ارتباطات مغزی

BRAIN CONNECTIVITY

شماره درس: ۲۵۶۳۰

یکشنبه و سه‌شنبه ۱۵–۱۳:۳۰

نیم‌سال اول ۱۴۰۱–۱۴۰۰

مقدمه

- در گذشته، بر پایه نظریه تفکیک کارکردی مغز (Brain Functional Segregation)، بسیاری از تحقیقات حوزه علوم اعصاب به دنبال یافتن نواحی مجزای مغزی بودند که هر یک مسئول انجام فعالیت‌های متمایزی هستند.
- پیدایش نظریه مکملی با عنوان نظریه یکپارچگی کارکردی مغز (Brain Functional Integration) بسیاری از تحقیقات اخیر در حوزه علوم اعصاب را به سمت حوزه ارتباطات مغزی سوق داده است.
- بر پایه نظریه یکپارچگی کارکردی مغز، حتی در ساده‌ترین فعالیت‌های مغزی مانند وضعیت استراحت مغز (Brain resting state)، ارتباطاتی میان فعالیت نواحی مجزای مغزی وجود دارد تا سیستم مغزی عملکرد هماهنگ و یکپارچه‌ای داشته باشد.

مقدمه

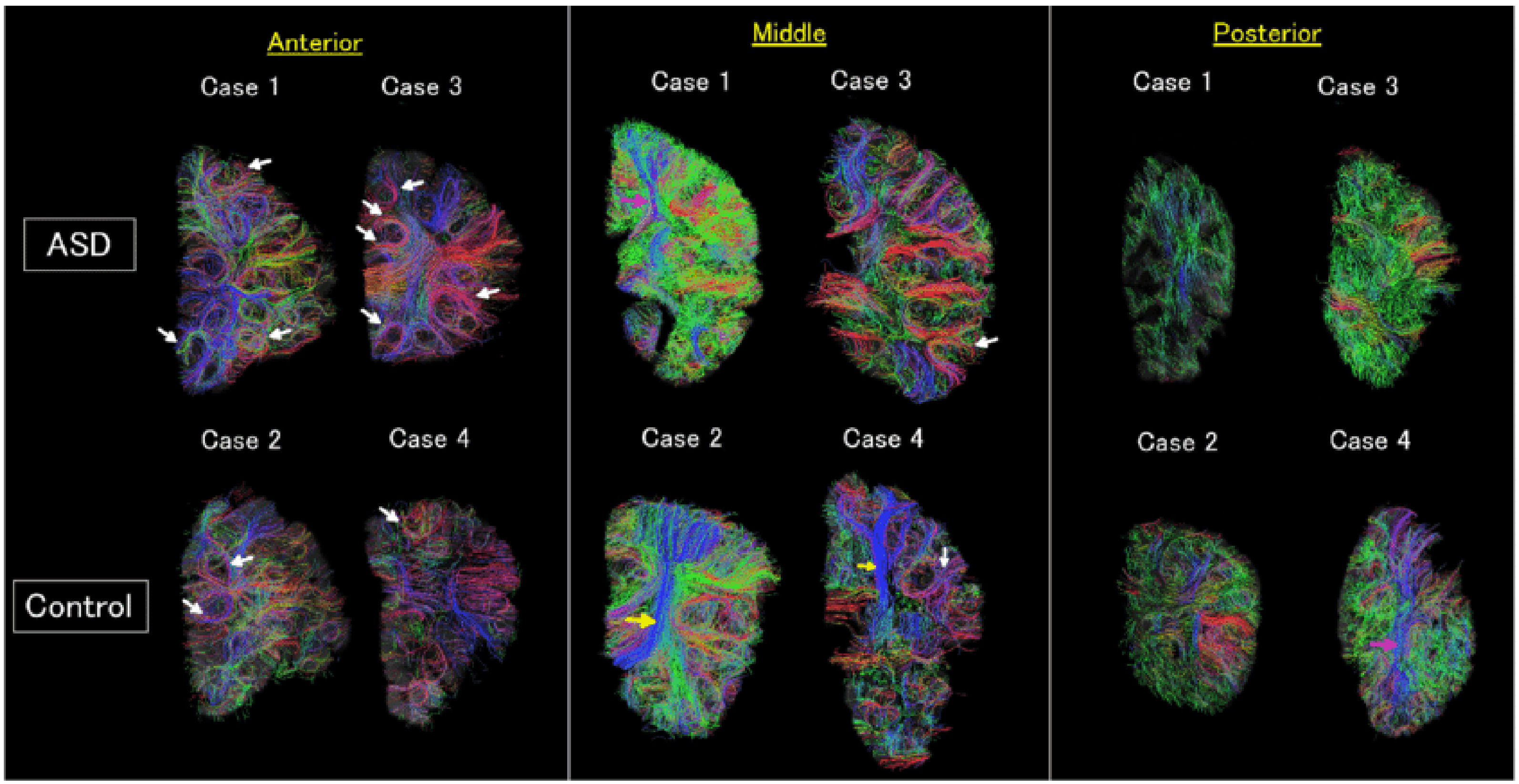
- سه شاخه اصلی ارتباطات مغزی:
  - ارتباطات ساختاری (Structural connectivity)
    - شبکه فیبرهای عصبی که نواحی مغزی را به هم مرتبط می‌کنند
  - ارتباطات کارکردی (Functional connectivity)
    - وابستگی‌های آماری و بدون جهت بین فعالیت نواحی مغزی
  - ارتباطات مؤثر (Effective connectivity)
    - ارتباطات علّی و جهت‌دار بین فعالیت نواحی مغزی
- از بین تکنیک‌های مختلف تصویربرداری کارکردی مغز، EEG به دلیل رزولوشن زمانی بالا، در دسترس بودن، کم حجم بودن و ارزان بودن مناسب‌ترین گزینه برای تخمین ارتباطات کارکردی/مؤثر مغزی است.

ارتباطات ساختاری

- ارتباطات ساختاری (آناتومیکی) به مجموعه‌ای از ارتباطات فیزیکی یا ساختاری (سیناپسی) اطلاق می‌شود که نورون‌های شبکه عصبی را به یکدیگر مرتبط می‌کنند.
- ارتباطات ساختاری را می‌توان در مقیاس‌های مختلفی مورد بررسی قرار داد:
   - از ارتباطات ساختاری بین شبکه‌های محلی مغز تا ارتباطات ساختاری بین شبکه‌های بزرگ مغزی و در واقع ارتباطات آناتومیکی بین نواحی مغزی.
- همچنین این ارتباطات در مقیاس‌های زمانی کوتاه (از مرتبه ثانیه تا دقیقه) غالباً استاتیک در نظر گرفته شوند در حالی‌که در مقیاس‌های زمانی بزرگ (از مرتبه ساعت تا روز برای مثال طی یک فرایند یادگیری) ممکن است پلاستیک یا دینامیک باشند.

ارتباطات ساختاری

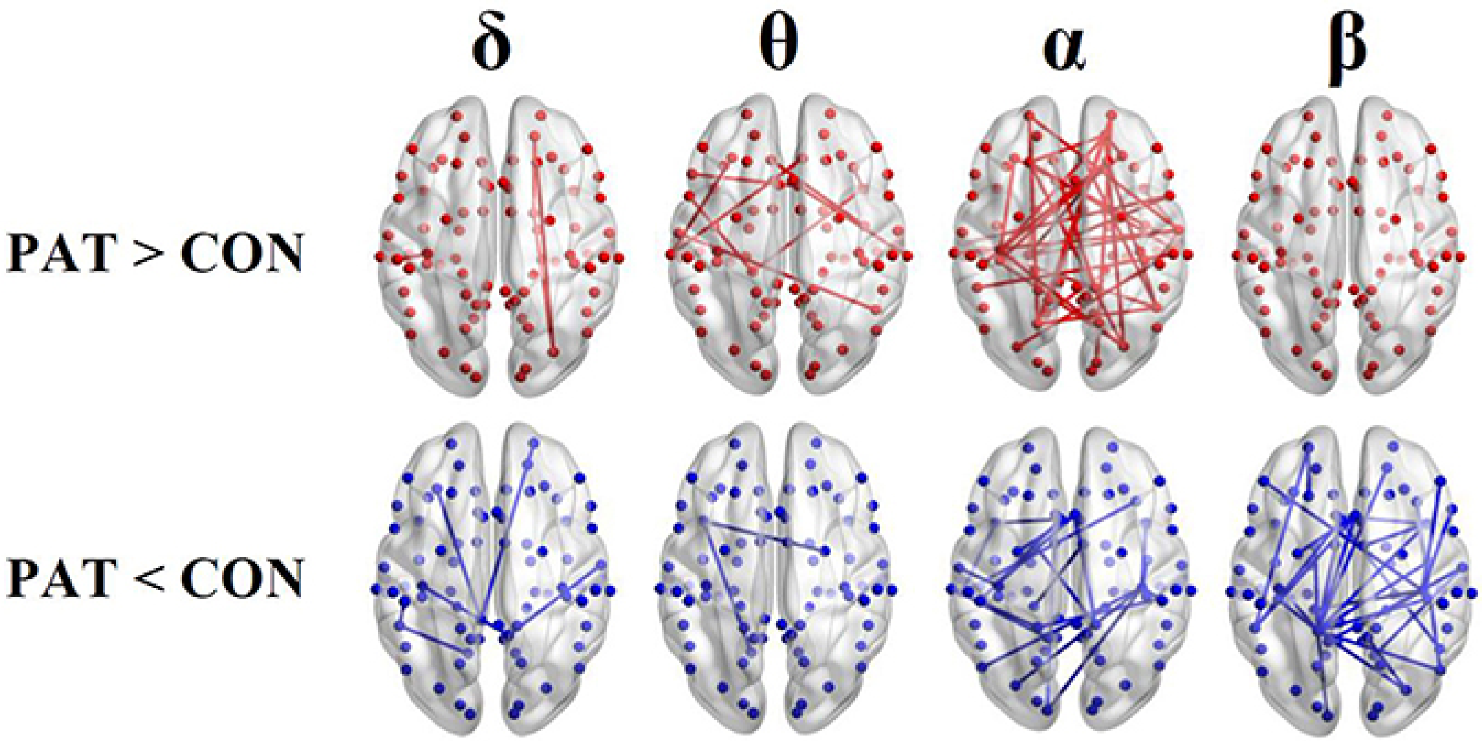
- یکی از ابزارهایی که برای بررسی ارتباطات ساختاری مورد استفاده قرار می‌گیرد تصویربرداری تانسور انتشار (DTI) است که به کمک آن مسیر فیبرهای عصبی تخمین زده می‌شود.



ارتباطات کارکردی

- ارتباطات کارکردی به الگوهای وابستگی آماری بین فعالیت واحدهای نورونی توزیع شده در مغز اطلاق می‌شود.
- یکی از ساده‌ترین ابزارهای مورد استفاده در این زمینه، بررسی همبستگی بین فعالیت واحدهای نورونی است.
- ارتباطات کارکردی معمولاً به شدت وابسته به زمان هستند (از مقیاس چند صد میلی‌ثانیه) و هیچ اطلاعاتی در زمینه ارتباطات علّت و معلولی فعالیت نواحی مغزی فراهم نمی‌کنند.
- نظریه یکپارچگی کارکردی: نواحی مجزا و تخصیص‌یافته مغزی در حین انجام عملیات و فرایندهای مشخص با یکدیگر ارتباط و هماهنگی دارند به نحوی‌که در نتیجه این ارتباطات، همگی این نواحی در قالب سیستمی یکپارچه و هماهنگ فعالیت می‌کنند.

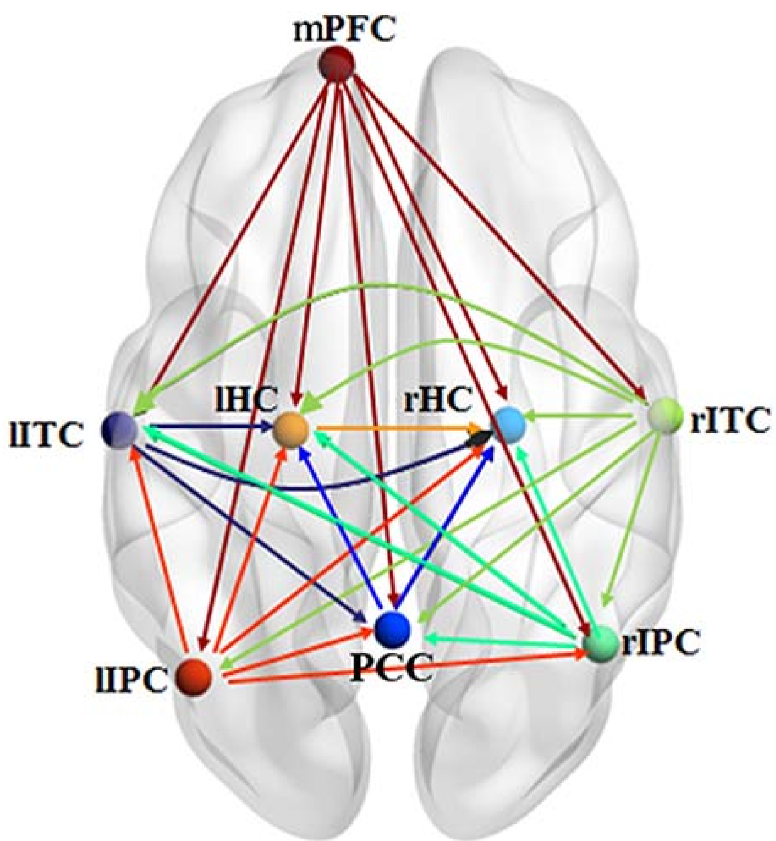
ارتباطات کارکردی

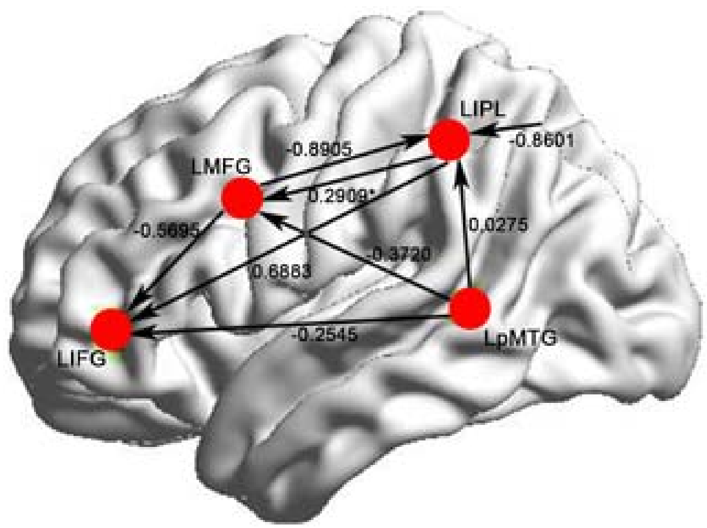


ارتباطات مؤثر

- ارتباطات مؤثر به شبکه ارتباطات علّت و معلولی در یک سیستم عصبی اطلاق می‌شود.
- در واقع این ارتباطات به بررسی تأثیرات علّت و معلولی نواحی مغزی بر هم می‌پردازد و به دنبال تعیین این است که کدام سیستم عصبی کدام سیستم دیگر را راه‌اندازی و فعال می‌کند.
- اگر چه بررسی ارتباطات مؤثر نسبت به ارتباطات کارکردی غالباً نیازمند روش‌های پیچیده‌تر با حجم محاسباتی بالاتری است ولی به نحوی اطلاعات سطح بالاتری نسبت به ارتباطات کارکردی فراهم می‌کند.
- با توجه به این که علت باید مقدم بر معلول باشد، در بسیاری مواقع از واژه ارتباطات تأخیری مترادف با ارتباطات علّی یا ارتباطات مؤثر استفاده می‌شود.

ارتباطات مؤثر





سیگنال‌های EEG

- داده‌های EEG به دلیل رزولوشن زمانی بالا (از مرتبه میلی ثانیه) ساختار دینامیکی فعالیت‌های مغزی را به خوبی ثبت می‌کنند و از این حیث داده‌های مناسبی برای بررسی ارتباطات کارکردی/مؤثر مغزی هستند.

- محدودیت اساسی سیگنال‌های EEG این است که از طریق آنها برآیند فعالیت‌های الکتریکی سیستم عصبی در کانال‌ها ثبت می‌شود. در نتیجه امکان تشخیص دقیق و قطعی محل منابع فعال مغز و دامنه فعالیت آنها امکان‌پذیر نیست.

- برای کاهش این مشکل، گاهی از رویکرد حل مسأله معکوس استفاده می‌شود تا با در نظر گرفتن یک سری قیود و فرضیات، نگاشت یکتایی از توزیع مکانی فعالیت‌های مغزی حاصل شود. البته این رویکرد از فرضیات محدود کننده‌ای استفاده می‌کند که ممکن است در عمل صادق نباشند.

سیگنال‌های EEG

- همچنین حتی در صورتی‌که از تعداد زیادی کانال EEG (مثلاً ۲۵۶ تا) استفاده شود، رزولوشن مکانی پاسخ مسأله معکوس بسیار کم و از مرتبه سانتیمتر می‌باشد که چندان مناسب نیست.
- در ضمن، حتی اگر محل منابع مغزی با دقت قابل قبولی تخمین زده شود، تخمین دقیق دامنه فعالیت آنها -که برای تخمین ارتباطات کارکردی/مؤثر مغزی ضروری است- غالباً تضمین نمی‌شود.

- مشکلات استفاده از داده‌های EEG برای تخمین ارتباطات مغزی:
  - اثر هدایت حجمی (Volume conduction effect)
  - اثر مرجع مشترک فعال (Active common reference effect)

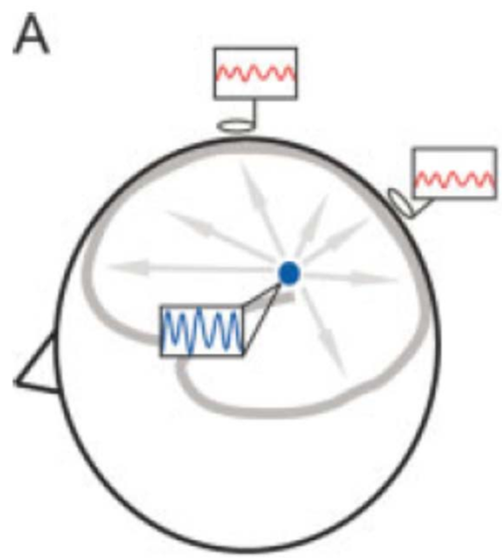
اثر هدایت حجمی

- با توجه به این که سیگنال‌های EEG عمدتاً توسط فعالیت الکتریکی نورون‌های قشر مخ ایجاد می‌شوند، معمولاً سیگنال هر کانال EEG را معرّف فعالیت منابع نورونی ناحیه قشری زیر آن در نظر می‌گیرند و مستقیماً ارتباطات کارکردی یا مؤثر بین کانال‌ها را تخمین می‌زنند.
- با این وجود یک پخش‌شدگی میدان الکتریکی وجود دارد که معمولاً از آن با عنوان اثر هدایت حجمی یاد می‌شود.
- در نتیجه، ممکن است سیگنال هر کانال برآیندی از فعالیت الکتریکی چندین منبع نورونی مغز باشد (به خصوص منابع مغزی نزدیک‌تر به هر کانال).
- این امر می‌تواند منجر به تعبیر نادرست وجود ارتباط بین نواحی قشری زیر این کانال‌ها شود. حتی در حالتی‌که فعالیت منابع نورونی زیر کانال‌ها از هم مستقل باشند، یا حتی منابع مذکور فعالیتی نداشته باشند، ممکن است ارتباط کارکردی/مؤثر نامطلوبی بین کانال‌های روی آنها مشاهده شود.

اثر هدایت حجمی

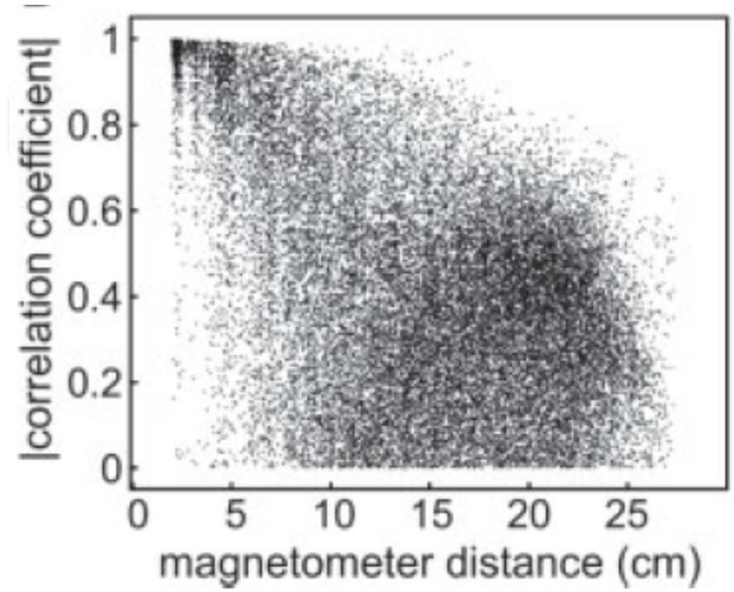
- در نتیجه ممکن است ارتباطات تخمینی بین کانالی به جای آنکه نشانگر ارتباطات منابع مغزی نزدیک آنها باشند، تنها به دلیل آرتیفکت هدایت حجمی ایجاد شده باشند.

- شماتیک نشان‌دهنده اینکه فعالیت یک منبع نورونی می‌تواند توسط سنسورهای مختلف ثبت شود:

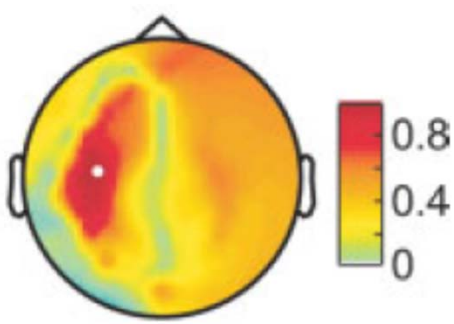


اثر هدایت حجمی

- مقدار قدرمطلق ضریب همبستگی بین همه زوج سیگنال‌های ثبت‌شده بر حسب فاصله:



- نمایش توپوگرافیک میزان همبستگی با سنسور مرجع که با رنگ سفید مشخص شده است:



اثر هدایت حجمی

- اثر هدایت حجمی به صورت یک برهم‌نهی خطی و لحظه‌ای (بدون تأخیر زمانی یا به طور معادل بدون شیفت فازی در حوزه فرکانس) فعالیت منابع نورونی مغز می‌باشد.

- سیگنال کانال i–ام با فرض وجود M منبع مغزی و با صرف‌نظر از نویز اندازه‌گیری:

$$x_i(t) = \sum_{m=1}^{M} a_{im} . s_m(t);\; i = 1,\ldots,N$$

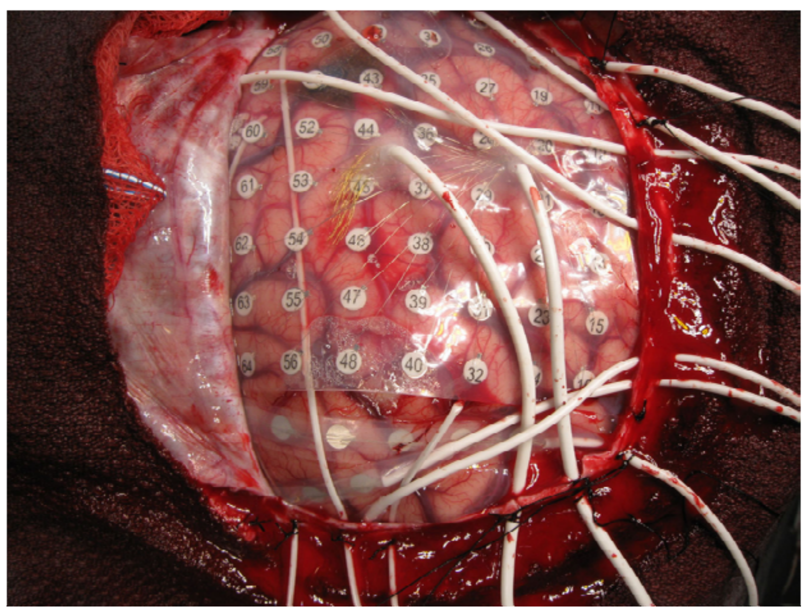
- ضرایب a_{im} معرف اثر هدایت حجمی از منبع m–ام به کانال i–ام هستند.

اثر هدایت حجمی

- راهکارهای کاهش اثرات نامطلوب هدایت حجمی در تخمین ارتباطات مغزی:

۱- ثبت سیگنال‌های تهاجمی الکتروکورتیکوگرام (ECoG) و محاسبه ارتباطات کارکردی/مؤثر بین این سیگنال‌ها

  - این روش برای انسان معمولاً به کار نمی‌رود.
  - نقش اثر هدایت حجمی تضعیف می‌شود ولی به طور کامل از بین نمی‌رود.



اثر هدایت حجمی

۲- تخمین موقعیت مکانی و دامنه فعالیت منابع مغزی با حل مسأله معکوس و سپس تخمین ارتباطات کارکردی/مؤثر بین دامنه فعالیت تخمینی نواحی به دست آمده

- مسئله معکوس جواب یکتایی ندارد و برای دستیابی به جواب یکتا لازم است قیود زمانی و مکانی مختلفی برای فعالیت منابع مغزی فرض شود. قیود مربوطه ممکن است در عمل صادق نباشند و این امر می‌تواند منجر به تخمین نادرست مکان و دامنه فعالیت منابع مغزی شود.
- همچنین روش‌های متداول حل مسئله معکوس EEG (مانند LORETA) عمدتاً بر تخمین دقیق موقعیت مکانی منابع مغزی تمرکز دارند و لزوماً تخمین دقیق دامنه فعالیت منابع مغزی را که برای تخمین ارتباطات بین منابع حیاتی است، تضمین نمی‌کنند.
- برای دستیابی به پاسخ‌های قابل اطمینان‌تر به ثبت‌هایی با تعداد کانال بالا (بیشتر از ۶۴ کانال)، مدل واقعی سر هر سوژه (به دست آمده از تصاویر MRI) و تعیین محل دقیق کانال‌ها نیازمندیم.

اثر هدایت حجمی

۳- استفاده از تخمینگرهای ارتباطی «مقاوم به آرتیفکت هدایت حجمی»

- این تخمینگرها به صورت نظری هرگز به خاطر هدایت حجمی منابع مستقل تخمین‌های معنی‌داری نمی‌دهند. در نتیجه، هر ارتباط تخمینی معنی‌دار آن‌ها ناشی از اندرکنش‌های تأخیری مغزی و نشانگر ارتباطات واقعی بین منابع مغزی است.
- اثر هدایت حجمی، به صورت ترکیب خطی بوده و برای باندهای فرکانسی معمول EEG (کوچکتر از ۱۰۰ هرتز) ماهیت لحظه‌ای دارد. در نتیجه، برای این که تخمینگری مقاوم به اثر هدایت حجمی باشد کافی است که ترکیب خطی لحظه‌ای منابع مستقل هرگز تخمین‌های ارتباطی معنی‌داری بین کانال‌ها ایجاد نکند.
- به بیان دیگر، برای مقاوم بودن یک تخمینگر به آرتیفکت هدایت حجمی کافی است به هرگونه ترکیب خطی لحظه‌ای مؤلفه‌های مستقل، مقاوم باشد.

اثر مرجع مشترک فعال

- ثبت سیگنال‌های EEG از طریق تقویت کننده‌های تفاضلی که دو ورودی هر یک از آنها سیگنال یک کانال و سیگنال یک مرجع مشترک می‌باشد، انجام می‌گیرد.
- مطلوب این است که سیگنال مرجع (که معمولاً سیگنال یک کانال روی سر است) هیچ فعالیت مغزی را ثبت نکند و تنها آرتیفکت‌های ناشی از سیگنال‌های قلبی و ... را که در سایر کانال‌های EEG تقریباً مشترک است در بر داشته باشد تا بدین وسیله حذف آرتیفکت انجام شود.
- اگر سیگنال مرجع مشترک، فعالیت مغزی قابل توجهی را ثبت کند و در واقع اگر مرجع مشترک فعال باشد، می تواند منجر به ظهور ارتباطات کارکردی/مؤثر نادرستی بین سیگنال‌های ثبت شده شود.

اثر مرجع مشترک فعال

○ راهکارهای کاهش اثرات نامطلوب مرجع مشترک فعال در تخمین ارتباطات مغزی:

۱- انتخاب یک کانال به عنوان مرجع مشترک از بین کانال‌هایی که در ثبت مورد نظر EEG فعالیت مغزی خاصی را ثبت نمی‌کنند (کانال های غیرفعال).

۲- انتخاب کانال‌هایی که روی استخوان ماستوئید گوش نصب شده باشند چون معمولاً به دلیل ضخامت بیشتر استخوان زیر آن‌ها، فعالیت مغزی ضعیف‌تری ثبت می‌کنند و در عین حال روی سر قرار دارند و آرتیفکت‌های مشترک را ثبت می‌کنند.

۳- عدم استفاده از مرجع مشترک میانگین (متوسط همه کانال‌ها به عنوان مرجع مشترک) مگر در حالتی که تعداد کانال‌ها حداقل ٦٤ باشد.

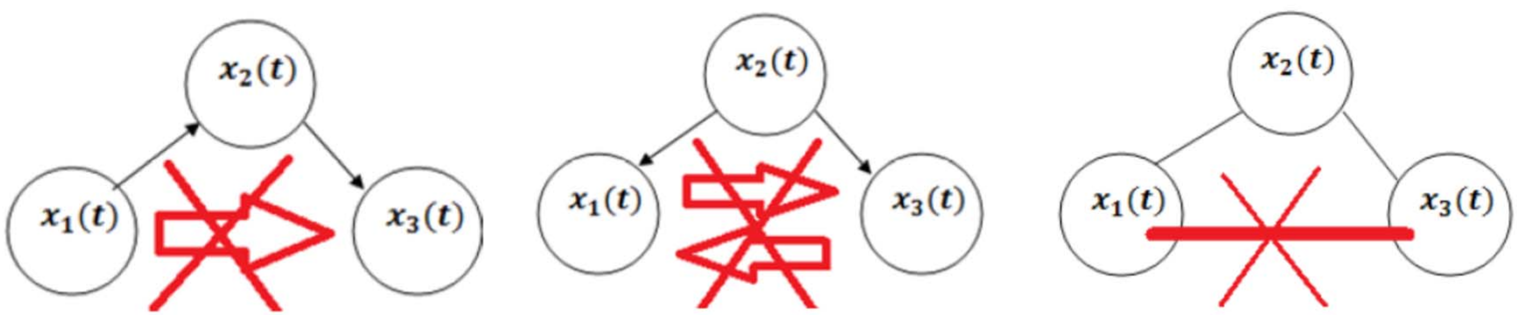
٤- استفاده از تخمینگرهای ارتباطی که حساسیت کمتری به اثر مرجع مشترک فعال داشته باشند.

تخمینگرهای ارتباطات EEG

○ مهمترین قابلیت‌های مطلوب تخمینگرهای ارتباطات کارکردی/مؤثر:

۱- قابلیت تخمین کلیه ارتباطات خطی و غیرخطی

۲- قابلیت حذف ارتباطات غیرمستقیم



۳- قابلیت تخمین شبکه ارتباطات داده‌های با بعد بالا یا دارای تاخیر طولانی

٤- حساسیت بالا به ارتباطات منابع مغزی در حضور اثر هدایت حجمی

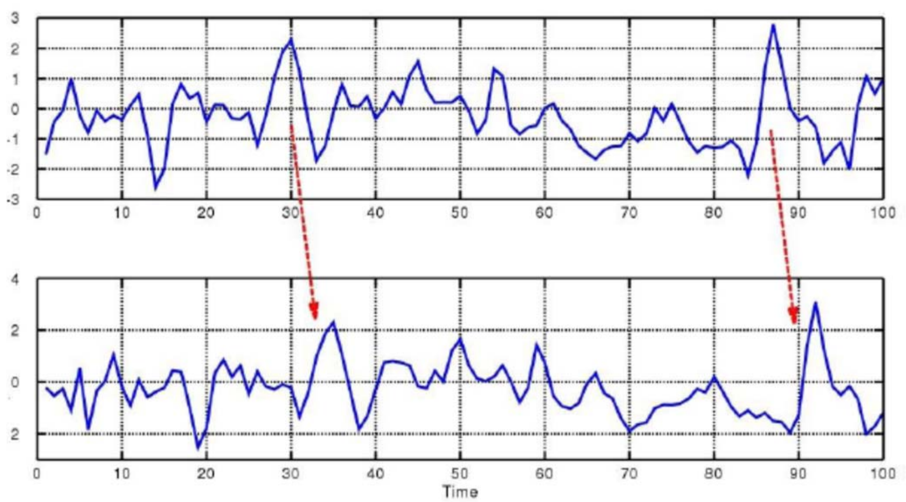
تخمینگرهای ارتباطات EEG

- تخمینگرهای ارتباطات کارکردی:
  - ضریب همبستگی پیرسون (Correlation Coefficient)
  - اطلاعات متقابل (Mutual Information)
  - اطلاعات متقابل جزئی (Partial Mutual Information)
  - همدوسی (Coherence)
  - همدوسی جزئی (Partial Coherence)
  - مقدار قفل‌شدگی فاز (Phase Locked Value)
  - جزءِ موهومی کوهرنسی (ImC)
  - شاخص تأخیر فاز (Phase Lag Index)

تخمینگرهای ارتباطات EEG

- تخمینگرهای ارتباطات مؤثر:

  - علّیت، مفهومی پیچیده و چالش برانگیز است که به سادگی قابل اندازه‌گیری و کمّی کردن نیست.
  - نظریه علّیت وینر: درصورتی که افزودن اطلاعات گذشته و حال (سیستم/سیگنال) X به اطلاعات گذشته و حال (سیستم/سیگنال) Y، دقت پیش‌بینی وضعیت آینده (سیستم/سیگنال) Y را بهبود دهد، X علت Y می‌باشد.
  - در مفهوم علّیت وینر، همیشه علّت باید از نظر زمانی مقدم بر معلول باشد.
  - در واقع ارتباطی که به صورت لحظه‌ای و بدون تأخیر رخ می‌دهد نشانگر یک اندرکنش (وابستگی) متقابل بدون جهت است و نه یک ارتباط علّی.



تخمینگرهای ارتباطات EEG

- تخمینگرهای ارتباطات مؤثر:
  - شاخص علّیت گرنجر (Granger Causality Index)
  - تابع تبدیل جهت‌دار (Directed Transfer Function)
  - همدوسی جهت‌دار جزئی (Partial Directed Coherence)
  - شاخص تأخیر فاز جهت‌دار (Directed Phase Lag Index)
  - آنتروپی انتقال (Transfer Entropy)
  - آنتروپی انتقال جزئی (Partial Transfer Entropy)

ضریب همبستگی پیرسون

- ضریب همبستگی پیرسون (Pearson Correlation Coefficient) بین دو کانال X و Y:

$$\rho_{X,Y}=\frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}=\frac{\sum_{n=1}^{N}(x[n]-\bar{x})(y[n]-\bar{y})}{\sqrt{\sum_{n=1}^{N}(x[n]-\bar{x})^2}\sqrt{\sum_{n=1}^{N}(y[n]-\bar{y})^2}}$$

- این معیار وابستگی خطی بین کانال‌های X و Y را اندازه‌گیری می‌کند.

- قدرمطلق ضریب همبستگی عددی کوچکتر یا مساوی ۱ است. عدد صفر نشان‌دهنده این است که ارتباط خطی بین دو متغیر وجود ندارد و عدد یک بیانگر این است که دو متغیر کاملا به هم وابسته خطی‌اند.

- اشکال عمده موجود در معیار ضریب همبستگی این است که اثر بعد زمان در این معیار دیده نمی‌شود. به عبارتی اگر سری‌های زمانی را درهم بریزیم و همبستگی را محاسبه کنیم، حاصل تغییری نخواهد کرد.

تابع همبستگی متقابل

- تابع همبستگی متقابل (Cross-Correlation Function):
- با در نظرگرفتن تأخیرهای مختلف پیش از محاسبه ضریب همبستگی، تلاش می‌کند ساختار زمانی داده را نیز مدنظر قراردهد.

$$\rho_{XY}(m) = \frac{1}{\sigma_X \sigma_Y} \mathrm{E}\left[\left(X_n - \mu_X \right) \left(Y_{n+m} - \mu_Y \right) \right]$$

- در همبستگی متقابل این مساله با درنظرگرفتن مقادیر تأخیر مختلف به نوعی پوشش داده می‌شود. در کاربردهای خاص می‌توان از تأخیر زمانی‌ای که حداکثر همبستگی متقابل را ایجاد می‌کند و بزرگی همبستگی بین دو سری زمانی، درباره شار اطلاعات بین دو ناحیه اظهار نظرهایی نمود.
- به طور مشخص این معیار در سناریوهایی کاربرد دارد که تعامل عمده یک جهته‌ای بین دو ناحیه وجود داشته باشد که عمده تأثیر خود را با اختلاف زمانی مشخص نشان می‌دهد.

کوهرنسی و همدوسی

- طیف قدرت متقابل سیگنال‌های $X(t)$ و $Y(t)$:

$$S_{XY}(f) = E\left[\widetilde{X}(f).\widetilde{Y}^*(f)\right]$$

- $\tilde{X}(f)$ و $\tilde{Y}(f)$ به ترتیب تبدیل فوریه سیگنال‌های $X(t)$ و $Y(t)$ هستند.
- می‌توان E را با میانگین‌گیری روی تعداد به قدر کافی زیاد از سگمنت‌های داده تخمین زد.

- کوهرنسی (coherency) یا طیف قدرت متقابل نرمالیزه شده کانال‌های X و Y:

$$C_{XY}(f) = \frac{S_{XY}(f)}{\sqrt{S_{XX}(f).S_{YY}(f)}}$$

- یک تخمین‌گر مختلط از ارتباط خطی کانال‌های مذکور در فرکانس f

کوهرنسی و همدوسی

- همدوسی (coherence) بین دو کانال X و Y:

$$Coh_{XY}(f) = \left|C_{XY}(f)\right|^2 \qquad 0 \le Coh_{XY}(f) \le 1$$

- تخمین ارتباط خطی بین کانال‌های X و Y در فرکانس f

همدوسی جزئی

- معیار همدوسی جزئی (Partial Coherence) برای چند متغیره کردن معیار همدوسی به منظور حذف ارتباطات غیر مستقیم ارائه شده است.
- مثال: یک سیستم سه متغیره $X_1(t)$، $X_2(t)$ و $X_3(t)$
- طیف توان متقابل جزئی $S_{12|3}(f)$:

$$S_{12|3}(f)=S_{12}(f)-S_{13}(f).S_{33}^{-1}(f).S_{23}(f)$$

- تابع همدوسی جزئی:

$$PCoh_{12}(f)=\frac{\left|S_{12|3}(f)\right|^2}{\left|S_{11|3}(f)\right|.\left|S_{22|3}(f)\right|}$$

- $PCoh_{12}(f)$ نشان‌گر کسری از $Coh_{12}(f)$ است که با X_3 به اشتراک گذاشته نشده است.
- امکان تعمیم به سیستم با تعداد متغیر بیشتر

مقدار قفل‌شدگی فاز

- مقدار قفل‌شدگی فاز (Phase Locked Value)
- محاسبه به صورت تصویر کردن اختلاف فاز لحظه‌ای دو سیگنال بر روی دایره واحد و محاسبه میانگین طول بردار برآیند
- نیاز به فاز لحظه‌ای سیگنال‌ها:
  - استفاده از تبدیل هیلبرت

$$\phi_i(t) = arctan\frac{\widetilde{X}_\imath(t)}{X_i(t)}$$

$$\widetilde{X}_\imath(t) = \frac{1}{\pi}p.v.\int_{-\infty}^{+\infty}\frac{X_i(\tau)}{t-\tau}d\tau$$

- $p.v.$ به معنی مقدار اصلی کوشی است.

مقدار قفل‌شدگی فاز

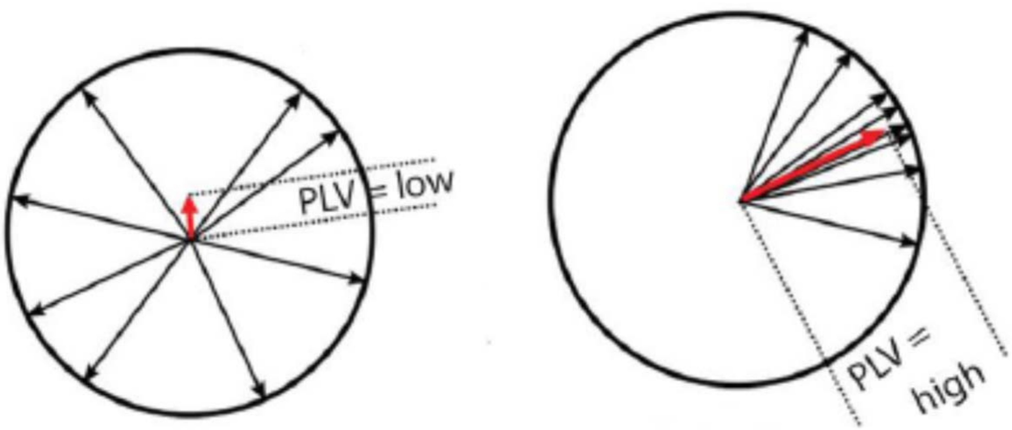
○ محاسبه اختلاف فاز دو سیگنال:

$$\Phi_{X,Y}(t) = \Phi_X(t) - \Phi_Y(t)$$

○ محاسبه PLV:

$$PLV_{X,Y} = \left| \frac{1}{N} \sum_{t=1}^{N} e^{j\Phi_{X,Y}(t)} \right|$$

○ PLV در صورت قفل‌شدگی کامل فاز دو سیستم برابر با ۱ و در حالت توزیع تصادفی اختلاف فاز بر روی دایره واحد به صفر میل خواهد کرد (غیرسنکرون بودن دو سیگنال).



شاخص تاخیر فاز

- شاخص تاخیر فاز (Phase Lag Index)
  - معیاری برای عدم تقارن توزیع اختلاف فاز دو سیگنال
  - PLI مقداری بین ۰ و ۱ دارد که هر چه‌قدر این مقدار به یک نزدیک‌تر باشد یعنی دو سیگنال هم‌فازتر بوده‌اند.
- محاسبه PLI:

$$PLI_{X,Y} = \left| \frac{1}{N} \sum_{t=1}^{N} sign(\Phi_{X,Y}(t)) \right|$$

اطلاعات متقابل

- اطلاعات متقابل (Mutual Information)
  - معیاری مبتنی بر تئوری اطلاعات
- آنتروپی شنون متغیر تصادفی گسسته X با تابع جرم احتمال $\mathrm{p}(x)$:

$$H(X) = -\sum_{x} p(x)\log(p(x))$$

  - همواره مقداری مثبت دارد.
  - نشان‌دهنده میزان اطلاعات موجود در متغیر تصادفی X
  - اگر مبنای لگاریتم ۲ باشد، $H(X)$ میانگین تعداد بیت‌های لازم برای کد کردن متغیر تصادفی گسسته X را نشان می‌دهد و واحد آن bits است.
  - برای متغیرهای پیوسته می‌توان با استفاده از گسسته‌سازی مقادیر X، تخمینی از $p(x)$ را با استفاده از روش‌های مبتنی بر هیستوگرام محاسبه کرد.

اطلاعات متقابل

- اطلاعات متقابل (Mutual Information)
  - معیاری مبتنی بر تئوری اطلاعات
- آنتروپی شنون متغیر تصادفی گسسته X با تابع جرم احتمال $\mathrm{p}(x)$:

$$H(X) = -\sum_{x} p(x)\log(p(x))$$

- می‌توان $-\log(p(x))$ را به عنوان نایقینی یا غیرمنتظره‌بودن مشاهده x از متغیر تصادفی X با تابع جرم احتمال $p(x)$ تعبیر کرد. درنتیجه آنتروپی شنون را می‌توان امید ریاضی نایقینی (Uncertainty) متغیر تصادفی X دانست.
- بنابراین توزیع یکنواخت که همه رخدادهای آن احتمال یکسان دارند بیشترین آنتروپی و یک توزیع احتمال تک مقداره کمترین آنتروپی را خواهند داشت.

اطلاعات متقابل

- دیورژانس اطلاعات Kullback-Leibler (KLD)

$$D_{p\|q}(X)=\sum_{x} p(x)\log\left(\frac{p(x)}{q(x)}\right)$$

- KLD معیار فاصله‌ای غیرمتریک بین تابع جرم احتمال $p(x)$ و تابع جرم احتمال $q(x)$ است.
- KLD همواره نامنفی است و تنها در صورتی که $p(x) = q(x)$ باشد برابر صفر است.
- KLD نامتقارن است و نامساوی مثلث را هم برآورده نمی‌کند.
- KLD آنتروپی نسبی (Rational Entropy) هم نامیده می‌شود. KLD اطلاعات موجود برای تمایز قائل‌شدن بین تابع جرم احتمال $p(x)$ از تابع جرم احتمال $q(x)$ را اندازه‌گیری می‌کند.

اطلاعات متقابل

- اطلاعات متقابل (مشترک) دو متغیر تصادفی گسسته X و Y:

$$MI(X,Y)=D(p(x,y)\,\|\,p(x).p(y))=\sum_{x,y}p(x,y).\log\left(\frac{p(x,y)}{p(x).p(y)}\right)$$

- میزان اطلاعاتی را که با مشاهده هر یک از متغیرهای تصادفی X یا Y نسبت به دیگری به دست می‌آید، اندازه‌گیری می‌کند.
- $MI(X,Y)$ همواره نامنفی و کران‌دار است.
- $MI(X,Y)$ برابر صفر است اگر و تنها اگر متغیرهای تصادفی X و Y از هم مستقل باشند.
- $MI(X,Y)$ می‌تواند برآیند وابستگی‌های خطی و غیرخطی بین متغیرهای تصادفی X و Y را اندازه‌گیری کند.
- مناسب برای بررسی کلیه ارتباطات کارکردی خطی و غیرخطی دو متغیرتصادفی X و Y (ارتباطات کارکردی مغزی)

اطلاعات متقابل

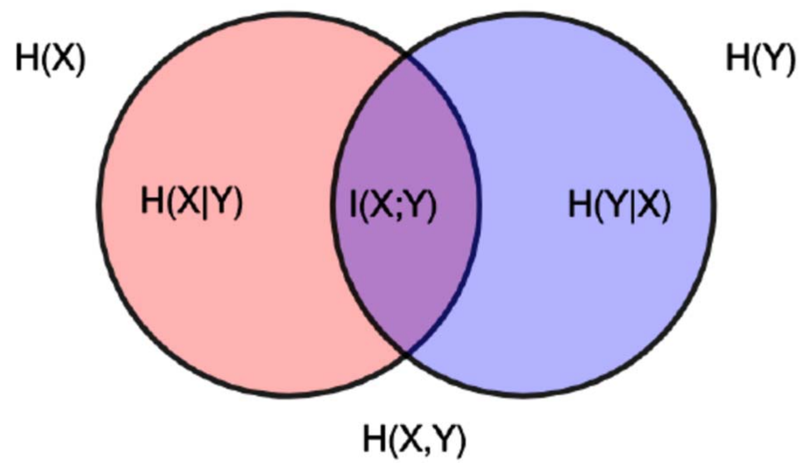
- بازنویسی $MI(X,Y)$ با استفاده از مفهوم آنتروپی شنون:

$$
\begin{aligned}
MI(X,Y) &= H(X) + H(Y) - H(X,Y) \\
&= H(X) - H(X \mid Y) \\
&= H(Y) - H(Y \mid X)
\end{aligned}
$$

$$H(X,Y) = -\sum_{x,y} p(x,y).\log(p(x,y))$$

$$H(X \mid Y) = -\sum_{x,y} p(x,y).\log(p(x \mid y))$$

$$H(Y \mid X) = -\sum_{x,y} p(x,y).\log(p(y \mid x))$$



اطلاعات متقابل جزئی

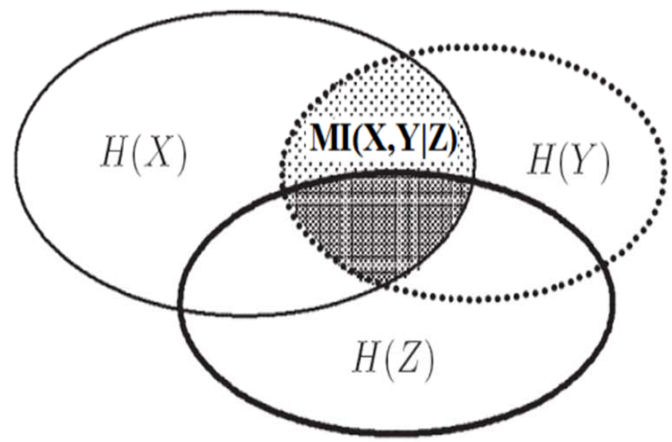
- اطلاعات متقابل جزئی (Partial Mutual Information)

- تعمیم معیار اطلاعات متقابل برای سیستم‌های چند متغیره (بیش از دو متغیر) به منظور حذف اثر وابستگی های غیرمستقیم

- $MI(X,Y|Z)$: اطلاعات متقابل جزئی (شرطی) متغیرهای تصادفی X و Y مشروط به مشاهده متغیر تصادفی Z

$$PMI(X,Y) = MI(X,Y|Z) = \sum_{x,y,z} p(x,y,z)\log\left(\frac{p(x,y|z)}{p(x|z).p(y|z)}\right)$$

$$= \sum_{x,y,z} p(x,y,z).\log\left(\frac{p(x,y,z).p(z)}{p(x,z).p(y,z)}\right)$$



$$PMI(X,Y) = MI(X,Y|Z) = H(X,Z) + H(Y,Z) - H(Z) - H(X,Y,Z)$$

علّیت گرنجر

- علّیت Granger
- ریشه این معیار: مدلسازی سری‌های زمانی توسط مدل‌های خودبازگشتی (AR)
- **یادآوری: مدل‌های پارامتری**
- مدل خودبازگشتی (Autoregressive-AR)
- $AR(p)$: $x[n] = \sum_{i=1}^{p} a_i x[n-i] + w[n]$
  - $x[n]$: سری زمانی ورودی
  - $w[n]$: نویز سفید
  - p: مرتبه مدل
  - a_iها: پارامترهای مدل (a_1 تا a_p)

علّیت گرنجر

- یادآوری: مدل‌های پارامتری
- مدل میانگین متحرک (Moving Average-MA)

- $MA(q)$: $x[n] = \sum_{i=1}^{q} b_i w[n-i] + w[n]$

   - $x[n]$: سری زمانی ورودی
   - $w[n]$: نویز سفید
   - p: مرتبه مدل
   - b_iها: پارامترهای مدل (b_1 تا b_q)

علّیت گرنجر

- **یادآوری: مدل‌های پارامتری**
- مدل خودبازگشتی-میانگین متحرک (ARMA)

- $ARMA(p,q)$: $x[n] = \sum_{i=1}^{p} a_i x[n-i] + \sum_{i=1}^{q} b_i w[n-i] + w[n]$
  - $x[n]$: سری زمانی ورودی
  - $w[n]$: نویز سفید
  - p و q: مرتبه مدل
  - b_iها و a_iها: پارامترهای مدل

علّیت گرنجر

- به طور مشخص، کیفیت یک مدل پارامتری توسط واریانس باقی‌مانده‌های آن مشخص می‌شود.

- با این پیش فرض، معیار علیت Granger به صورت لگاریتم نسبت باقی مانده‌های دو مدل AR مشخص می‌گردد که یکی از آنها سعی می‌کند مقدار سری زمانی x را بر حسب گذشته خود آن پیش‌بینی کند و دیگری همین کار را با استفاده از گذشته y در کنار گذشته x انجام دهد.

- کاهش قابل توجه واریانس در مدل دوم نسبت به مدل اول نشان می‌دهد مقادیر گذشته y توانسته‌اند به پیش‌بینی x کمک کنند و در نتیجه x تابعیت علّی از y دارد.

علّیت گرنجر

$$x[n] = \sum_{m=1}^{p} a_{x,m} x[n-m] + e_x[n]$$

$$x[n] = \sum_{m=1}^{p} a_{xx,m} x[n-m] + \sum_{m=1}^{p} a_{xy,m} y[n-m] + e_{xy}[n]$$

$$GC_{y\to x} = \ln\left(\frac{Cov(e_x)}{Cov(e_{xy})}\right)$$

- تابعیت علّی y از x نیز به صورت مشابه قابل تعریف است.

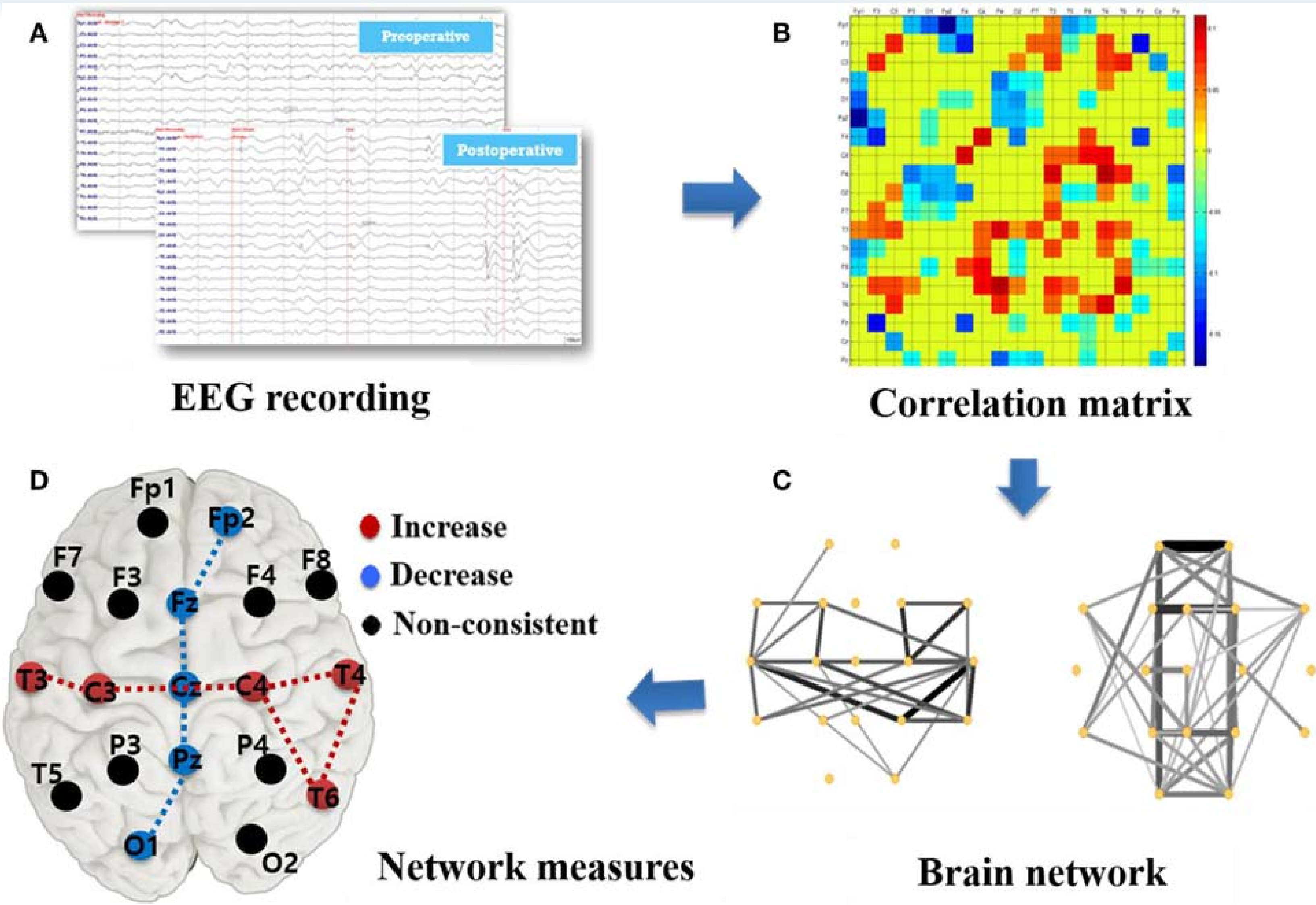
ویژگی‌های مبتنی بر ارتباطات کارکردی و علّی

- محاسبه معیار ارتباطات مغزی بین هر دو کانال ثبت EEG
  - محاسبه معیار می‌تواند در باندهای مختلف فرکانسی یا بازه‌های مختلف زمانی صورت پذیرد.

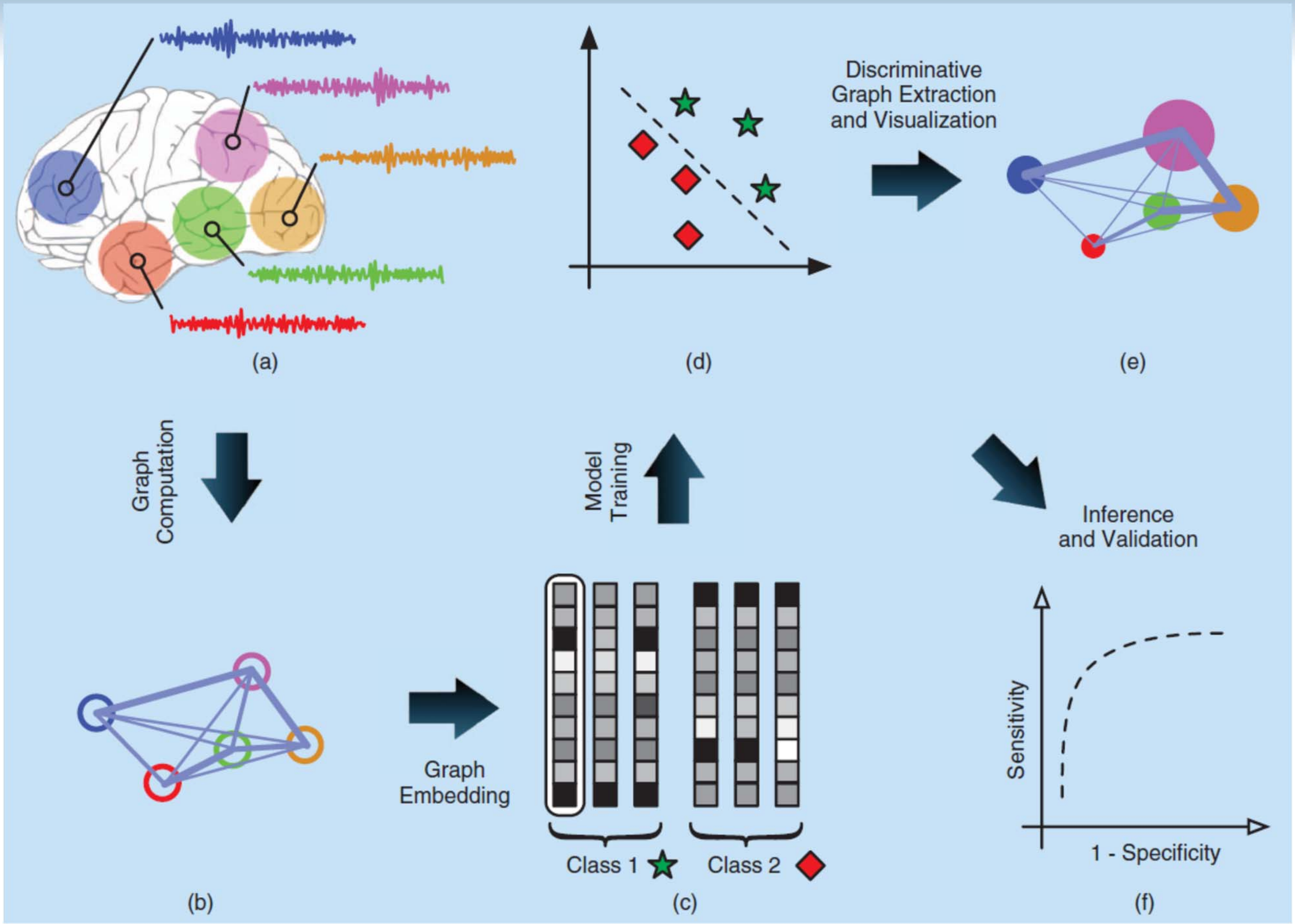
- نحوه استفاده از اطلاعات به دست آمده از ارتباطات مغزی:
  - استفاده از مقادیر ارتباطات به صورت خام و تشکیل بردار ویژگی‌ها
  - تبدیل ماتریس ارتباطات مغزی به گراف متناظر
    - گراف وزن‌دار
    - گراف باینری: مثلاً با استفاده از آستانه‌گذاری یا با اضافه کردن یال‌ها با وزن بیشتر تا زمانی که گراف همبند شود.

- استفاده از گراف ارتباطات مغزی:
  - استخراج ویژگی‌های گرافی
  - استفاده از کرنل‌های گرافی

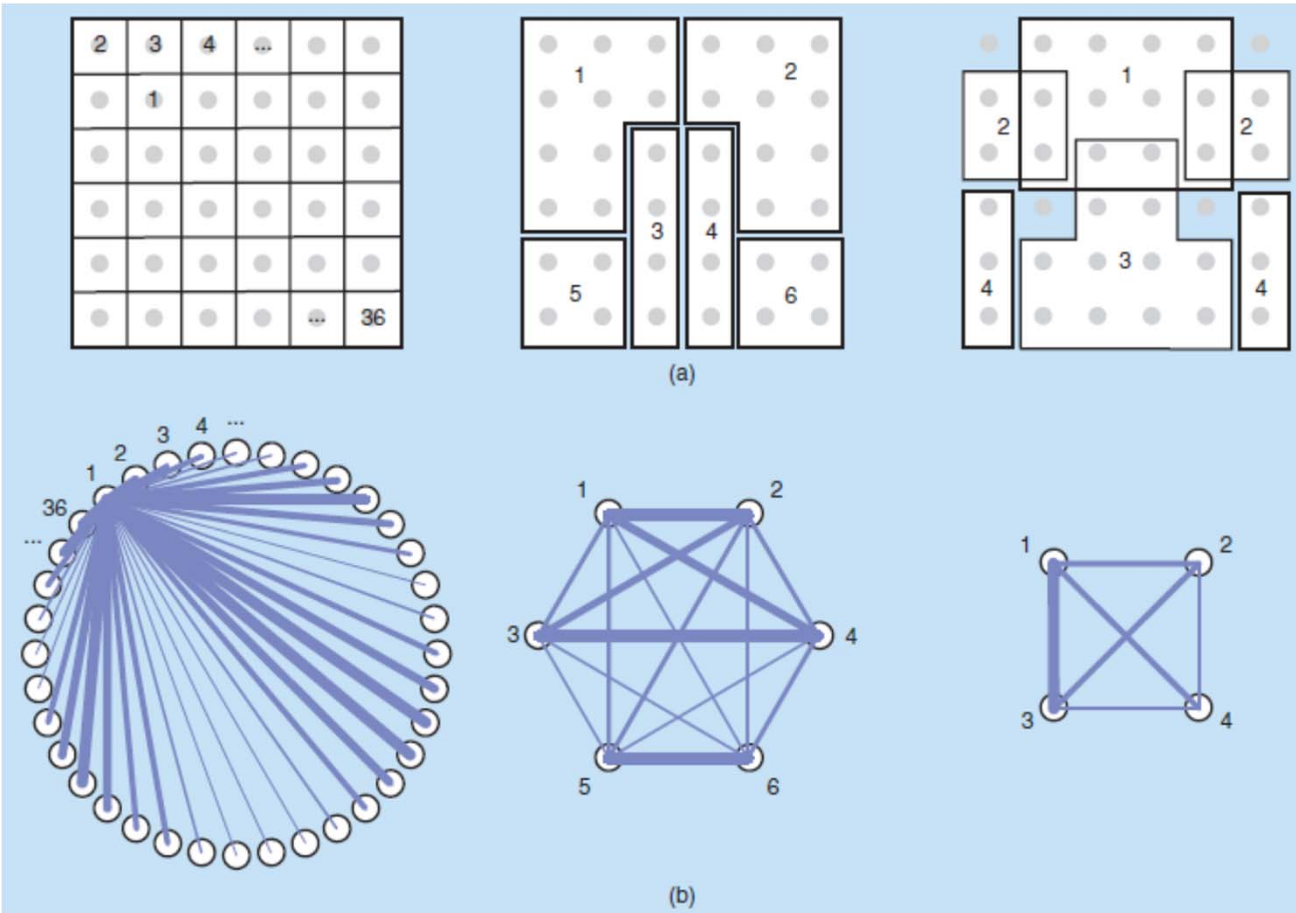
ساخت گراف ارتباطات مغزی



گراف ارتباطات مغزی

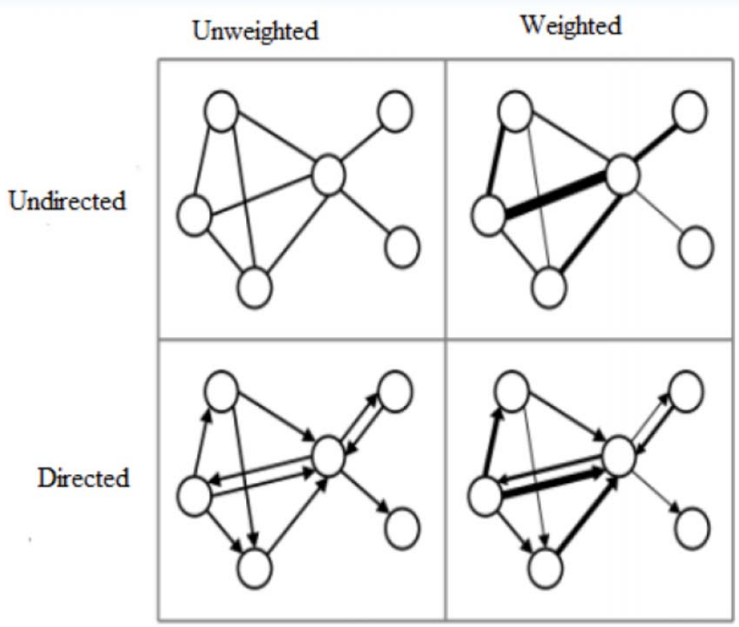


تعیین رأس‌های گراف مغزی



استخراج ویژگی‌های گراف

- انواع گراف‌ها:
  - وزن‌دار، بدون وزن (باینری)
  - جهتدار، بدون جهت



- ویژگی‌های گرافی:
  - سراسری: بیان‌کننده ویژگی‌های کلی یک گراف
  - محلی: ویژگی‌های گره‌های گراف (یک بردار به بعد تعداد گره‌های گراف)

  - کمّی‌سازی تفکیک کارکردی
  - کمّی‌سازی یکپارچگی کارکردی
  - تعیین اهمیت مناطق مختلف مغز

ویژگی‌های سنجش تفکیک کارکردی

- نظریه تفکیک کارکردی:
  - برای انجام عملیات و فرایندهای مشخص، نواحی جداگانه‌ای از مغز فعال می‌شوند.
  - یک تخصیص کارکردی وجود دارد که هر ناحیه مغز تنها در حین فعالیت‌های خاصی که به آن تخصیص یافته است، فعال می‌شود.

- ویژگی‌های سنجش تفکیک‌پذیری:
  - تعیین حضور گروه‌هایی که به عنوان خوشه یا ماژول شناخته می‌شوند در داخل شبکه گرافی

- حضور خوشه‌ها در شبکه‌های ساختاری نشان دهنده این است که پتانسیل تفکیک کارکردی در این شبکه‌ها وجود دارد، در حالی‌که وجود خوشه‌ها در شبکه‌های کارکردی بیان‌گر پردازش‌های عصبی جدا از هم است.

- متداول‌ترین ویژگی‌ها:
  - ضریب خوشه‌بندی (Clustering coefficient)
  - بازده محلی (Local efficiency)

ضریب خوشه‌بندی

- ضریب خوشه‌بندی: درجه تمایل گره‌ها به ایجاد یک خوشه

- در گراف‌های بدون جهت باینری، ضریب خوشه‌بندی گره i به صورت نسبت تعداد مثلث‌هایی که در اطراف گره i وجود دارد به تمام یال‌های ممکن میان گره‌های همسایه گره i تعریف می‌شود:

$$c_i = \frac{2t_i}{k_i(k_i - 1)}$$

  - t_i تعداد مثلث‌های موجود در اطراف گره i را نشان می‌دهد: $t_i = \frac{1}{2} \sum_{i \to j \to h} a_{ij}\, a_{ih} a_{jh}$

  - k_i درجه گره i است: $k_i = \sum_{i \neq j} a_{ij}$

  - a_{ij} بیان‌گر وجود یا عدم وجود یال بین دو گره i و j در گراف است.

- ضریب خوشه‌بندی کلی شبکه، میانگین ضریب خوشه‌بندی تمام گره‌ها است:

$$C = \frac{1}{N} \sum_{i=1}^{N} c_i = \frac{1}{N} \sum_{i=1}^{N} \frac{2t_i}{k_i(k_i - 1)}$$

بازده محلی

- بازده محلی گره i:

$$E_{loc,i} = \sum_{i \to j \to h} \frac{a_{ij} a_{ih} [d_{jh}(N_i)]^{-1}}{k_i(k_i - 1)}$$

- $d_{jh}(N_i)$ طول کوتاهترین مسیر بین گره j و h که در همسایگی گره i قرار دارند.

ویژگی‌های سنجش یکپارچگی کارکردی

- نظریه یکپارچگی کارکردی:
  - نواحی مجزا و تخصیص یافته مغزی در حین انجام عملیات و فرآیندهای مشخص با یکدیگر ارتباط و هماهنگی دارند به نحوی که در نتیجه این ارتباطات، همگی این نواحی در قالب سیستمی یکپارچه و هماهنگ فعالیت می‌کنند.
  - منظور از یکپارچگی کارکردی در مغز توانایی ترکیب سریع اطلاعات از نواحی مجزای مغز است.
  - ویژگی‌های سنجش یکپارچگی معمولا مبتنی بر مسیر تعریف می‌شوند، مخصوصا طول کوتاهترین مسیر که دو گره مختلف را به هم وصل می‌کند. به طور کلی مسیرهای کوتاه‌تر به صورت مطلوب‌تری اطلاعات را انتقال می‌دهند.

- متداول‌ترین ویژگی‌ها:
  - طول مسیر مشخصه
  - بازده عمومی
  - شعاع و قطر شبکه

طول مسیر مشخصه

- طول مسیر مشخصه:
  - نشان‌گر میزان اتحاد و یکپارچگی شبکه است و به وسیله آن مشخص می‌شود که اطلاعات درون شبکه تا چه اندازه می‌توانند به راحتی و باسرعت جابه‌جا شوند.
- طول مسیر (یا کوتاه‌ترین مسیر) یا فاصله بین دو گره i و j (d_{ij}):
  - کمترین تعداد یال‌های متصل‌کننده گره i به j
- طول مسیر مشخصه گراف (L):
  - میانگین فاصله میان همه زوج رأس‌ها

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} d_{i,j}$$

  - N تعداد رأس‌های گراف

بازده عمومی

- بازده عمومی:
  - اندازه‌گیری درهم‌تنیدگی شبکه و کارایی کلی آن برای انتقال اطلاعات

$$E_{global} = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{i,j}}$$

شعاع و قطر شبکه

- معیار خروج از مرکز برای گره i:
  - فاصله گره i با تمام گره‌های موجود در گراف محاسبه می‌شود. بیشترین این مقادیر محاسبه شده برای گره i، خروج از مرکز گره i را نشان می‌دهد.

$$ecc(v_i) = \max_{v_j \in V, i \neq j} d(v_i, v_j)$$

- شعاع شبکه:

$$radius = \min_{v_i \in V} ecc(v_i) = \min_{v_i \in V} \max_{v_j \in V, i \neq j} d(v_i, v_j)$$

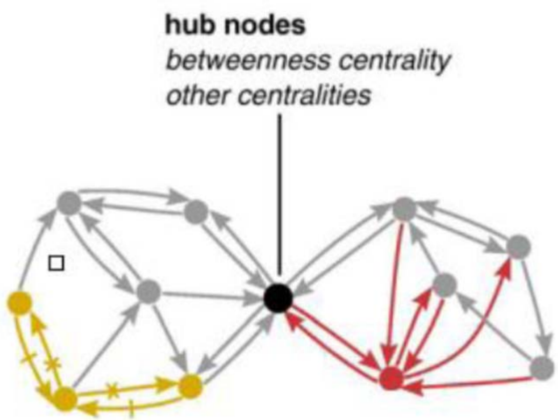
- قطر شبکه:

$$diameter = \max_{v_i \in V} ecc(v_i) = \max_{v_i \in V} \max_{v_j \in V, i \neq j} d(v_i, v_j)$$

- بازه تغییرات هر سه معیار بین صفر تا بی‌نهایت است.

معیارهای سنجش مرکزیت

- معیارهای سنجش مرکزیت (Centrality):
  - اندازه‌گیری اهمیت گره‌ها در یک شبکه مغزی و شناسایی رأس‌های مهم درون گراف
  - هرچه مرکزیت یک گره بیشتر باشد، آن گره در انتقال اطلاعات شبکه مغزی موثرتر است.
  - مناطق مهم مغز (قطب‌ها) اغلب با بسیاری از مناطق دیگر مغز ارتباط دارند، یکپارچگی کارکردی را تسهیل می‌کنند و نقش اساسی در مقاوت شبکه به تخریب‌ها دارند.
  - گره‌های مرکزی در بسیای از مسیرهای کوتاه درون یک شبکه شرکت می‌کنند و در نتیجه به عنوان یک جزء مهم در کنترل جریان اطلاعات عمل می‌کنند.



- ویژگی‌های متداول:
  - درجه
  - مابینی (Betweenness)
  - نزدیکی (Closeness)

معیارهای سنجش مرکزیت

- درجه یک گره:

$$C_d(i) = k_i = \sum_{j \neq i} a_{ij}$$

- نزدیکی:
  - برابر با معکوس میانگین فاصله طول مسیر مشخصه گره i تا همه گره‌های دیگر است.

$$C_c(i) = \frac{N-1}{\sum_{j \neq i} l_{ij}}$$

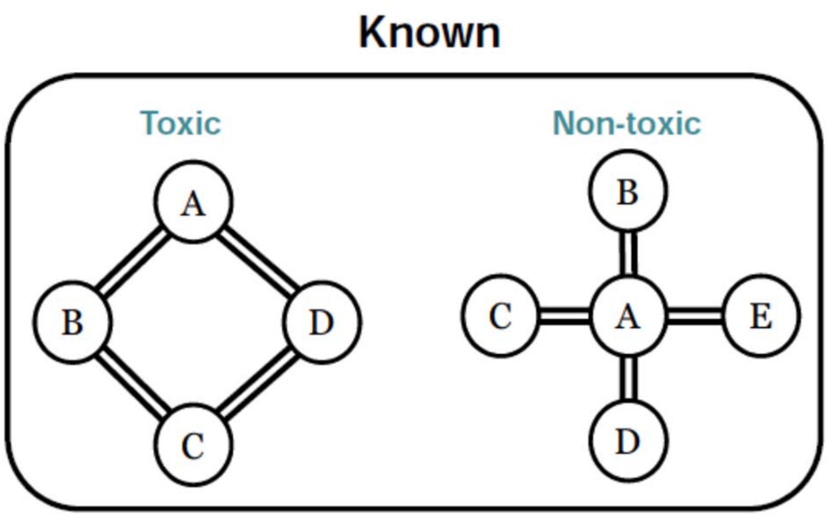
- مابینی:
  - برابر با تعداد دفعاتی است که یک گره در طول کوتاهترین مسیر بین دو گره دیگر به عنوان یک پل عمل کرده است.

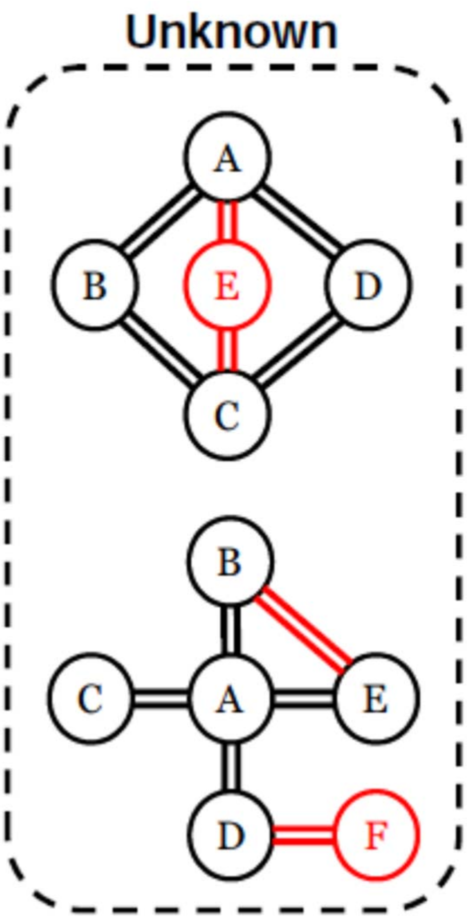
$$C_b(i) = \frac{2}{(N-1)(N-2)} \sum_{j \neq h \neq i} \frac{n_{hj}(i)}{n_{hj}}$$

  - $n_{hj}(i)$: تعداد کوتاه‌ترین مسیرها بین گره h و j که از i می‌گذرند.
  - n_{hj}: تعداد کل کوتاه‌ترین مسیرها بین گره h و j

طبقه‌بندی گراف‌ها بر اساس ساختارهای گرافی

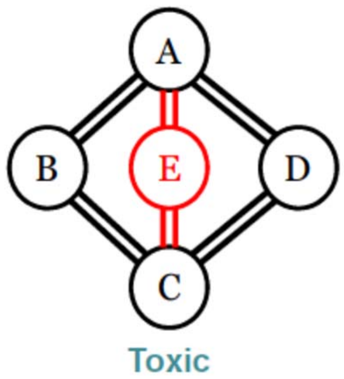
- مثال: پیش‌بینی سمّی یا غیرسمّی بودن مولکول‌ها با داشتن مجموعه‌ای از نمونه‌های سمّی و غیرسمّی با برچسب مشخص



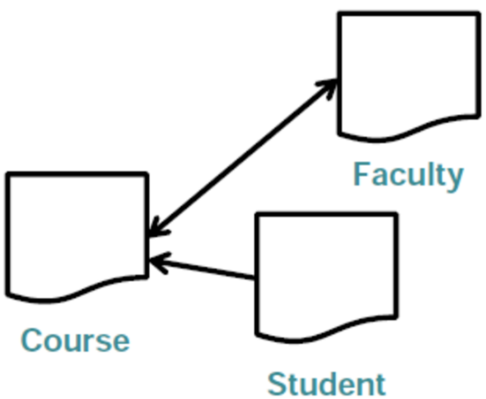


طبقه‌بندی گراف‌ها بر اساس ساختارهای گرافی

- طبقه‌بندی گرافی (Graph classification) (بین-گرافی):
  - به هر گراف کامل یک برچسب کلاس تخصیص داده می‌شود.
  - مثال: گراف‌های مولکولی



- طبقه‌بندی رأس‌ها (Vertex classification) (درون-گرافی):
  - دریک گراف واحد به هر رأس یک برچسب کلاس تخصیص داده می‌شود.
  - مثال: گراف‌های Webpage (vertex) / hyperlink (edge)



S.Hajipour, Sharif University of Technology, 2021

یادآوری: یادگیری ماشین

- طبقه‌بندی باینری:
  - داده‌های آموزشی:

$$(\boldsymbol{x}_i, y_i) \text{ for } i = 1 \dots N \qquad \boldsymbol{x}_i \in \mathbb{R}^d \qquad y_i \in \{-1, 1\}$$

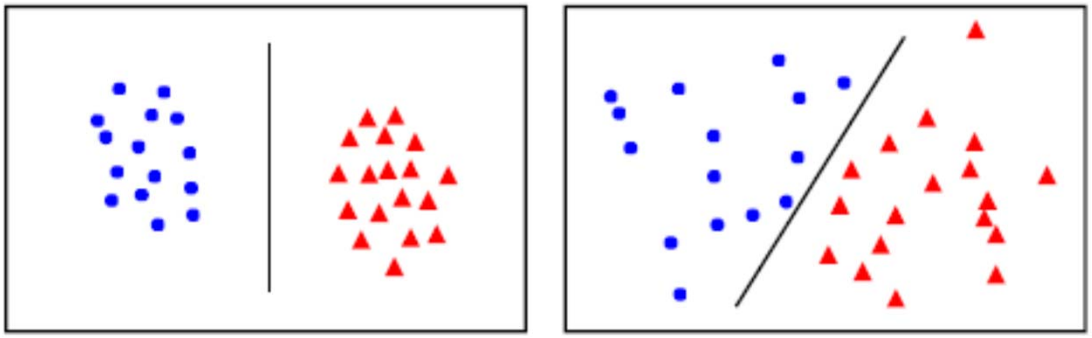
  - هدف: آموزش یک طبقه‌بند $f(x)$ به‌گونه‌ای که:

$$f(x_i)\begin{cases} \geq 0 & y_i = +1 \\ < 0 & y_i = -1 \end{cases}$$

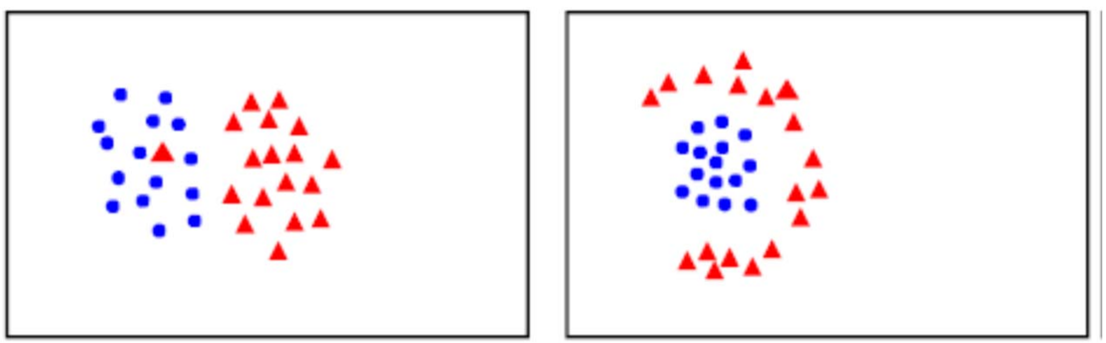
  - یعنی: $y_i f(x_i) > 0$

یادآوری: یادگیری ماشین

- خطی جدایی‌پذیر



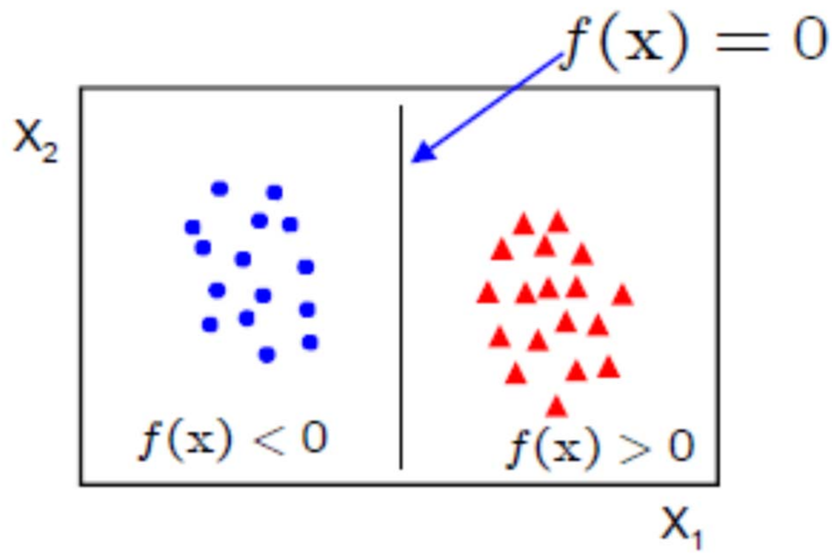
- خطی جدایی‌ناپذیر:



یادآوری: یادگیری ماشین

- طبقه‌بند خطی:

$$f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b$$



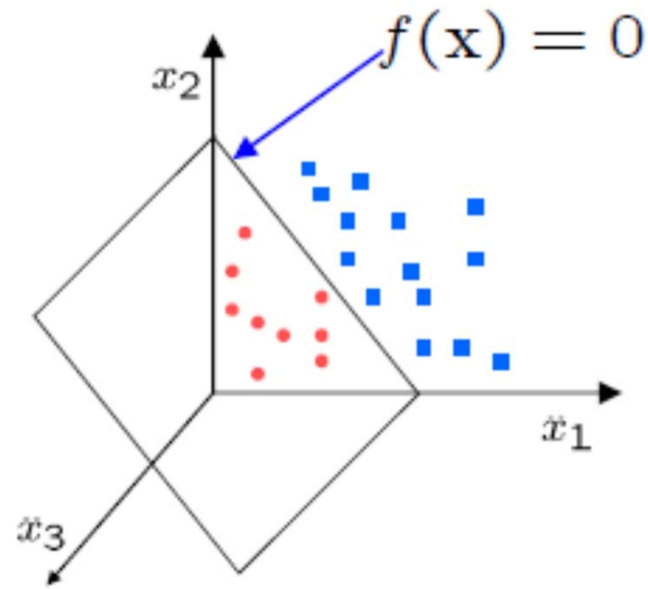
- در حالت داده‌های دوبعدی جداکننده یک خط است که $\boldsymbol{w}$ بردار نرمال خط بوده و $\boldsymbol{b}$ بایاس است.

- به بردار $\boldsymbol{w}$ بردار وزن می‌گوییم.

یادآوری: یادگیری ماشین

- طبقه‌بند خطی:

$$f(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + b$$



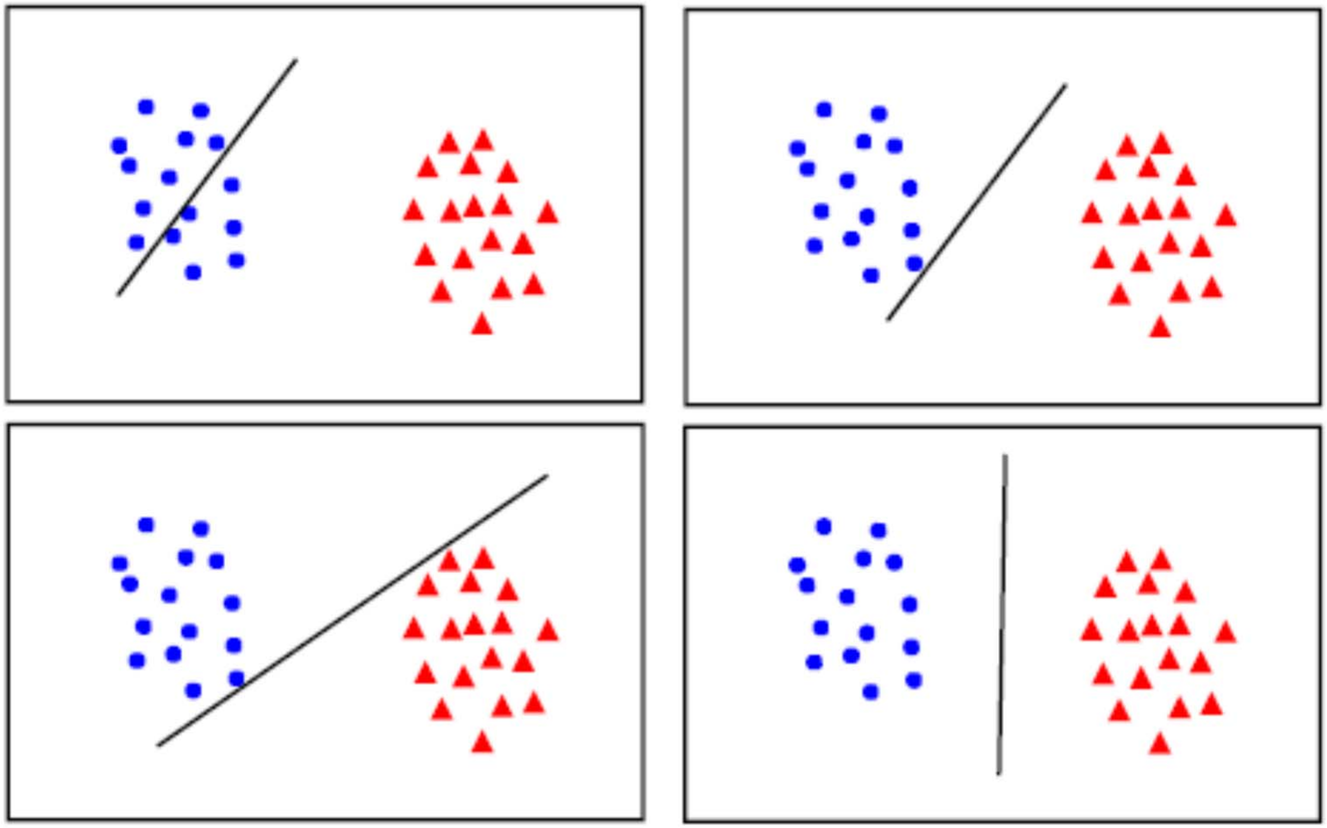
- در حالت داده‌های سه‌بعدی جداکننده یک صفحه است و برای داده‌های n بعدی یک ابرصفحه است.

- برای یک طبقه‌بند خطی، داده‌های آموزش برای آموزش $\boldsymbol{w}$ استفاده می‌شوند و بعد دور ریخته می‌شوند.

- فقط $\boldsymbol{w}$ برای طبقه‌بندی داده‌ها لازم است.

یادآوری: یادگیری ماشین

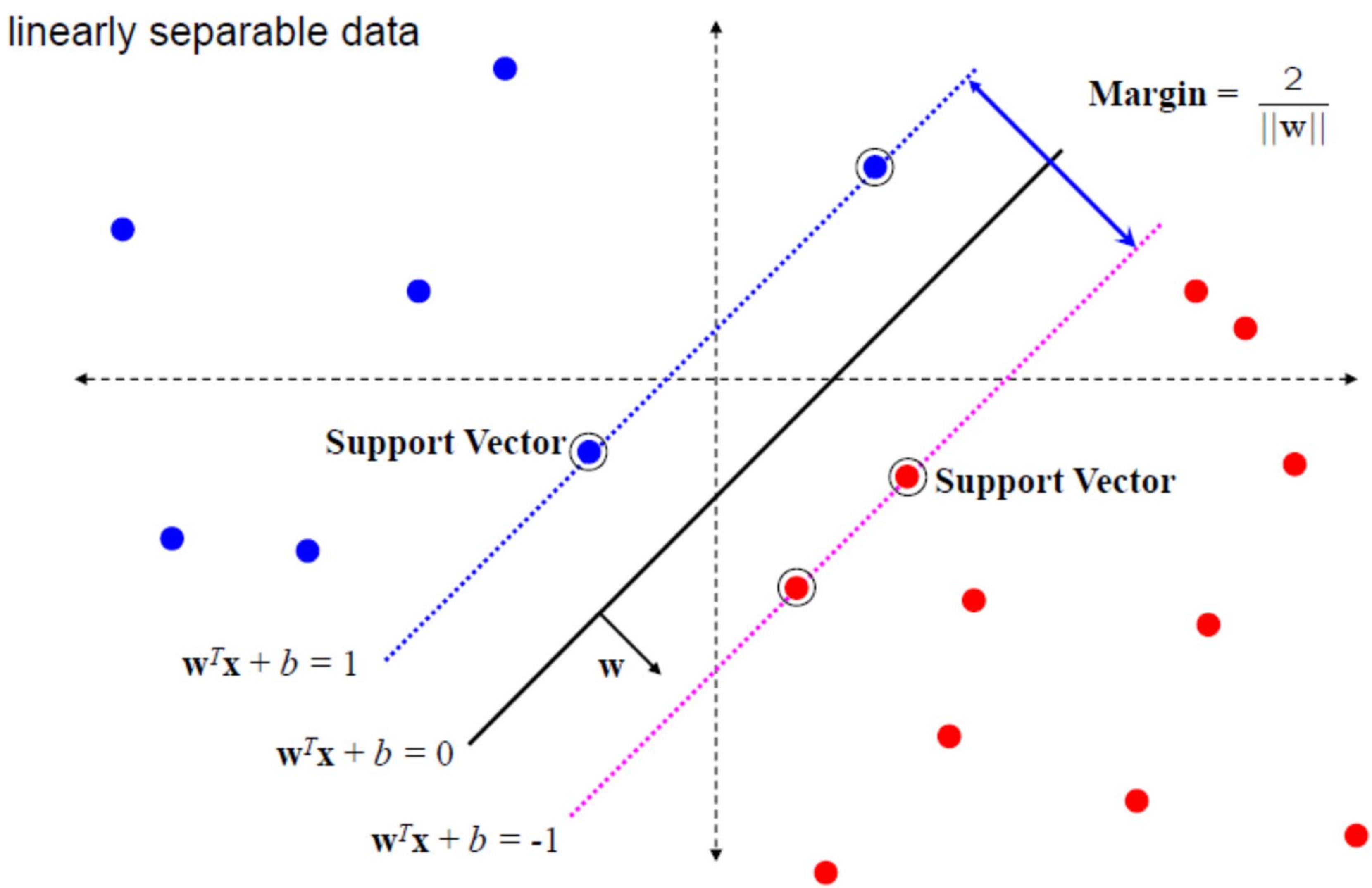
- بهترین مقدار برای $\boldsymbol{w}$ چیست؟



- پاسخ با بیشترین حاشیه (maximum margin): پایدارترین پاسخ تحت اغتشاشات ورودی

یادآوری: یادگیری ماشین

- ماشین بردار پشتیبان (Support Vector Machine) (SVM)



یادآوری: یادگیری ماشین

- آموزش طبقه‌بند SVM را می‌توان به صورت یک مسئله بهینه‌سازی فرموله کرد:

$$\max_{\mathbf{w}} \frac{2}{||\mathbf{w}||} \text{ subject to } \mathbf{w}^\top \mathbf{x}_i + b \begin{array}{ll} \geq 1 & \text{if } y_i = +1 \\ \leq -1 & \text{if } y_i = -1 \end{array} \text{ for } i = 1 \ldots N$$

- یا به طور معادل:

$$\min_{\mathbf{w}} ||\mathbf{w}||^2 \text{ subject to } y_i \left(\mathbf{w}^\top \mathbf{x}_i + b \right) \geq 1 \text{ for } i = 1 \ldots N$$

- یک مسئله بهینه‌سازی درجه دوم با قیدهای خطی: یک مینیموم یکتا دارد.

یادآوری: یادگیری ماشین

- مسئله اولیه (Primal):
  - یک طبقه‌بند خطی به صورت:

$$f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b$$

  - که منجر به حل یک مسئله بهینه‌سازی روی $\boldsymbol{w}$ می‌شود:

$$\min_{\mathbf{w}\in\mathbb{R}^d} ||\mathbf{w}||^2 + C\sum_{i}^{N} \max\left(0, 1 - y_i f(\mathbf{x}_i)\right)$$

- مسئله دوگان (Dual):
  - یک طبقه‌بندی خطی به صورت:

$$f(\mathbf{x}) = \sum_{i}^{N} \alpha_i y_i (\mathbf{x}_i{}^\top \mathbf{x}) + b$$

  - که منجر به حل یک مسئله بهینه‌سازی روی مقادیر α_i می‌شود.

یادآوری: یادگیری ماشین

- مسئله اولیه و مسئله دوگان:
  - تعداد داده‌های آموزش: N
  - بعد بردار ویژگی‌ها: d

Primal version of classifier:

$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$$

Dual version of classifier:

$$f(\mathbf{x}) = \sum_i^N \alpha_i y_i (\mathbf{x}_i{}^\top \mathbf{x}) + b$$

- مسئله اولیه و مسئله دوگان به ترتیب به آموزش d و N پارامتر نیاز دارند.
- فرم دوگان فقط شامل ترم‌هایی به صورت $(x_j^T x_k)$ است.

- مسئله اولیه و مسئله دوگان:

Primal problem: for $\mathbf{w} \in \mathbb{R}^d$

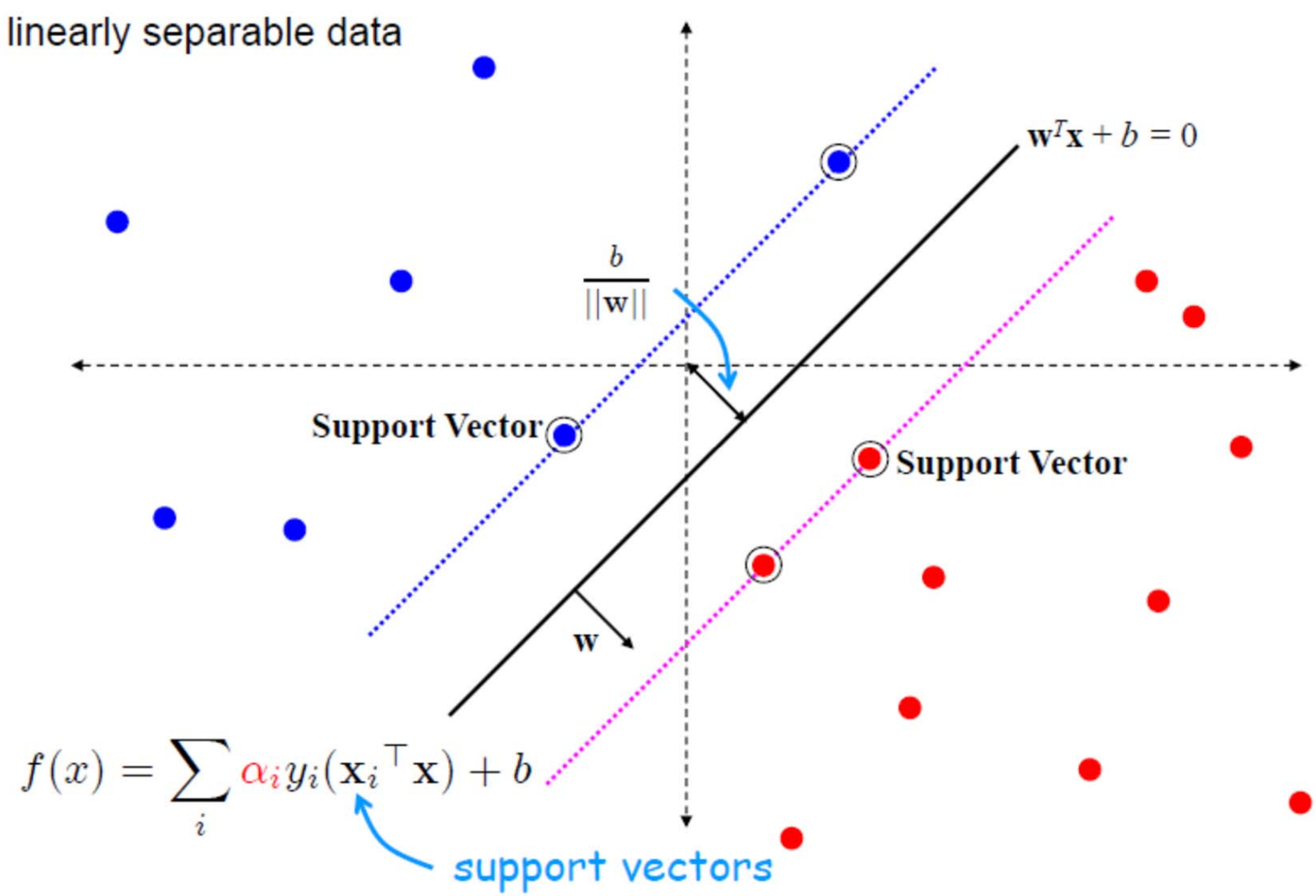
$$\min_{\mathbf{w}\in\mathbb{R}^d} ||\mathbf{w}||^2 + C\sum_{i}^{N} \max\left(0, 1 - y_i f(\mathbf{x}_i)\right)$$

Dual problem: for $\boldsymbol{\alpha} \in \mathbb{R}^N$ (stated without proof):

$$\max_{\alpha_i \geq 0} \sum_i \alpha_i - \frac{1}{2}\sum_{jk} \alpha_j \alpha_k y_j y_k (\mathbf{x}_j{}^\top \mathbf{x}_k) \text{ subject to } 0 \leq \alpha_i \leq C \text{ for } \forall i, \text{ and } \sum_i \alpha_i y_i = 0$$

یادآوری: یادگیری ماشین

- ماشین بردار پشتیبان (Support Vector Machine) (SVM)



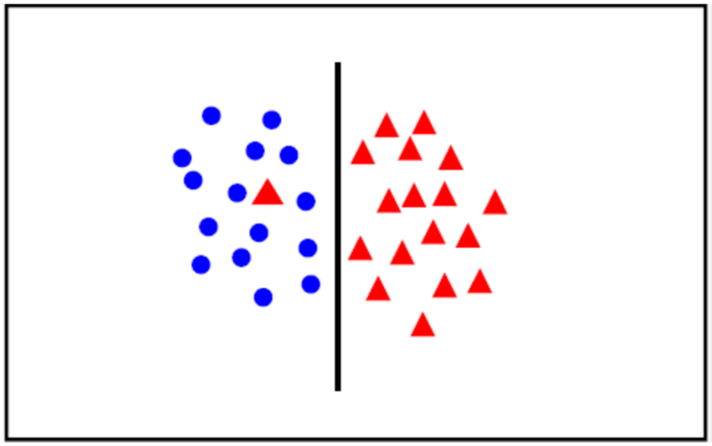
یادآوری: یادگیری ماشین

- اگر داده‌ها به صورت خطی جدایی‌پذیر نباشند:
- حالت اول: استفاده از متغیرهای کمکی

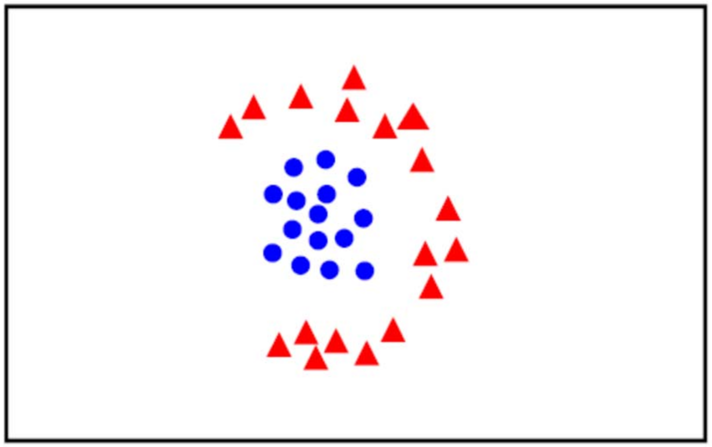
$$\min_{\mathbf{w}\in\mathbb{R}^d,\xi_i\in\mathbb{R}^+} ||\mathbf{w}||^2 + C\sum_i^N \xi_i$$

subject to

$$y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right) \geq 1 - \xi_i \text{ for } i = 1 \dots N$$

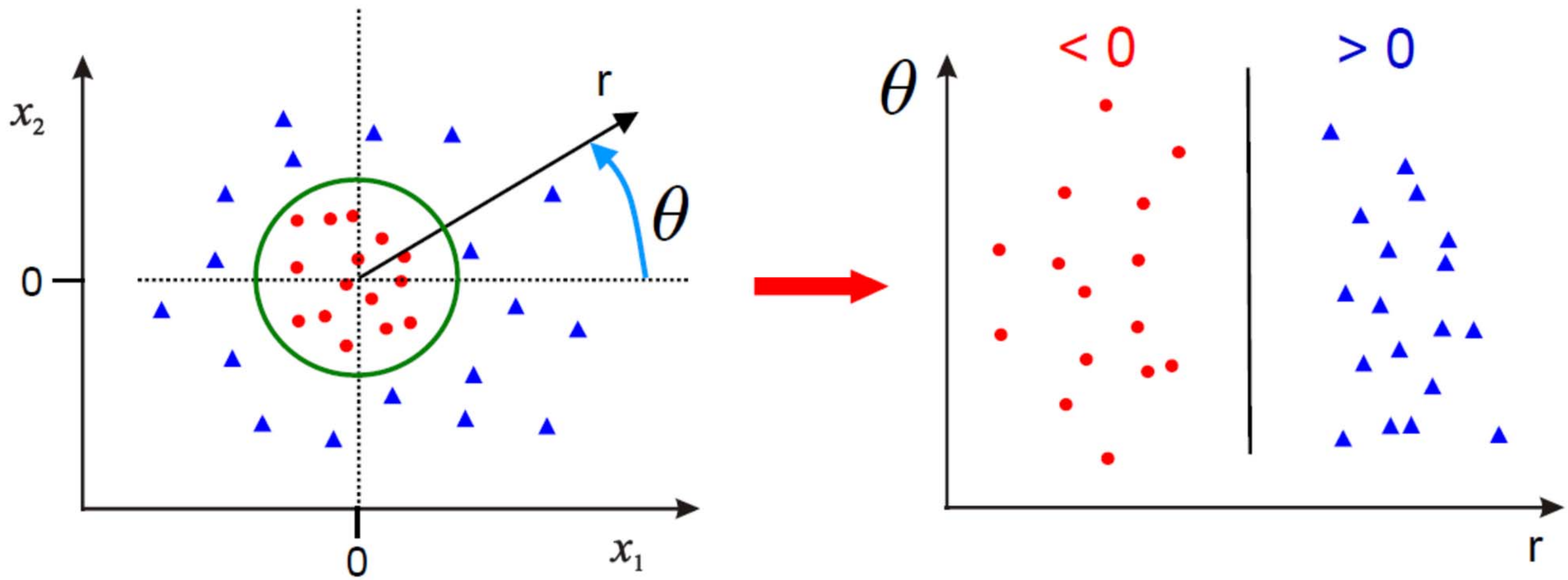


- حالت دوم: طبقه‌بندهای خطی مناسب نیستند



یادآوری: یادگیری ماشین

- راه‌حل ۱: استفاده از مختصات قطبی



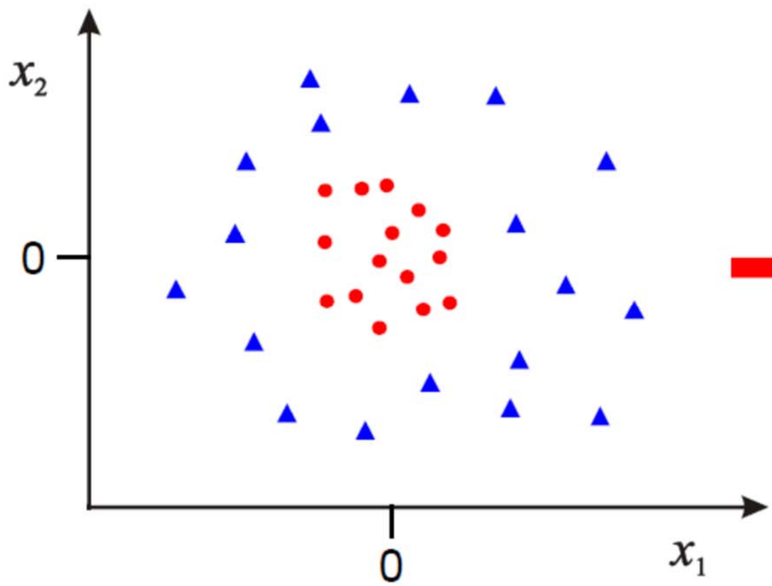
- داده‌هایی که در مختصات اولیه غیرخطی بودند، در مختصات قطبی خطی جدایی‌پذیر می‌شوند.

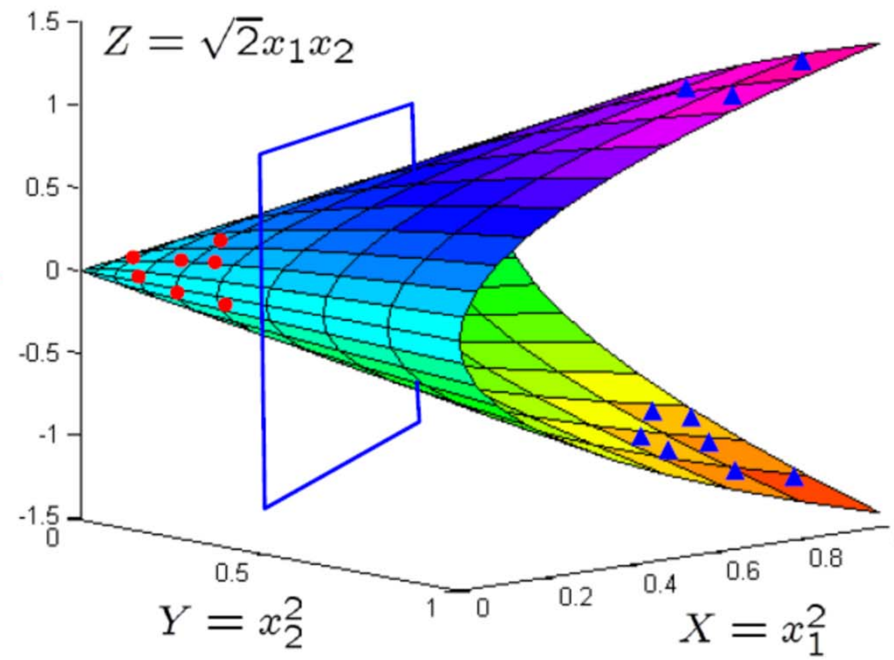
$$\Phi : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} r \\ \theta \end{pmatrix} \qquad \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

یادآوری: یادگیری ماشین

- راه‌حل ۲: نگاشت داده‌ها به فضا با بعد بیشتر

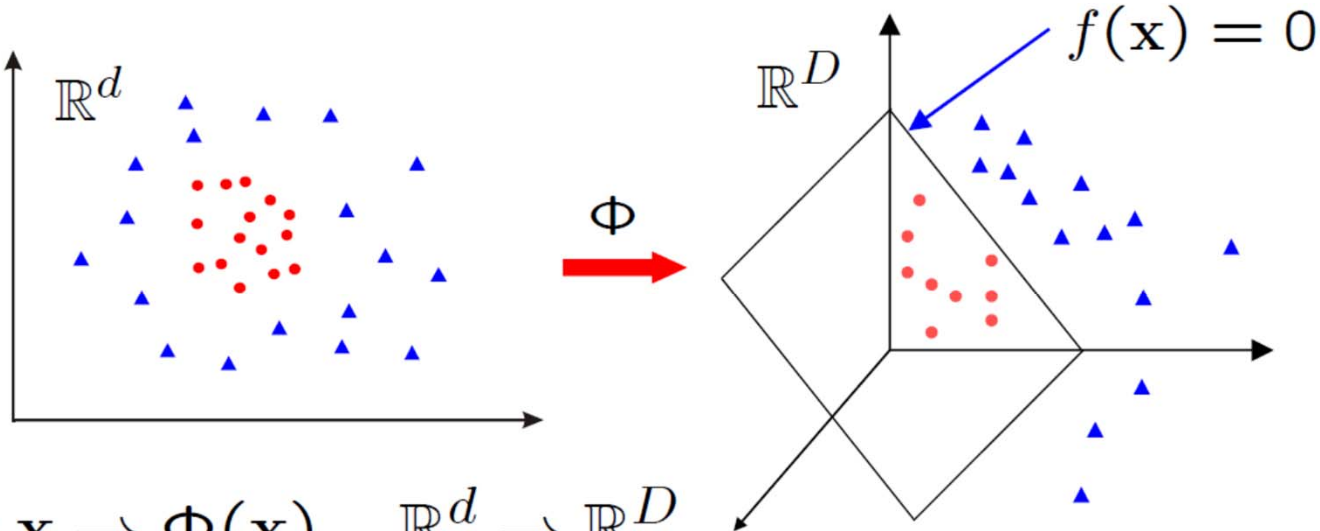
$$\Phi : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \end{pmatrix} \qquad \mathbb{R}^2 \rightarrow \mathbb{R}^3$$





یادآوری: یادگیری ماشین

- طبقه‌بند SVM در فضای ویژگی تبدیل‌یافته:



$$\Phi : \mathbf{x} \to \Phi(\mathbf{x}) \qquad \mathbb{R}^d \to \mathbb{R}^D$$

- آموزش طبقه‌بند خطی روی $\boldsymbol{w}$ در فضای $\mathbb{R}^D$:

$$f(\mathbf{x}) = \mathbf{w}^\top \Phi(\mathbf{x}) + b$$

- $\phi(x)$ یک نگاشت ویژگی است.

یادآوری: یادگیری ماشین

- طبقه‌بند اولیه در فضای ویژگی‌های تبدیل‌یافته:

Classifier, with $\mathbf{w} \in \mathbb{R}^D$:

$$f(\mathbf{x}) = \mathbf{w}^\top \Phi(\mathbf{x}) + b$$

Learning, for $\mathbf{w} \in \mathbb{R}^D$

$$\min_{\mathbf{w} \in \mathbb{R}^D} ||\mathbf{w}||^2 + C \sum_i^N \max(0, 1 - y_i f(\mathbf{x}_i))$$

- بردار ویژگی‌های $\boldsymbol{x}$ را به $\Phi(\boldsymbol{x})$ نگاشت می‌کنیم.
- مسئله را برای $\boldsymbol{w}$ در فضای با بعد بالاتر $\mathbb{R}^D$ حل می‌کنیم.
- اگر $d \gg D$ باشد تعداد پارامترهایی که برای آموزش $\boldsymbol{w}$ نیاز است، خیلی زیاد می‌شود. آیا می‌توان به‌گونه‌ای از آن اجتناب کرد؟

○ طبقه‌بند دوگان در فضای ویژگی‌های تبدیل‌یافته:

Classifier:

$$f(\mathbf{x}) = \sum_i^N \alpha_i y_i \mathbf{x}_i^\top \mathbf{x} + b$$

$$\rightarrow f(\mathbf{x}) = \sum_i^N \alpha_i y_i \Phi(\mathbf{x}_i)^\top \Phi(\mathbf{x}) + b$$

Learning:

$$\max_{\alpha_i \geq 0} \sum_i \alpha_i - \frac{1}{2} \sum_{jk} \alpha_j \alpha_k y_j y_k \mathbf{x}_j^\top \mathbf{x}_k$$

$$\rightarrow \max_{\alpha_i \geq 0} \sum_i \alpha_i - \frac{1}{2} \sum_{jk} \alpha_j \alpha_k y_j y_k \Phi(\mathbf{x}_j)^\top \Phi(\mathbf{x}_k)$$

subject to

$$0 \leq \alpha_i \leq C \text{ for } \forall i, \text{ and} \sum_i \alpha_i y_i = 0$$

یادآوری: یادگیری ماشین

- طبقه‌بند دوگان در فضای ویژگی‌های تبدیل‌یافته:
  - $\Phi(\boldsymbol{x})$ تنها به صورت جفت‌های $\Phi(\boldsymbol{x}_j)^{\mathrm{T}}\Phi(\boldsymbol{x}_i)$ ظاهر می‌شوند.
  - وقتی ضرب‌های اسکالر محاسبه شدند، فقط بایستی بردار N-بعدی α آموزش داده شود.
  - نیازی نیست مانند مسئله اولیه آموزش در فضای D بعدی انجام شود.
- **کرنل** $k(x_j, x_i)$ را به صورت $k(x_j, x_i) = \Phi(x_j)^{\mathrm{T}}\Phi(x_i)$ تعریف می‌کنیم.

Classifier:

$$f(\mathrm{x}) = \sum_{i}^{N} \alpha_i y_i \, k(\mathrm{x}_i, \mathrm{x}) + b$$

Learning:

$$\max_{\alpha_i \geq 0} \sum_i \alpha_i - \frac{1}{2} \sum_{jk} \alpha_j \alpha_k y_j y_k \, k(\mathrm{x}_j, \mathrm{x}_k)$$

subject to

$$0 \leq \alpha_i \leq C \text{ for } \forall i, \text{ and } \sum_i \alpha_i y_i = 0$$

○ تبدیل‌های خاص:

$$\Phi : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \end{pmatrix} \quad \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

$$\begin{aligned} \Phi(\mathbf{x})^\top \Phi(\mathbf{z}) &= \left(x_1^2, x_2^2, \sqrt{2}x_1x_2\right) \begin{pmatrix} z_1^2 \\ z_2^2 \\ \sqrt{2}z_1z_2 \end{pmatrix} \\ &= x_1^2z_1^2 + x_2^2z_2^2 + 2x_1x_2z_1z_2 \\ &= (x_1z_1 + x_2z_2)^2 \\ &= (\mathbf{x}^\top \mathbf{z})^2 \end{aligned}$$

○ Kernel Trick:

○ طبقه‌بند را می‌توان بدون محاسبه صریح $\Phi(x)$ آموزش داده و بر داده‌ها اعمال نمود.

○ تنها چیزی که نیاز است محاسبه کرنل $k(x,z) = (x^T z)^2$ است.

یادآوری: یادگیری ماشین

- چند نوع کرنل:
  - کرنل خطی:

$$k(x,x') = x^T x'$$

  - کرنل چندجمله‌ای:

$$k(x,x') = (1 + x^T x')^d \text{ for any } d > 0$$

  - کرنل گوسی:

$$k(x,x') = exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \text{ for } \sigma > 0$$

یادآوری: یادگیری ماشین

- طبقه‌بند SVM با کرنل گوسی:

N = size of training data

$$f(\mathbf{x}) = \sum_{i}^{N} \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b$$

weight (may be zero)

support vector

Gaussian kernel $k(\mathbf{x}, \mathbf{x}') = \exp\left(-||\mathbf{x} - \mathbf{x}'||^2 / 2\sigma^2\right)$

Radial Basis Function (RBF) SVM

$$f(\mathbf{x}) = \sum_{i}^{N} \alpha_i y_i \exp\left(-||\mathbf{x} - \mathbf{x}_i||^2 / 2\sigma^2\right) + b$$

کرنل‌های گرافی

- مسئله مقایسه گراف‌ها (Graph Comparison Problem):
- دو گراف G و G' را در فضای گرافی $\mathcal{G}$ در نظر می‌گیریم. مسئله مقایسه گراف‌ها به صورت تعیین نگاشت زیر تعریف می‌شود:

$$s: \ \mathcal{G} \times \mathcal{G} \ \rightarrow \ \mathbb{R}$$

- به‌گونه‌ای که $s(G,G')$ شباهت یا عدم شباهت بین دو گراف G و G' را کمّی کند.

کرنل‌های گرافی

- فاصله‌های گرافی:
  - فاصله‌های ویرایش گرافی:
    - تعداد عملگرهای لازم برای تبدیل G به G' شمرده می‌شود.
    - به انواع مختلف عملگرها هزینه‌ای تخصیص داده می‌شود (حذف/اضافه کردن یال/رأس، تصحیح برچسب‌ها، ...).
  - تعاریف توپولوژیک
    - هر گراف به یک بردار ویژگی نگاشت می‌شود.
    - از فواصل مختلف برداری برای محاسبه فاصله گراف‌ها استفاده می‌شود.
  - گشت تصادفی (Random Walks):
    - گشت‌های یکسان در دو گراف ورودی G و G' شمرده می‌شود.
    - گشت‌ها دنباله‌هایی از رأس‌ها هستند که در آنها اجازه تکرار رأس‌ها وجود دارد.

کرنل‌های گرافی

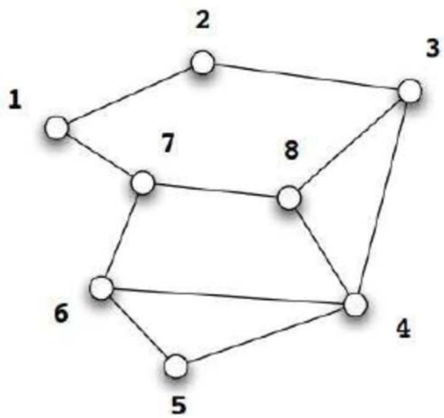
- ماتریس کرنل

$$\begin{bmatrix} K(G_1, G_1), K(G_1, G_2), \dots, K(G_1, G_n) \\ K(G_2, G_1), K(G_2, G_2), \dots, K(G_2, G_n) \\ \dots \\ K(G_n, G_1), K(G_n, G_2), \dots, K(G_n, G_n) \end{bmatrix}$$

- کرنل گرافی برای هر جفت گراف از مجموعه داده‌های آموزشی محاسبه می‌شود.
- این ماتریس به عنوان ورودی تابع SVM برای ایجاد مدل طبقه‌بندی استفاده می‌شود.
  - یا هر روش طبقه‌بندی کرنل‌دار دیگری

کرنل‌های گرافی

- مثال: کرنل انتشار مسیر بین دو گراف باینری



- ماتریس مجاورت یک گراف: A

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

- درایه‌های ماتریس A^K:
  - A_{ij}^K: تعداد مسیرهای به طول K از رأس i به رأس j

$$A^2 = \begin{bmatrix} 2 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 2 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 3 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 4 & 1 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 3 & 0 & 2 \\ 0 & 1 & 1 & 2 & 1 & 0 & 3 & 0 \\ 1 & 1 & 1 & 1 & 1 & 2 & 0 & 3 \end{bmatrix}$$

کرنل‌های گرافی

- مثال: کرنل انتشار مسیر بین دو گراف باینری

- معیاری از شباهت بین دو رأس:
  - مجموع تعداد مسیرهای یک‌یالی، دو یالی و ... از یک رأس به رأس دیگر با در نظر گرفتن یک ضریب کاهشی برای مسیرهای طولانی‌تر

$$k(i,j) = \left[\sum_k \frac{\lambda^k}{k!} A^k\right]_{ij} = [\exp(\lambda A)]_{ij} \qquad 0 < \lambda < 1$$

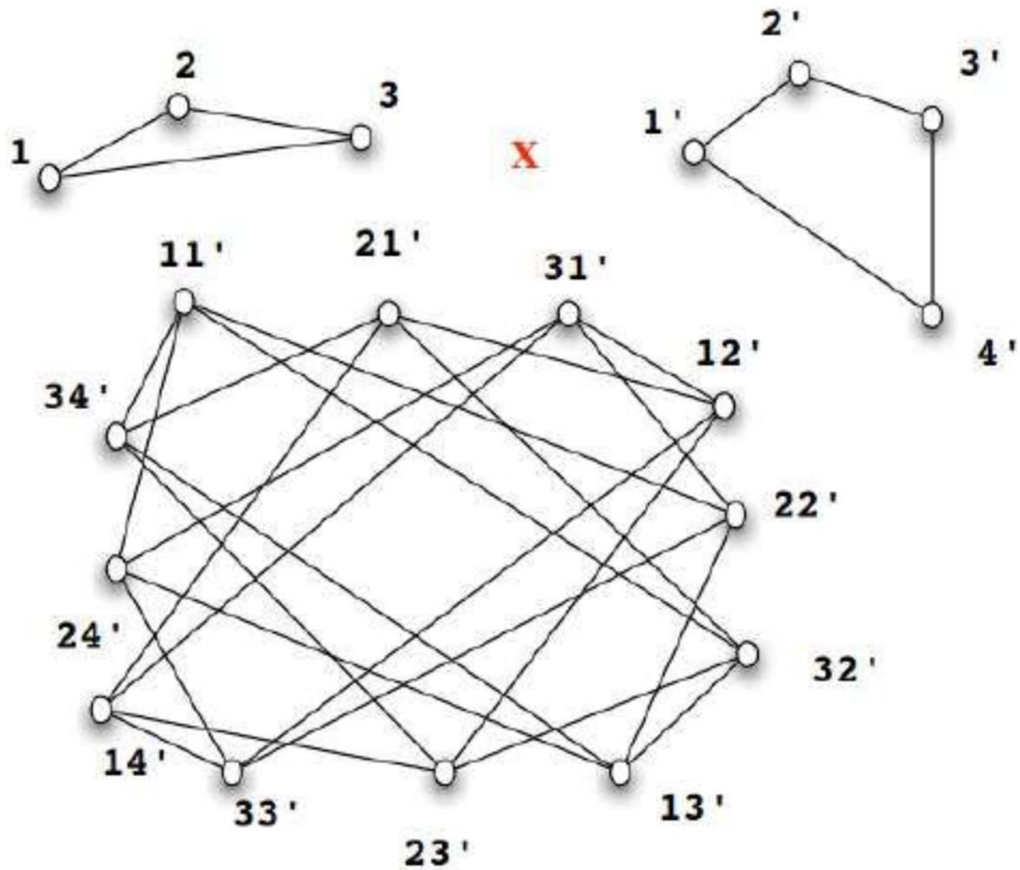
- تعمیم این ایده برای محاسبه شباهت بین دو گراف؟
  - یک گراف جدید با استفاده از ضرب کرونکر بین دو گراف می‌سازیم:

$$V_\times(G \times G') = \{(v, v') : v \in V, v' \in V'\}$$
$$E_\times(G \times G') = \{((v, v'), (w, w')) : (v, w) \in E, (v', w') \in E'\}$$

کرنل‌های گرافی

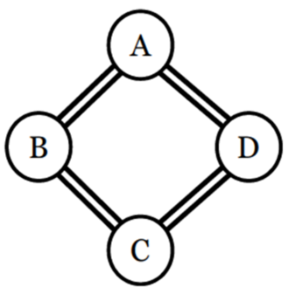
- مثال: کرنل انتشار مسیر بین دو گراف باینری

- مثالی از ضرب کرونکر دو گراف:

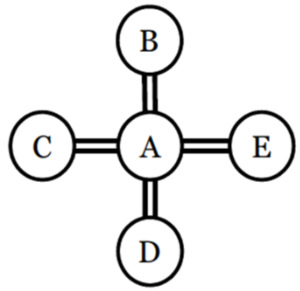


کرنل‌های گراف

- مثال: کرنل انتشار مسیر بین دو گراف باینری



Type-A



Type-B

Type-A	A	B	C	D
A	0	1	1	0
B	1	0	0	1
C	1	0	0	1
D	0	1	1	0

Type-B	A	B	C	D	E
A	0	1	1	1	1
B	1	0	0	0	0
C	1	0	0	0	0
D	1	0	0	0	0
E	1	0	0	0	0

Intuition: multiply each entry of Type-A by ***entire matrix*** of Type-B

Type-A		A					B					C					D				
	Type-B	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
A	A	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1	0	0	0	0	0
	B	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	C	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	D	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	E	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B	A	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
	B	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	C	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	D	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	E	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
C	A	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
	B	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	C	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	D	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	E	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
D	A	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1	0	0	0	0	0
	B	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	C	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	D	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	E	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0

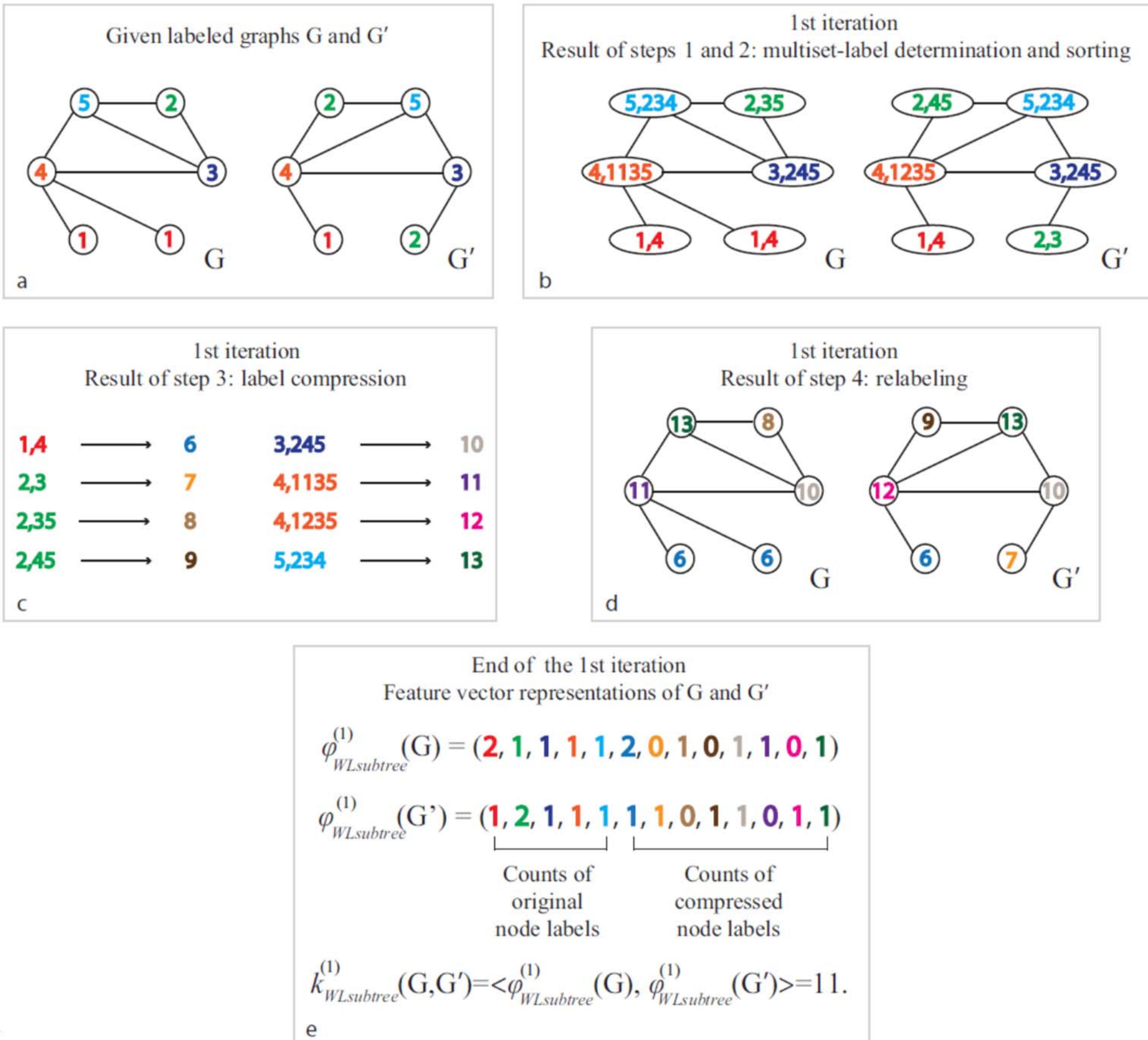
کرنل‌های گراف

- مثال: کرنل انتشار مسیر بین دو گراف باینری

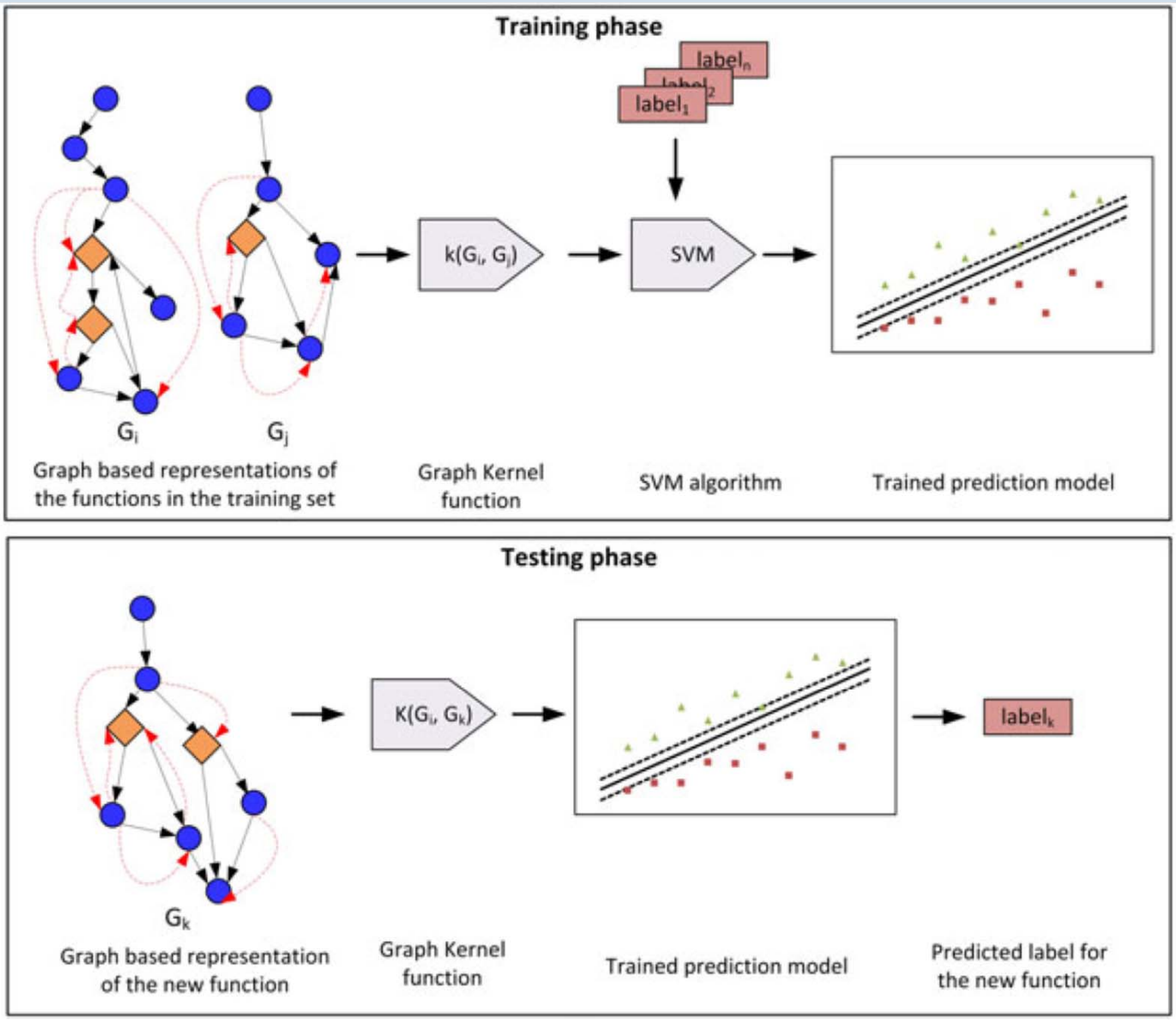
$$k(G,G') = \frac{1}{|G|\,|G'|}\sum_k \frac{\lambda^k}{k!}\,\mathbf{e}^\top A_\times^k\,\mathbf{e} = \frac{1}{|G|\,|G'|}\,\mathbf{e}^\top \exp(\lambda A_\times)\,\mathbf{e}$$

- $A_\times$: ماتریس مجاورت ضرب کرونکر دو گراف G و G'
- $\boldsymbol{e}$: بردار تمام ۱

○ مثال: کرنل Weisfeiler-Lehman subtree



کرنل‌های گرافی



کاربردها

- بررسی تغییرات شبکه‌های ارتباطات مغزی بر اثر اختلالات مغزی مانند:
  - اختلالات طیف اوتیسم
  - آلزایمر
  - افسردگی
  - پارکینسون
  - اسکیزوفرنی
  - صرع
- رابط‌های مغز-رایانه
- تشخیص احساسات