ کاهش ابعاد داده با استفاده از PCA:

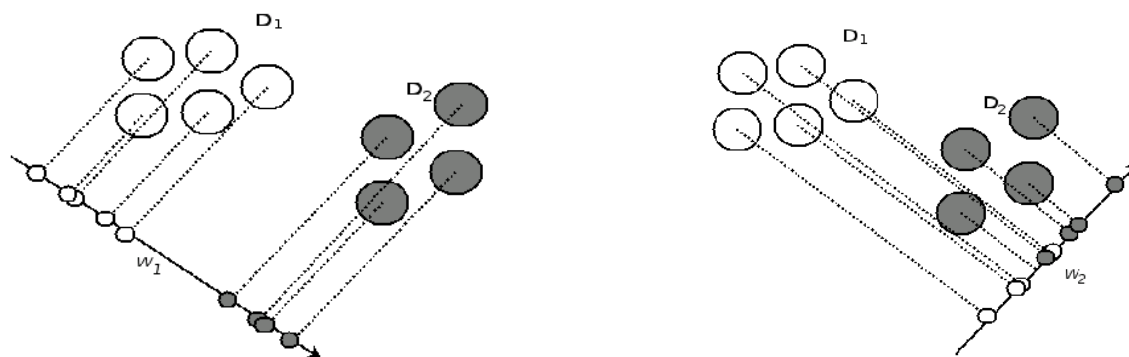
□ مولفه‌هایی از مجموعه داده را که بیشترین تاثیر در واریانس را دارند حفظ می‌کنیم و بقیه مولفه‌ها را حذف می‌کنیم.

# کاهش بعد به روش ترکیبی



20

## (Fisher's linear Discriminant) FLD □



□ با داشتن ویژگی‌های استخراج شده از داده‌های آموزشی و کلاس متناظر با هر داده (دو کلاس)، هدف یافتن نگاشت خطی داده‌ها به فضای یک بعدی است به گونه‌ای که بیشترین تفکیک‌پذیری را داشته باشند.

□ مجموعه ویژگی‌های استخراج شده از داده‌های آموزشی:  $\{x \in R^D\}$

□ نگاشت خطی:  $y = w^T x$

# کاهش بعد به روش ترکیبی



21

□ میانگین کلاس  $i$ -ام:

$$\mu_i = \frac{1}{N_i} \sum_{x \in C_i} x$$

□  $N_i$  تعداد داده‌ها در کلاس  $i$ -ام و  $C_i$  کلاس  $i$ -ام است.

□ برای نقاط نگاشت شده، میانگین از رابطه زیر به دست می‌آید:

$$\tilde{\mu}_i = \frac{1}{N_i} \sum_{y \in Y_i} y = \frac{1}{N_i} \sum_{x \in C_i} w^T x = w^T \mu_i$$

□ می‌دانیم که اختلاف بین میانگین‌ها، همیشه معیار مناسبی از میزان جدایی کلاس‌ها در فضای جدید نیست:

$$|\tilde{\mu}_1 - \tilde{\mu}_2| = |w^T (\mu_1 - \mu_2)|$$

$$\|w\| \uparrow \Rightarrow |\tilde{\mu}_1 - \tilde{\mu}_2| \uparrow$$

# کاهش بعد به روش ترکیبی



22

□ ماتریس پخشی درون کلاسی:

$$\tilde{S}_1^2 + \tilde{S}_2^2$$

$$\tilde{S}_i^2 = \sum_{y \in Y_i} (y - \tilde{\mu}_i)^2 = w^T \underbrace{\left( \sum_{x \in c_i} (x - \mu_i)(x - \mu_i)^T \right)}_{S_i} w$$

□ FLD،  $w$  را به گونه‌ای پیدا می‌کند که:

$$w = \arg \max_w \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{S}_1^2 + \tilde{S}_2^2} = \arg \max_w \frac{w^T S_B w}{w^T S_w w}$$

□ ماتریس پخشی درون کلاسی

$$S_w = S_1 + S_2$$

□ ماتریس پخشی میان کلاسی

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

# کاهش بعد به روش ترکیبی



23

□ یافتن  $w$  ای که کسر را ماکزیمم کند:

$$\arg \max_w \frac{w^T S_B w}{w^T S_w w} \Rightarrow S_B w = \lambda S_w w$$

□ مسئله مقدار ویژه تعمیم یافته Generalized EigenValue Problem

# کاهش بعد به روش گزینشی



24

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_P \end{bmatrix} \longrightarrow \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iQ} \end{bmatrix}$$

□ کاهش بعد به روش گزینشی

□ معیارهای جستجو

□ الگوریتم‌های جستجو

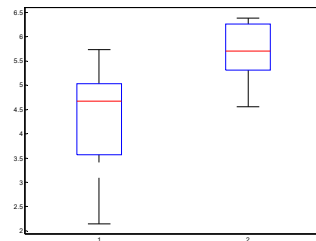


# کاهش بعد به روش گزینشی

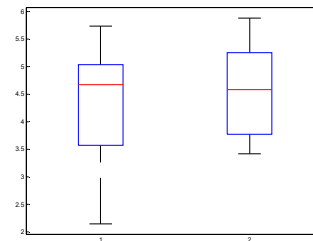


25

□ معیار مناسب برای انتخاب ویژگی‌ها



ویژگی z-ام



ویژگی i-ام

□ معیارهای بدون سرپرست، غیروابسته به طبقه‌بندی‌کننده

□ معیارهای باسرپرست، غیروابسته به طبقه‌بندی‌کننده

□ معیارهای باسرپرست، وابسته به طبقه‌بندی‌کننده

# کاهش بعد به روش گزینشی



26

□ معیارهای با سرپرست، غیر وابسته به طبقه‌بندی کننده:

□ معیار وابسته به ماتریس‌های پخشی (Scattering Matrices)

□ یک بردار ویژگی شامل  $P$  ویژگی داریم. می‌خواهیم بررسی کنیم آیا این دسته ویژگی، مناسب است یا خیر. برچسب هر آزمایش را نیز داریم.

□ ماتریس پخشی درون کلاسی:

$$S_1 = \sum_{i \in C_1} (\mathbf{x}_i - \boldsymbol{\mu}_1)(\mathbf{x}_i - \boldsymbol{\mu}_1)^T$$
$$S_W = S_1 + S_2$$

$$S_2 = \sum_{i \in C_2} (\mathbf{x}_i - \boldsymbol{\mu}_2)(\mathbf{x}_i - \boldsymbol{\mu}_2)^T$$

□ ماتریس پخشی میان کلاسی:

$$S_b = \sum_{i=1}^2 (\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)^T$$

□ معیارهای تعریف شده:

$$J_1 = \frac{\text{trace}(S_b)}{\text{trace}(S_W)} \quad J_2 = \frac{|S_b|}{|S_W|}$$

□ هر چه این معیارها بزرگتر باشند نشان می‌دهد که دسته ویژگی مورد بررسی مناسب‌تر است.

# کاهش بعد به روش گزینشی



27

□ معیارهای با سرپرست، وابسته به طبقه‌بندی کننده:

□ معیار مبتنی بر صحت طبقه‌بندی کننده

□ یک بردار ویژگی شامل  $P$  ویژگی داریم. می‌خواهیم بررسی کنیم آیا این دسته ویژگی، مناسب است یا خیر. برچسب هر آزمایش را نیز داریم. هم‌چنین می‌خواهیم از طبقه‌بندی کننده  $C$  استفاده کنیم.

□ تقسیم داده‌های آموزشی به دو دسته:

داده‌های آموزشی

آموزشی-آزمون	آموزشی-آموزشی
--------------	---------------

□ داده‌های آموزشی-آموزشی

□ داده‌های آموزشی-آزمون

□ آموزش طبقه‌بندی کننده بر مبنای داده‌های آموزشی-آموزشی

□ محاسبه صحت طبقه‌بندی کننده بر روی داده‌های آموزشی-آزمون

□ هر چه قدر صحت طبقه‌بندی کننده بیشتر باشد، بردار ویژگی مناسب‌تر است.

# انتخاب ویژگی به روش گزینشی



28

□ الگوریتم‌های جستجو برای یافتن مناسب‌ترین دسته ویژگی:

□ روش جستجوی فیلتری

□ معیارها بر روی تک تک ویژگی‌ها اعمال شده و ویژگی‌هایی که بهترین معیار را دارند، انتخاب می‌شوند.

□ معیارهای به کار رفته در این روش معمولاً معیارهای غیروابسته به طبقه‌بندی‌کننده می‌باشند.

□ روش جستجوی پوششی

□ معیارها معمولاً معیارهای وابسته به طبقه‌بندی‌کننده هستند.

□ دسته ویژگی‌های مناسب با توجه به اثر آنها بر روی یکدیگر انتخاب می‌شوند.

□ بررسی تمام حالت‌های ممکن، غیر عملی است، در نتیجه از روش‌های جستجو که زمان کمتری نیاز دارند استفاده می‌شود مانند:

■ روش‌های جستجوی پی‌درپی

■ روش‌های جستجوی تصادفی مانند الگوریتم ژنتیک

# انتخاب ویژگی به روش گزینشی



29

□ مثالی از روش‌های جستجوی پی‌درپی: جستجوی پیشروی پی‌درپی

- ۱. از مجموعه تهی شروع می‌کنیم.
- ۲. ویژگی‌ای را به مجموعه ویژگی‌های فعلی اضافه می‌کنیم که به کمک آن بهترین معیار به دست آید.
- ۳. اگر به یک شرط نهایی (مثلاً تعداد ویژگی‌ها) برای اتمام الگوریتم نرسیده‌ایم، به مرحله ۲ بازگشته و در غیر این صورت الگوریتم پایان می‌یابد.

□ دیگر الگوریتم‌های جستجوی پی‌درپی

- جستجوی پسروی پی‌درپی
- جستجوی شناور پی‌درپی

## چند نکته مهم



30

- بر چسب داده‌های آزمون را در اختیار نداریم، از کجا بفهمیم ویژگی‌ها و طبقه‌بندی‌کننده‌ای که استفاده کرده‌ایم مناسب است؟
  - تقسیم داده‌های آموزشی به دو دسته آموزشی-آموزشی و آموزشی-آزمون (Hold out)
  - استفاده از روش **k-fold cross-validation**
  - استفاده از روش **Leave one-out cross-validation**
- در روش‌های کاهش بعد ترکیبی، ماتریسی که برای کاهش بعد داده‌های آموزشی تعیین شد باید برای داده‌های آزمون استفاده شود.
  - نباید برای داده‌های آزمون دوباره محاسبه شود.
- از آنجا که ویژگی‌های مختلف در محدوده‌های متفاوتی هستند، برای هر چه بهتر عمل کردن طبقه‌بندی‌کننده‌ها باید ویژگی‌های استخراج شده نرمالیزه شوند:
  - مثلاً می‌توان تمام ویژگی‌ها را در محدوده  $[-1, 1]$  نرمالیزه کرد یا **z-score** را محاسبه کرد.



با تشکر