

# آنالیز چند متغیره‌ی مخلوط‌ها در سیستم‌های شیمیایی فاقد گزینش پذیری طیفی بر اساس سنجش بر مبنای جایگزینی شنا ساگر :

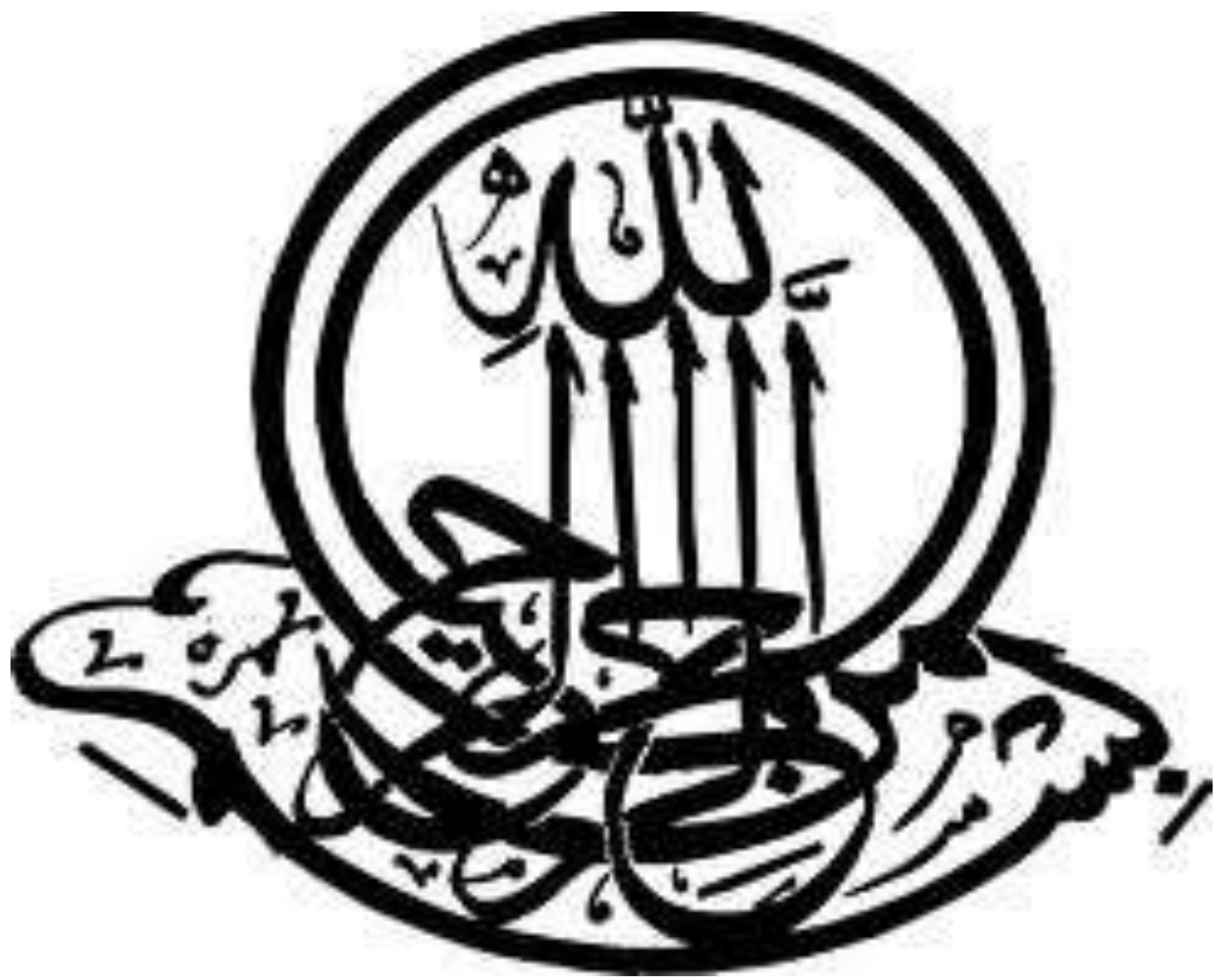
## اندازه‌گیری همزمان هیستیدین و سیستئین

پایان‌نامه کارشناسی ارشد  
سمیه خدادادی کریموند

استاد راهنما: دکتر حمید عبداللّهی

استاد مشاور: اکرم رستمی

شهریور ۱۳۹۲



تقدیم به عزیزانہ :

پدر و مادرم

خواہر و برادرانم

اساتید و معلمانم

## تشکر و قدردانی

همد و سپاس کردگاری را سزااست که فرصت کسب علم و دانش را به ما عطا فرمود.

.....مول و نادانی من و عصبیان و گستاخی من تو را باز نداشت از اینکه راهنمایی ام کنی به سوی صراط قربیت و موفقم

گردانی به آنچه رضا و فوشنودی توست.

زالترین سپاسها تقدیم به آنانی که گوهر وجودشان، نسیم کلامشان و باران مہبتشان را همواره بی هیچ منت و ادعایی ارزانیم داشتند:

از خانواده‌ی عزیزم به خاطر حمایت های عاشقانه‌شان در تمام دوران زندگی ام بینهایت سپاسگزارم.

از استاد راهنمای عزیزم جناب آقای دکتر حمید عبداللہی، که دلسوزی، تلاش و کوششان در تعلیم و تربیت و انتقال معلومات و تجربیات ارزشمند در کنار برقراری رابطه صمیمی و دوستانه با دانشجویان و ایجاد فضائی دلنشین برای کسب علم و دانش و درک شرایط دانشجویان ستودنی است، کمال تشکر را دارم.

لازم می دانم مراتب قدردانی و تشکر خود را همراه با بهترین آرزوها به استاد مشاور عزیزم خانم اکرم رستمی که از تجربیات و دانسته هایشان استفادہ نمودم بفضورشان تقدیم نمایم.

## چکیده

در روش سنجش بر مبنای جایگزینی شناساگر، یک شناساگر رنگی یا فلئورسانس کننده به صورت کاملاً برگشت پذیر با گیرنده پیوند برقرار می‌کند. با اضافه کردن آنالیتی رقابتی به این سیستم، آنالیت و شناساگر برای پیوند با گیرنده وارد رقابت می‌شوند. از آنجا که میل ترکیبی آنالیت با گیرنده بیشتر از شناساگر با گیرنده است، آنالیت جایگزین شناساگر پیوند داده شده با میزبان می‌شود. این جایگزینی سبب آزاد شدن شناساگر و تغییر سیگنال نوری آن می‌شود. شرط اساسی در این روش این است که میل ترکیبی بین شناساگر و گیرنده باید قابل مقایسه با میل ترکیبی بین آنالیت و گیرنده باشد. در آنالیز مخلوط‌های شیمیایی نیاز به گزینش پذیری طیفی الزامیست. بدین معنا که تنها گونه‌هایی که یا خود و یا محصولات حاصل از تعادلات شیمیایی‌شان دارای طیف‌های منحصر به فردی باشند، قابلیت اندازه‌گیری همزمان را دارا می‌باشند.

در روش سنجش بر مبنای جایگزینی شناساگر، تفاوت در ثابت تعادل گونه‌های رقیب تحت شرایط تعادلی مختلف، باعث ایجاد داده‌های اسپکتروسکوپی می‌شود که آنالیز همزمان آنها را در سیستم‌های شیمیایی فاقد گزینش پذیری طیفی مقدور می‌سازد.

در این تحقیق، با استفاده از روش سنجش بر مبنای جایگزینی شناساگر اندازه‌گیری همزمان دو آمینواسید هیستیدین و سیستئین بر اساس برهم‌کنش رقابتی‌شان با کمپلکس نیکل-موراکساید تحت شرایط تعادلی متفاوت و با به کارگیری روش‌های کالیبراسیون چندمتغیره مانند روش حداقل مربعات جزئی و شبکه‌ی عصبی مصنوعی تابع پایه شعاعی، با موفقیت انجام گرفت.

## فهرست

I .....	فهرست مطالب
IV.....	فهرست شکل‌ها
VI.....	فهرست جدول‌ها

فهرست مطالب ..... شماره صفحه

۱.....	فصل اول
۱.....	مقدمه
۱.....	۱-۱ مقدمه‌ای بر روش سنجش بر مبنای جایگزینی شناساگر
۲.....	۱-۱-۱ روش سنجش بر مبنای جایگزینی شناساگر
۳.....	۲-۱ آنالیز مخلوط‌های شیمیایی در سیستم‌های تعادلی
۴.....	۳-۱ روش‌های کالیبراسیون چند متغیره
۴.....	۱-۳-۱ روش حداقل مربعات جزئی
۹.....	۲-۳-۱ شبکه‌های عصبی مصنوعی
۹.....	۱-۲-۳-۱ شبکه‌ی عصبی زیستی
۱۰.....	۲-۲-۳-۱ ساختار شبکه‌ی عصبی مصنوعی
۱۱.....	۱-۲-۲-۳-۱ وزن‌ها
۱۱.....	۲-۲-۲-۳-۱ توابع انتقال
۱۲.....	۳-۲-۳-۱ انواع یادگیری شبکه‌های عصبی مصنوعی
۱۲.....	۱-۳-۲-۳-۱ یادگیری تحت نظارت
۱۲.....	۲-۳-۲-۳-۱ یادگیری بدون نظارت (یادگیری خود سامانده)
۱۲.....	۴-۲-۳-۱ انواع شبکه‌های عصبی بر اساس تعداد لایه‌ها
۱۲.....	۱-۴-۲-۳-۱ شبکه‌های عصبی تک لایه
۱۲.....	۲-۴-۲-۳-۱ شبکه‌های عصبی چند لایه

۱۲.....	شبکه‌ی عصبی پس انتشار ۱-۲-۴-۲-۳-۱
۱۳.....	شبکه‌ی عصبی تابع پایه‌ی شعاعی ۲-۲-۴-۲-۳-۱
۱۴.....	توپولوژی شبکه‌ی عصبی مصنوعی ۵-۲-۴-۱
۱۵.....	کاهش تعداد متغیر ۴-۱
۱۵.....	آنالیز بر پایه‌ی مدل ۵-۱
۱۶.....	الگوریتم نیوتون-رافسون ۱-۵-۱
۱۷.....	الگوریتم نیوتون-گوس ۲-۵-۱
۲۰.....	معیار پایانی ۱-۲-۵-۱
۲۰.....	توسعه‌ی لونبرگ/مارکوات ۲-۲-۵-۱
۲۲.....	نتیجه‌گیری ۶-۱
۲۳.....	فصل دوم
۲۳.....	مروری بر تحقیقات گذشته
۲۳.....	تحقیقات انجام شده در زمینه‌ی اندازه‌گیری همزمان هیستیدین و سیستئین ۱-۲
۲۳.....	مقدمه ۱-۱-۲
۲۶.....	برخی از تحقیقات انجام شده بر اساس روش سنجش بر مبنای جایگزینی شناساگر (IDA): ۲-۲
۳۰.....	نتیجه‌گیری ۳-۲
۲۳.....	فصل سوم
۳۱.....	مواد، دستگاه‌ها و روش کار
۳۱.....	دستگاه‌ها و نرم‌افزارها ۱-۳
۳۱.....	تهیه‌ی محلول‌ها ۲-۳
۳۳.....	روش کار ۳-۳

فصل چهارم .....	۳۸
بحث و نتیجه‌گیری .....	۳۸
۱-۴ مقدمه .....	۳۸
۲-۴ بررسی سیستم شبیه‌سازی شده .....	۳۸
۳-۴ اندازه‌گیری همزمان دو آمینواسید H و C .....	۴۴
۱-۳-۴ شبکه‌ی عصبی مصنوعی تابع پایه‌ی شعاعی (RBFNN) .....	۴۴
۲-۳-۴ روش حداقل مربعات جزئی (PLS) .....	۴۸
۴-۴ اندازه‌گیری همزمان هیستیدین و سیستین در سیستم واقعی .....	۵۱
۵-۴ بهینه‌سازی شرایط آزمایش .....	۵۴
۱-۵-۴ بهینه‌سازی pH .....	۵۴
۶-۴ تعیین گستره‌ی کالیبراسیون .....	۵۶
۷-۴ ارتباط غیرخطی بین غلظت‌های کل و تعادلی .....	۵۸
۸-۴ اندازه‌گیری همزمان دو آمینواسید هیستیدین و سیستین با استفاده از دو تکنیک کالیبراسیون چند متغیره :	
PLS و RBFNN .....	۵۹
۱-۸-۴ کالیبراسیون چند متغیره‌ی غیر خطی با استفاده از روش RBFNN .....	۵۹
۱-۱-۸-۴ کاهش تعداد متغیرها .....	۶۰
۲-۸-۴ کالیبراسیون چند متغیره‌ی غیرخطی با استفاده از روش حداقل مربعات جزئی (PLS) .....	۶۵
۹-۴ بررسی اثر مزاحمت در اندازه‌گیری همزمان هیستیدین و سیستین .....	۶۹
۱۰-۴ اندازه‌گیری در نمونه‌های حقیقی .....	۷۳
۱۱-۴ نتیجه‌گیری .....	۷۴



فهرست شکل‌ها .....	شماره صفحه
فصل اول .....	۱
۱-۱ شماتیکی از روش شناساگر- فضا ساز- گیرنده .....	۲
۲-۱ شماتیکی از روش سنجش بر مبنای جایگزینی شناساگر .....	۲
۳-۱ نمونه‌ای از یک شبکه‌ی عصبی زیستی .....	۱۰
۴-۱ مدل نورون در یک شبکه‌ی عصبی مصنوعی .....	۱۱
۵-۱ طرحی از یک RBFNN .....	۱۴
۶-۱ شماتیکی از الگوریتم نیوتون- گوس- لونبرگ / مارکوات .....	۲۲
فصل دوم .....	۲۴
۱-۲ شماتیکی از مجموعه‌ی $Hg^{+2}/DNA/TO$ .....	۲۵
۶-۲ نمودار امتیاز حاصل از شناسایی آمینواسیدهای گروه دوم .....	۲۹
۷-۲ پاسخ کمپلکس Ni-Mur به آمینواسیدهای متفاوت .....	۳۰
فصل سوم .....	۳۱
۱-۳ ساختار تریس .....	۳۲
۲-۳ ساختار شناساگر موراکساید .....	۳۲
فصل چهارم .....	۳۸
۱-۴ سیگنال مشابه‌سازی برای شناساگر I و کمپلکس آن MI .....	۴۰
۲-۴ طیف مشابه‌سازی برای کمپلکس‌های MH2 و MC2 در غلظت‌هایی تجزیه‌ای برابر از H و C .....	۴۰
۳-۴ پروفایل‌های غلظتی در داده‌ی شبیه‌سازی شده .....	۴۲
۴-۴ طیف شبیه‌سازی از دو نمونه با ترکیب غلظتی متفاوت از H و C اما پاسخ طیفی مشابه .....	۴۳
۵-۴ طیف مشابه‌سازی و سر هم زده شده از دو مخلوط تحت دو شرایط مختلف .....	۴۴
۶-۴ منحنی همبستگی بین غلظت واقعی و غلظت پیش‌بینی با RBFNN برای دو آمینواسید H و C .....	۴۷
۷-۴ منحنی PRESS بر حسب تعداد فاکتورها برای دو آمینواسید H و C .....	۴۹
۸-۴ منحنی همبستگی بین غلظت واقعی و غلظت پیش‌بینی با PLS1 برای دو آمینواسید H و C .....	۵۰
۹-۴ طیف خالص شناساگر موراکساید و کمپلکس نیکل- موراکساید .....	۵۲
۱۰-۴ منحنی نسبت مولی برای سه کمپلکس Ni-Mur و Ni-His2 و Ni-Cys2 .....	۵۳

- ۱۱-۴ نمایش تغییر سیگنال در pH های ۶ تا ۹ به ازای تیتراژ شدن کمپلکس نیکل-موراکساید با هیستیدین ..... ۵۴
- ۱۲-۴ نمایش تغییر سیگنال در pH های ۶ تا ۹ به ازای تیتراژ شدن کمپلکس نیکل-موراکساید با سیستئین ..... ۵۴
- ۱۳-۴ نمایش سیگنال حاصل از سری کالیبراسیون تحت دو شرایط pH ۷ و ۸ ..... ۵۷
- ۱۴-۴ نمایش سیگنال حاصل از سری کنترل تحت دو شرایط pH ۷ و ۸ ..... ۵۷
- ۱۵-۴ نمایش سیگنال حاصل از سری ارزیابی تحت دو شرایط pH ۷ و ۸ ..... ۵۸
- ۱۶-۴ نمودار سه‌بعدی غلظت تعادلی شناساگر موراکساید بر حسب، هیستیدین و سیستئین ..... ۵۹
- ۱۷-۴ نمایش منحنی همبستگی بین غلظت واقعی و غلظت پیش‌بینی شده با RBFNN در سری ارزیابی برای آمینواسید هیستیدین و سیستئین ..... ۶۲
- ۱۸-۴ منحنی PRESS بر حسب تعداد فاکتورها برای هیستیدین و سیستئین ..... ۶۶
- ۱۹-۴ نمایش منحنی همبستگی بین غلظت واقعی و غلظت پیش‌بینی شده با PLS1 برای آمینواسید هیستیدین و سیستئین ..... ۶۷
- ۲۰-۴ سیگنال حاصل از تأثیر آمینواسیدهای مختلف بر کمپلکس نیکل-موراکساید ..... ۷۲

فهرست جدول‌ها	شماره صفحه
فصل سوم	۳۱
۳.۱ مقادیر متفاوت غلظت هیستیدین و سیستئین برای سری کالیبراسیون بر حسب میکرو مولار ( $\mu\text{M}$ )	۳۵
۳.۲ مقادیر متفاوت غلظت هیستیدین و سیستئین برای سری کنترل بر حسب میکرو مولار ( $\mu\text{M}$ )	۳۶
۳.۳ مقادیر متفاوت غلظت هیستیدین و سیستئین برای سری ارزیابی بر حسب میکرو مولار ( $\mu\text{M}$ )	۳۷
فصل چهارم	۳۸
۴.۱ مدل پروفایل‌های غلظتی گونه‌های موجود در تعادل مذکور با توجه به الگوریتم نیوتون-رافسون	۴۲
۴.۲ مقادیر بهینه پارامترهای مختلف شبکه‌ی <b>RBFNN</b> به دست آمده از داده‌ی شبیه‌سازی شده، برای دو آمینواسید <b>H</b> و <b>C</b>	۴۶
۴.۳ مقادیر خطای به دست آمده از سری پیش‌بینی برای هر دو آمینواسید <b>H</b> و <b>C</b> ، برای داده شبیه‌سازی و با روش <b>PLS1</b> و <b>RBFNN</b>	۵۱
۴.۴ مقادیر ثابت‌های تشکیل کمپلکس‌ها تحت دو شرایط ۷ و ۸ $\text{pH}$	۵۶
۴.۵ نتایج مقادیر بهینه پارامترهای مختلف شبکه <b>RBFNN</b> برای دو آمینواسید هیستیدین و سیستئین	۶۲
۴.۶ مقادیر مختلف غلظت واقعی و غلظت پیش‌بینی برای هیستیدین با <b>RBFNN</b> در سری ارزیابی	۶۴
۴.۷ مقادیر مختلف غلظت واقعی و غلظت پیش‌بینی برای سیستئین با <b>RBFNN</b> در سری ارزیابی	۶۵
۴.۸ مقادیر مختلف غلظت واقعی و غلظت پیش‌بینی برای هیستیدین با <b>PLS1</b>	۶۹
۴.۹ مقادیر مختلف غلظت واقعی و غلظت پیش‌بینی برای سیستئین با <b>PLS1</b>	۷۰
۴.۱۰ حداقل غلظت آمینواسیدهای متفاوت برای ایجاد خطای بالاتر از ۵٪ در اندازه‌گیری جذب کمپلکس نیکل-موراکساید	۷۳
۴.۱۱ مقادیر مختلف غلظت واقعی و غلظت پیش‌بینی برای آمینواسید هیستیدین در نمونه‌های حقیقی	۷۴
۴.۱۱ مقادیر مختلف غلظت واقعی و غلظت پیش‌بینی برای آمینواسید سیستئین در نمونه‌های حقیقی	۷۵

# فصل اول

## مقدمه

### ۱-۱ مقدمه‌ای بر روش سنجش بر مبنای جایگزینی شناساگر<sup>۱</sup>

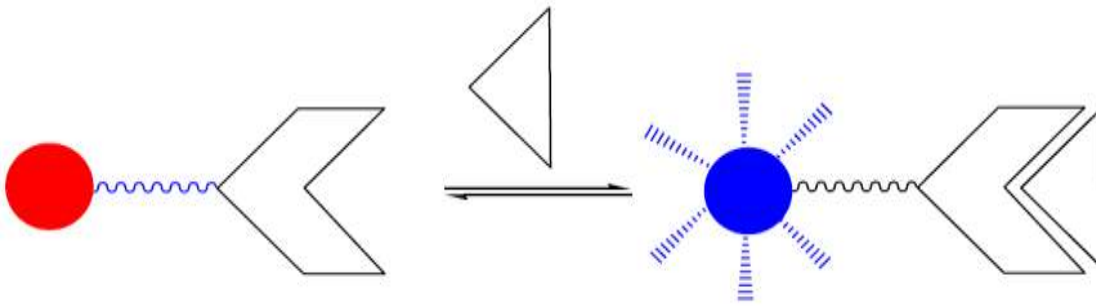
شناساگر-فاصله دهنده-گیرنده<sup>۲</sup>، و یا به اختصار ISR، روشی است که به صورت گسترده به عنوان حسگر شیمیایی استفاده می‌شود. در این روش همان‌گونه که در شکل (۱-۱) نشان داده شده است، یک شناساگر (رنگی یا فلئورسانس) از طریق یک فاصله دهنده و با پیوند کووالانسی به گیرنده متصل می‌شود. اضافه شدن آنالیتی که بتواند با گیرنده پیوند برقرار کند، باعث تغییر سیگنال فلئورسانس و یا جذب شناساگر می‌شود. از این تغییر سیگنال می‌توان برای اندازه‌گیری ثابت‌های تشکیل و یا استوکیومتری استفاده شود. گرچه ISR روشی رایج برای شناسایی و تشخیص مولکول-هاست، اما محدودیت‌ها و معایبی نیز می‌توان برای آن برشمرد، که از آن جمله می‌توان به موارد زیر اشاره کرد:

اول اینکه، پیوند بین شناساگر و گیرنده ممکن است نیازمند سنتز پیچیده و مشکل باشد. دوماً، استفاده از یک گیرنده برای چندین شناساگر و یا یک شناساگر برای چند میزبان، کار آسانی نیست. یک راه-حل مناسب برای رفع این مشکل، استفاده از روش سنجش بر مبنای جایگزینی شناساگر است [۱].

---

<sup>۱</sup> Indicator Displacement Assay (IDA)

<sup>۲</sup> Indicator-Spacer-Receptor (ISR)

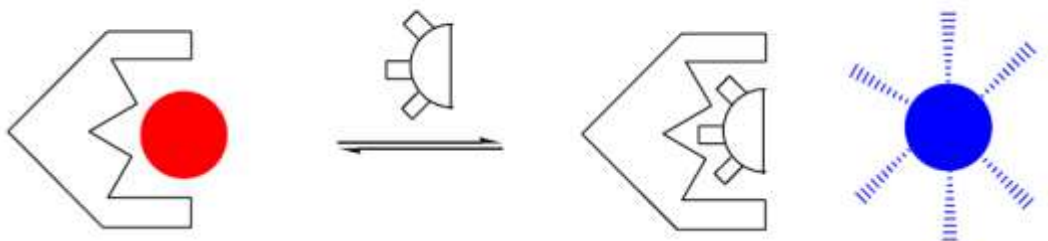


آنالیت- گیرنده- فضا ساز- شناساگر      آنالیت      گیرنده- فضا ساز- شناساگر

شکل ۱-۱. شماتیکی از روش شناساگر- فضا ساز- گیرنده. پیوند آنالیت با گیرنده سبب تغییر رنگ شناساگر از قرمز به آبی شده است.

### ۱-۱-۱ روش سنجش بر مبنای جایگزینی شناساگر

در روش سنجش بر مبنای جایگزینی شناساگر و یا به اختصار IDA، در مرحله اول، یک شناساگر به صورت کاملاً برگشت پذیر با گیرنده پیوند برقرار می کند. با اضافه شدن آنالیتی رقابتی به این سیستم، آنالیت مورد نظر جایگزین شناساگر می شود. این جایگزینی سبب آزاد شدن، و تغییر سیگنال نوری شناساگر می شود (شکل ۱-۲). بر این مبنای عامل اساسی و مهم در IDA، این است که میل ترکیبی بین شناساگر و گیرنده، قابل مقایسه با آنالیت و گیرنده باشد.



شناساگر- گیرنده      آنالیت      آنالیت- گیرنده      شناساگر آزاد

شکل ۱-۲. شماتیکی از روش سنجش بر مبنای جایگزینی شناساگر. پیوند رقابتی آنالیت با گیرنده سبب جایگزینی شناساگر از حفره گیرنده شده که نتیجه آن تغییر رنگ شناساگر از قرمز به آبی است.

متداول‌ترین برهمکنش‌ها بین شناساگر و یا آنالیت با میزبان (گیرنده)، از نوع پیوند هیدروژنی، نیروهای الکترواستاتیک و یا تشکیل کمپلکس با مراکز فلزی می‌باشد. این برهمکنش‌های غیرکووالانت به عوامل متعددی بستگی دارند که از آن جمله می‌توان به ساختار هندسی میزبان، بار و خاصیت هیدروفوبی آن، و همچنین محیط حلال اشاره کرد.

IDA مزایای متعددی بر روش قدیمی‌تر ISR دارد. اول اینکه، این روش نیازی به تشکیل پیوند کووالانسی شناساگر با گیرنده ندارد. دوماً، به این دلیل که هیچ پیوند کووالانسی بین شناساگر و گیرنده تشکیل نمی‌شود، می‌توان چندین شناساگر متفاوت را برای یک گیرنده‌ی مشخص به کار برد. سوماً، این روش سنجش در هر دو محیط آلی و آبی به خوبی قابل استفاده می‌باشد.

روش سنجش بر مبنای جایگزینی شناساگر به سه دسته تقسیم می‌شود. دسته‌ی اول، سنجش بر مبنای جایگزینی شناساگر رنگ‌سنجی<sup>۱</sup> است، که در آن از یک شناساگر رنگی استفاده می‌شود. دسته‌ی دوم، سنجش بر مبنای جایگزینی شناساگر فلئورسانس کننده<sup>۲</sup> است، که در آن از یک شناساگر با خاصیت فلئورسانس کننده‌گی استفاده می‌شود. و دسته‌ی سوم، سنجش بر مبنای جایگزینی شناساگر تشکیل کمپلکس با مراکز فلزی<sup>۳</sup> است، که در آن از یک مرکز فلزی و شناساگری رنگی یا فلئورسانس کننده با خاصیت کیلیت‌کنندگی استفاده می‌شود [۲].

## ۲-۱ آنالیز مخلوط‌های شیمیایی در سیستم‌های تعادلی

تعادل شیمیایی به حالتی گفته می‌شود که در آن، فعالیت شیمیایی واکنشگرها و محصولات در واکنش شیمیایی با گذشت زمان ثابت بماند و تغییر نکند. معمولاً این تعادل در شرایطی حاصل می‌شود که سرعت واکنش رفت با سرعت واکنش برگشت برابر باشد.

آنالیز مخلوط‌های شیمیایی در سیستم‌های تعادلی با استفاده از کمومتریکس کاملاً مرسوم می‌باشد. در همه‌ی این روش‌ها نیاز به گزینش‌پذیری طیفی الزامیست. بدین معنا که، تنها گونه‌هایی که یا خود و یا محصولات حاصل از تعادلات شیمیایی‌شان دارای طیف‌های منحصر به فرد باشند قابلیت اندازه‌گیری همزمان را دارا می‌باشند.

---

<sup>1</sup> Colorimetric IDA (C-IDA)

<sup>2</sup> Fluorescent IDA (F-IDA)

<sup>3</sup> Metal-complexing IDA (M-IDA)

همان‌گونه که در قسمت قبل بیان شد، در روش سنجش بر مبنای جایگزینی شناساگر، از رقابت بین شناساگر و آنالیت برای پیوند با مولکول گیرنده، جهت شناسایی و تشخیص آنالیت استفاده می‌شود. در واقع، تفاوت در ثابت‌های تعادل گونه‌های رقیب تحت شرایط تعادلی متفاوت، عامل ایجاد سیگنال‌های اسپکتروسکوپی منحصر به فرد برای گونه‌هایی می‌شود که فاقد گزینش‌پذیری طیفی بودند. در نتیجه آنالیز همزمان این گونه‌ها تحت شرایط تعادلی متفاوت امکان‌پذیر می‌باشد.

### ۳-۱ روش‌های کالیبراسیون چند متغیره<sup>۱</sup>

روش‌هایی هستند که در آنها به جای استفاده از یک متغیر (برای مثال استفاده از طول موج ماکسیمم)، از تعداد بیشتری طول موج (برای مثال استفاده از کل طیف)، جهت کالیبراسیون استفاده می‌شود. از مزایای این روش می‌توان به افزایش دقت این روش‌ها نسبت به کالیبراسیون تک متغیره اشاره کرد. با استفاده از این روش‌ها می‌توان مدل‌هایی ساخت که بین متغیرهای وابسته و متغیرهای مستقل ارتباط برقرار کند، و از طریق این ارتباط به اطلاعات نمونه‌های مجهول پی برد. مراحل کالیبراسیون شامل: طراحی آزمایش، انتخاب مدل، تخمین پارامترها و پیش‌بینی نمونه‌های مجهول می‌باشد. در روش‌های کالیبراسیون تک‌متغیره، امکان تصحیح مزاحمت‌ها بدون وارد کردن اطلاعات اضافی وجود ندارد. در حالی که روش‌های کالیبراسیون چندمتغیره، بدون وارد کردن اطلاعات اضافی قادر به تفکیک اطلاعات مفید مربوط به آنالیت از اطلاعات نامناسب مثل نویز و مزاحمت‌های شناخته شده می‌باشند. از جمله روش‌های کالیبراسیون چندمتغیره‌ی خطی می‌توان به رگرسیون جزء اصلی<sup>۲</sup> [۳] و حداقل مربعات جزئی<sup>۳</sup> [۴] اشاره کرد. همچنین شبکه‌های عصبی مصنوعی [۵] و حداقل مربعات جزئی غیرخطی، جزو روش‌های کالیبراسیون غیرخطی می‌باشند.

### ۱-۳-۱ روش حداقل مربعات جزئی

روش حداقل مربعات جزئی و یا به اختصار PLS، اولین بار در سال ۱۹۷۵ توسط هرمان ولد برای مدل کردن دسته‌ای از داده‌های پیچیده ارائه شد. در این مدل از روش NIPALS<sup>۴</sup> استفاده شده است که راهی ساده و مناسب جهت تخمین پارامترهای این مدل می‌باشد.

<sup>1</sup> Multivariate Calibration

<sup>2</sup> Principle Component Regression

<sup>3</sup> Partial Least Squares

<sup>4</sup> Non Linear Iterative Partial Least Squares

PCR ترکیبی از رگرسیون خطی چندمتغیره<sup>۱</sup> (MLR)، و PCA<sup>۲</sup> می‌باشد. MLR از جمله روش‌هایی است که برای سیستم‌های خطی کاربرد دارد، و در آن فاکتورهای را می‌یابد که بهترین همبستگی را بین متغیرهای مستقل و وابسته برقرار می‌کند. در PCA، فاکتورهای اصلی (PC)، از ماکسیمم مقدار واریانس روی پاسخ‌های دستگاهی تعیین می‌شوند که این فاکتورها قادر به دادن هیچگونه اطلاعاتی در مورد متغیرهای پیش‌بینی شونده نیستند. در PLS، هدف تعیین پارامترهایی است که بتواند علاوه بر ایجاد بیشترین واریانس، بهترین همبستگی را نیز بین متغیرهای وابسته و مستقل برقرار نماید. PLS از جمله روش‌های معکوس می‌باشد که در آن، ماتریس داده‌ی اندازه‌گیری شده متغیر مستقل، و ماتریس غلظت، متغیر وابسته است.

هدف روش PLS، پیش‌بینی متغیرهای وابسته یعنی غلظت، از متغیرهای مستقل می‌باشد. این پیش‌بینی توسط فاکتورهای عمود برهمی که متغیرهای پنهانی گفته می‌شوند انجام می‌گیرد. این روش، به ویژه برای زمانی که نیاز به پیش‌بینی یک مجموعه از متغیرهای وابسته از یک مجموعه‌ی خیلی بزرگ متغیرهای مستقل داریم، سودمند می‌باشد [۶].

برای فهم بهتر، پیش از پرداختن به روش PLS، روش حداقل مربعات به صورت مختصر توضیح داده می‌شود.

فرض کنید دو ماتریس، یکی حاوی داده‌ی جذبی و دیگری ماتریس غلظت‌های معلوم وجود دارند، که این دو ماتریس طبق رابطه‌ی (۱.۱) با یکدیگر مرتبط می‌باشند.

$$X_{(m \times n)} = Y_{(m \times p)} \times B_{(p \times n)} \quad (1.1)$$

که در این رابطه  $X$  ماتریس داده‌ی جذبی با ابعاد  $(m \times n)$  و  $Y$ ، ماتریس غلظت با ابعاد  $(m \times p)$ ، و  $B$  ضریب رگرسیون و با ابعاد  $(p \times n)$  می‌باشد.  $m$  تعداد نمونه‌ها،  $n$  تعداد طول موج‌ها و  $p$  تعداد گونه‌ها می‌باشد. این دو ماتریس توسط  $B$ ، به یکدیگر مرتبط می‌شوند. در واقع هدف، پیدا کردن این ضریب رگرسیون با استفاده از دو ماتریس شناخته شده‌ی جذب و غلظت، و سپس به کاربردن این ضریب برای داده‌ی ناشناخته، و به دست آوردن پارامتر مورد نظر می‌باشد.

برای به دست آوردن ضریب رگرسیون، طرفین معادله‌ی (۱.۱) در معکوس ماتریس غلظتی ضرب می‌شود (معادله‌ی ۱.۲). البته برای این مرحله، ماتریس غلظت باید مربع (مقارن و عناصر تصویری نسبت به قطر ماتریس یکسان باشد) و مرتبه کامل<sup>۳</sup> باشد. در این صورت از ضرب این ماتریس در

<sup>1</sup> Multivariate Linear Regression (MLR)

<sup>2</sup> Principal Component Analysis

<sup>3</sup> Full rank



معکوس خودش ماتریس همانی<sup>۱</sup> (**I**)، به دست می آید (معادله ی ۳.۱) که تمامی عناصر این ماتریس برابر یک می باشد.

$$Y^{-1}_{(p \times m)} \times X_{(m \times n)} = B_{(p \times n)} \quad (۲.۱)$$

$$Y^{-1}_{(p \times m)} \times Y_{(m \times p)} = I_{(p \times p)} \quad (۳.۱)$$

اما در صورتی که ماتریس غلظت، مربع نباشد، طبق معادله ی (۴.۱) عمل می کنیم.

$$(Y^T Y)^{-1} \times (Y^T Y) = Y^+ Y = I \quad (۴.۱)$$

و در نتیجه خواهیم داشت :

$$B_{(p \times n)} = Y^+_{(p \times m)} \times X_{(m \times n)} \quad (۵.۱)$$

حال می توان از این ضریب به دست آمده (**B**)، برای پیش بینی غلظت داده های ناشناخته طبق معادله ی (۶.۱) استفاده کرد.

$$X_{un} = Y_{un} \times B = Y_{un} (Y^+ X) \quad (۶.۱)$$

در PLS ماتریس داده (**X**)، متغیر مستقل و ماتریس غلظت (**Y**)، متغیر وابسته می باشد. در این روش، بالاترین کوواریانس بین **X** و **Y**، مدنظر است، و به دنبال بخش هایی از **X** می باشد که به **Y** مرتبط است.

هر دو ماتریس **X** و **Y** در این روش، به یک مجموعه فاکتورهای عمود بر هم و یک مجموعه بارهای ویژه تجزیه می شوند. بنابراین متغیرهای مستقل به صورت معادله ی (۷.۱) تجزیه می شوند.

$$X = TP^T \quad (۷.۱)$$

**T** ماتریس امتیاز<sup>۲</sup> و **P** ماتریس بار<sup>۳</sup> نامیده می شود.

همچنین متغیرهای وابسته نیز به صورت معادله ی (۸.۱) تجزیه می شوند.

$$Y = TQ^T \quad (۸.۱)$$

---

<sup>1</sup> Identity

<sup>2</sup> Score

<sup>3</sup> Loading

که در این معادله،  $Q$ ، ماتریس بار و ستون‌های  $T$ ، بردارهای امتیاز، می‌باشند و زمانی تجزیه‌ی دقیق  $X$  صورت می‌گیرد که تعدادشان برابر با مرتبه‌ی  $X$  باشد.

مراحل انجام PLS با الگوریتم NIPALS به شرح زیر می‌باشد:

۱- اولین سطر یا ستون ماتریس  $Y$  را به عنوان بار ( $u$ ) در نظر گرفته و سپس طبق معادله‌ی (۹.۱) آن را نرمال می‌کنیم. هدف از این مرحله، قرار دادن داده‌ها در یک بازه می‌باشد.

$$u_1 = \frac{u}{\|u\|} \quad (9.1)$$

با ضرب  $u$  نرمال شده در  $X$  طبق معادله‌ی (۱۰.۱) ماتریس بار  $X$  به دست می‌آید.

$$w_{1(n \times 1)} = X^T_{(n \times m)} \times u_{1(m \times 1)} \quad (10.1)$$

۲- در مرحله‌ی بعد اولین بردار ماتریس امتیاز  $X$  طبق معادله‌ی زیر حاصل می‌شود.

$$t_{1(m \times 1)} = X_{(m \times n)} \times w_{1(n \times 1)} \quad (11.1)$$

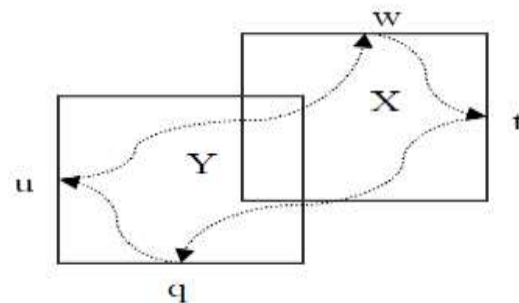
۳- در مرحله‌ی سوم، اولین بردار ماتریس بار  $Y$ ، با استفاده از وارد کردن ماتریس امتیاز  $X$  در فضای ماتریس  $Y$  حاصل می‌شود.

$$q_{1(m \times 1)} = Y^T_{(m \times 1)} \times t_{1(m \times 1)} \quad (12.1)$$

۴- در مرحله‌ی آخر، ماتریس امتیاز  $Y$  طبق رابطه‌ی زیر حاصل می‌شود.

$$u_{1(m \times 1)} = Y_{(m \times p)} \times q_{1(p \times 1)} \quad (13.1)$$

این چهار مرحله را می‌توان به صورت شماتیک در شکل زیر نشان داد.



فاکتور بار برای ماتریس  $X$  به این صورت محاسبه می‌شود.

$$p_{(n \times 1)} = X^T_{(n \times m)} \times t_{(m \times 1)} \quad (14.1)$$

سپس با فاکتورهای به دست آمده، هر ماتریس بازسازی و از ماتریس اصلی کم می شود.

$$E_{(m \times n)} = X_{(m \times n)} - t_{(m \times 1)} \times p_{(1 \times n)}^T \quad (15.1)$$

$$F_{(m \times p)} = Y_{(m \times p)} - t_{(m \times 1)} \times q_{(1 \times p)}^T \quad (16.1)$$

اگر ماتریس باقیمانده در حد نویز باشد، تعداد فاکتورهای به دست آمده درست انتخاب شده است. در غیر این صورت، ماتریس های باقیمانده دوباره وارد چرخه ای چهار مرحله ای می شوند و این کار تا جایی ادامه پیدا خواهد کرد که ماتریس باقیمانده در حد نویز شود.

برای تعیین تعداد صحیح فاکتورها، از دو روش ارزیابی<sup>۱</sup> و ارزیابی<sup>۲</sup> - متقاطع استفاده می شود. در روش ارزیابی، یک مجموعه داده با غلظت و سیگنال مشخص به نام سری ارزیابی وجود دارد. مدل کالیبراسیون برای داده های سری کالیبراسیون هر بار با تعداد معینی فاکتور از یک تا تعداد فاکتور بیشتری ساخته می شود و در هر کدام از مدل های ساخته شده به ازای تعداد فاکتورهای به کار برده شده، غلظت سری ارزیابی پیش بینی می شود و یک سری غلظت برای این مجموعه بر اساس این تعداد فاکتور پیش بینی می شود. این غلظت های به دست آمده، بر اساس رابطه ی مجموع مربع خطای باقیمانده<sup>۳</sup> (PRESS)، با غلظت های واقعی سری ارزیابی مقایسه می شود.

$$PRESS = \sum (\hat{y}_i - y_i)^2 \quad (17.1)$$

که در این معادله،  $\hat{y}_i$  غلظت پیش بینی شده و  $y_i$ ، غلظت واقعی برای سری ارزیابی می باشد. سپس مجموع مربع خطاهای حاصل (PRESS)، بر حسب تعداد فاکتورها رسم، و کمترین مقدار PRESS که نشان دهنده ی کمترین خطا در پیش بینی متغیرهای وابسته است، به عنوان تعداد فاکتور بهینه انتخاب می گردد. در روش ارزیابی - متقاطع، از همان مجموعه داده ی کالیبراسیون برای تعیین تعداد صحیح فاکتورها استفاده می شود. به این صورت که هر بار یک ردیف از ماتریس داده کنار گذاشته می شود و بر اساس مابقی ردیف های داده، مدل کالیبراسیون با یک تعداد معینی فاکتور ساخته می شود. این مدل به دست آمده، برای پیش بینی غلظت ردیف کنار گذاشته شده به کار برده می شود. در ادامه خطای نمونه های پیش بینی شده (PRESS)، بر اساس غلظت پیش بینی شده و غلظت واقعی محاسبه می شود. در مرحله ی بعد، تعداد فاکتورها را تغییر داده و این مراحل تکرار می شود. این عمل برای سایر ردیف ها تکرار شده، و در نهایت مجموع خطای به دست آمده برای هر ردیف محاسبه

<sup>1</sup> Validation

<sup>2</sup> Cross-validation

<sup>3</sup> Predicted Residual Error Sum Of Square (PRESS)

شده، و سپس مانند روش قبل، مجموع مربع خطاهای حاصل (PRESS) بر حسب تعداد فاکتورها رسم، و کمترین مقدار PRESS که نشان‌دهنده‌ی کمترین خطا در پیش‌بینی متغیرهای وابسته است را به عنوان تعداد فاکتور بهینه انتخاب می‌کنیم.

در ادامه، از تعداد صحیح فاکتورهای به‌دست آمده از PRESS، ماتریس‌های مستقل و وابسته بازسازی شده، و در نهایت ضریب رگرسیون طبق معادله‌ی زیر و مانند روش حداقل مربعات به‌دست می‌آید.

$$(t_{(m \times 1)} P^T_{(1 \times n)})^+ \times (t_{(m \times 1)} Q^T_{(1 \times p)}) = B_{PLS} \quad (18.1)$$

روش PLS به دو صورت PLS<sub>1</sub> و PLS<sub>2</sub> به کار برده می‌شود. در PLS<sub>1</sub> بردارهای غلظت برای هر کدام از گونه‌ها به‌صورت جداگانه مدل می‌شود. اما در PLS<sub>2</sub> بردارهای غلظت برای همه‌ی گونه‌ها با هم مدل می‌شوند.

## ۱-۳-۲ شبکه‌های عصبی مصنوعی

### ۱-۳-۲-۱ شبکه‌ی عصبی زیستی

این شبکه‌ها مجموعه‌ای بسیار منظم از پردازشگرهایی موازی به‌نام نورون می‌باشند که به صورت هماهنگ برای حل مسئله عمل می‌کنند، و توسط سیناپس‌ها<sup>۱</sup> (ارتباط‌های الکترومغناطیس) اطلاعات را منتقل می‌کنند. نورون‌ها از سه قسمت اصلی جسم سلولی<sup>۲</sup>، آکسون<sup>۳</sup> و دندریت<sup>۴</sup> تشکیل می‌شوند (شکل ۱-۳). دندریت‌ها سیگنال‌های الکتریکی را به هسته‌ی سلول منتقل می‌کنند. آکسون طول بیشتری دارد و سیگنال الکتروشیمیایی دریافتی از هسته‌ی سلول را به نورون دیگر منتقل می‌کند. نقاط اتصال بین سلول‌های عصبی مختلف که از طریق آنها پیام‌ها منتشر می‌شوند را سیناپس گوئیم. راندمان انتشار پیام‌ها از سلولی به سلول دیگر از طریق فاکتورهای زیادی (شیمیایی و الکتریکی)، تحت تأثیر قرار می‌گیرد، که به آن قدرت سیناپسی گفته می‌شود [۷].

<sup>1</sup> Synapse

<sup>2</sup> Cell body

<sup>3</sup> Axon

<sup>4</sup> Dendrites