اهمیت انتخاب مشخصه در کاربرد هوش مصنوعی در برآورد ژرفای تاقدیس توسط داده‌های گرانی

محمداحسان حکمتیان [۱*]، وحید ابراهیم‌زاده اردستانی [۲]، محمدعلی ریاحی [۲]، ایوب معمار کوچه‌باغ [۳] و جلال امینی [۴]

[۱] دانشجوی دکترا، دانشکده علوم پایه، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران
[۲] دانشیار، مؤسسه ژئوفیزیک دانشگاه تهران، تهران، ایران
[۳] استادیار، پژوهشکده چرخه سوخت هسته‌ای، پژوهشگاه علوم و فنون هسته‌ای، تهران، ایران
[۴] دانشیار، دانشکده فنی، دانشگاه تهران، تهران، ایران

تاریخ دریافت: ۲۰/ ۰۲/ ۱۳۹۱ تاریخ پذیرش: ۱۹/ ۱۲/ ۱۳۹۱

چکیده

استفاده از هوش مصنوعی یا شناسایی الگو ((Pattern Recognition (PR) در تفسیر داده‌های ژئوفیزیکی و یا دیگر علوم زمین چند سالی است که به‌ویژه با ورود شبکه عصبی در این عرصه مطرح شده است. در روش شبکه عصبی و دیگر روش‌های شناسایی الگو و از جمله در روش رده‌بندی کننده بردار حمایت‌شده ((Support Vector Classifier (SVC) که در این پژوهش از آن استفاده شده است، با استفاده از مقادیر مشخصه‌های استخراج‌شده از ساختارها یا اشیای مورد نظر به رده‌بندی آن اشیا یا ساختارها پرداخته می‌شود و به این ترتیب می‌توان به تفسیر مورد نظر دست یافت. به‌طور مرسوم درانتخاب این مشخصه‌ها نظر شخصی نقش اصلی را بر عهده داشته است. در این نوشتار با ارائه نرم‌افزاری که دارای قابلیت انتخاب مشخصه‌های ((Features Selection (FS) مناسب برای برآورد ژرفای ساختارهای طاقدیسی توسط داده‌های گرانی است و نشان دادن تفاوت استفاده از مشخصه‌های مناسب و نامناسب در این تفسیر، اهمیت انتخاب مشخصه در کاربرد هوش مصنوعی در تفسیر گرانی و دیگر علوم زمین نشان داده می‌شود. در اجرای این پژوهش، برای تربیت روش رده‌بندی کننده بردار حمایت‌شده مربوط، ۲۰ نیمرخ مصنوعی گرانی با منبع تاقدیسی تولید شد (مجموعه تربیت کننده). همچنین برای امتحان روش رده‌بندی کننده بردار حمایت‌شده یادشده، ۲۰ نیمرخ مصنوعی گرانی دیگر هم که آنها نیز دارای منابع به شکل تاقدیس بودند، تولید شد (مجموعه امتحان کننده). در این پژوهش نشان داده شد که برآورد ژرفای منابع تاقدیسی در حالت استفاده از مقادیر مشخصه‌های مناسب، خیلی دقیق‌تر از برآورد ژرفا در حالت استفاده از مقادیر مشخصه‌های نامناسب است. لازم به یادآوری است که نرم‌افزار انتخاب مشخصه یادشده به‌طور ویژه تنها برای انتخاب مشخصه تفسیر گرانی طراحی نشده، بلکه طراحی آن به‌گونه‌ای است که قابل استفاده برای انتخاب مشخصه در هر فعالیت تفسیری علوم زمین هم است.

کلیدواژه ها: انتخاب مشخصه، شناسایی الگو، برآورد ژرفای تاقدیس،گرانی، رده‌بندی کننده بردار حمایت‌شده.

***نویسنده مسئول:** محمداحسان حکمتیان
E-mail: mhekmatian@aeoi.org.ir

۱- پیش‌گفتار

نیمرخ‌های گرانی که به‌وسیله تاقدیس‌های زیرزمینی ایجاد می‌شوند وابسته به متغیرهای مختلف منبع، مانند ژرفای آن، الگوهای مختلفی از خود نشان می‌دهند و هر کدام از این الگوها را می‌توان یک رده در نظر گرفت و با درنظر گرفتن همین رده‌های مختلف رده‌بندی کرد. بنابراین با استفاده از یکی از الگوریتم‌های شناسایی الگو مانند شبکه عصبی یا الگوریتم رده‌بندی کننده بردار حمایت‌شده می‌توان به شناسایی این الگوها پرداخت و آنها را رده‌بندی کرد و بدین وسیله به برآورد متغیرهای آنها همچون ژرفای یک تاقدیس زیرزمینی پرداخت. این فعالیت‌های شناسایی الگو، قابل استفاده در انواع پژوهش‌های علوم زمین هم است. ولی مسئله اصلی مطرح در این پژوهش این است که در شناسایی این الگوها، مشخصه‌هایی که استفاده می‌شوند باید به شیوه مناسبی انتخاب شوند. در این مقاله نشان داده شده است که یکی از روش‌های مناسب انتخاب مشخصه، الگوریتم "نسبت فاصله میان رده‌ای به فاصله درون رده‌ای" ((Inter/Intra Class Distance (IICD) است (Heijden et al., 2004). در این راستا یک نرم‌افزار هم که در محیط Visual Basic 6 طراحی شده ارائه شده است.

همچنین بر پایه اطلاعات موجود، تاکنون برای برآورد ژرفای تاقدیس‌های زیر زمینی با داده‌های گرانی از الگوریتم رده‌بندی کننده بردار حمایت‌شده استفاده نشده است. بنابراین این موضوع هم یکی از موضوعات جدید مطرح‌شده در این مقاله می‌تواند در نظر گرفته شود. در مورد الگوریتم رده‌بندی کننده بردار حمایت‌شده در تئوری‌ها و الگوریتم‌ها، توضیحاتی ارائه خواهد شد.

ولی در اینجا به عنوان مقدمه باید گفت: هنگامی که اشیا (Objects) متفاوت هستند و این اشیا (در کار ما نیمرخ‌های گرانی هر کدام یک شئ محسوب می‌شوند.) الگوهای متفاوتی دارند (یا به بیان بهتر به رده‌های متفاوتی تعلق دارند) در این شرایط،" شناسایی الگو"، رده‌های مربوط به این اشیا را بر اساس بعضی از الگوریتم‌های ویژه پیدا می‌کند که این کار رده‌بندی (Classification) نامیده می‌شود. به بیان دیگر شناسایی الگو را می‌توان به عنوان جستجو در اشیا و یافتن الگوهای مختلف در آنها معنا کرد. در شناسایی الگو هر شی با مقادیر مربوط به مشخصه‌هایش شناخته می‌شود و همچنین هر رده هم با مجموعه مقادیر معینی از مشخصه‌ها شناخته می‌شود. رده‌بندی یا تشخیص رده‌ها در شناسایی الگو به دو صورت کلی می‌تواند انجام شود یا با یادگیری غیر نظارتی که در این حالت رده‌بندی از پیش شناخته شده‌ای وجود ندارد و بنابراین یادگیری با فرض وجود رابطه آماری میان مقادیر مشخصه‌ها صورت می‌گیرد و یا با یادگیری تحت نظارت که در این حالت بر پایه رده‌بندی از پیش شناخته شده، مقادیر مشخصه‌های یک مجموعه از اشیا رده‌بندی شده را به عنوان یک مجموعه آموزش دهنده به روش (یا الگوریتم) رده‌بندی کننده شناسایی الگومورد استفاده معرفی می‌شود تا با استفاده از آن به روش رده‌بندی آموزش داده شود که چگونه اشیا رده‌بندی نشده را بتواند رده‌بندی کند. درگرانی، با توجه به امکان پذیر بودن تولید نیمرخ‌های مصنوعی، ایجاد مجموعه آموزش دهنده برای تربیت کردن الگوریتم رده‌بندی کننده (در اینجا SVC) امکان‌پذیر است و بنابراین کاملاً منطقی است که از الگوریتم‌های با یادگیری تحت نظارت استفاده شود. رده‌بندی کننده بردار حمایت‌شده یا همان SVC یک نوع رده‌بندی کننده با یادگیری تحت نظارت است.

۲- تئوری‌ها و الگوریتم‌ها

۲- ۱. الگوریتم رده‌بندی کننده بردار حمایت‌شده (SVC)

پیش از توضیح درباره الگوریتم یادشده باید بیان شود که رده‌بندی کننده بردار حمایت‌شده از مفهومی در دانش رایانه گرفته شده است که "ماشین بردار حمایت شده" (Support Vector Machine (SVM)) نامیده می‌شود و آن همان‌گونه که بیان شد یکی از مفاهیم در دانش رایانه است که برای رده‌بندی و تحلیل "برگشتی" (Regression) استفاده می‌شود. وقتی SVM برای رده‌بندی استفاده شود به آن SVC گفته می‌شود. در ادامه SVC در دو حالت خطی وغیر خطی توضیح داده می‌شود.

در توضیح درباره الگوریتم SVC به شکل خطی باید گفت: در این حالت لازم است که دو رده از اشیا با یک مرز خطی قابل جدایش باشند. یک رده در سمت چپ خط مرزی و دیگری در سمت راست آن قرار می‌گیرد. ولی باید این را یادآور شد که در SVC مرز میان دو رده یک خط نیست بلکه نواری است که خط مرزی یادشده در مرکز پهنای آن قرار دارد. در این شرایط اگر مقادیر مشخصه‌های مورد نظر همه اشیا در مجموعه الگوهای آموزش دهنده در ماتریس z قرار داشته باشند، تابع جدایش مرز خطی در فضای مشخصه‌های (فضایی که هر یک محورهای مختصات آن نشان‌دهنده مقادیر یکی از مشخصه‌هاست) مورد نظر به‌صورت زیر است (Heijden et al., 2004):

$$g(z_n)= w^* z_n + b \qquad (۱$$

در رابطه (۱)، w بردار ضرایب و b یک مقدار اسکالر است. بردار z_n مقادیر مشخصه‌های شی nام (در این تحقیق هر شی یک نیمرخ گرانی است) است. همینطور $g(z_n)$ هم یک مقدار اسکالر است که مقدار تابع جدایش دو رده موجود در مجموعه تربیت کننده در فضای مشخصه‌ها را نشان می‌دهد (و همان‌گونه که در شکل (۱) دیده می‌شود مقادیر منهای یک و یک را با کمک بردارهای حمایت‌شده در SVC به خود می‌گیرد). دو رده جدا شونده به‌وسیله SVC در حالت استفاده از تابع جدایش خطی بر پایه مرجع (Heijden et al., 2004) به‌صورت دو رابطه زیر (برای همه n ها) مشخص می‌شوند:

$$w^T z_n + b \geq 1 \quad \text{if} \quad C_n = +1 \qquad (۲$$

$$w^T z_n + b \leq -1 \quad \text{if} \quad C_n = -1 \qquad (۳$$

در SVC برای جدایش دو رده، به یکی از رده‌ها عدد یک و به رده دیگر عدد منهای یک را اختصاص می‌دهند واین عددهای یک و منهای یک در واقع برچسب‌هایی هستند برای هر یک از دو رده موجود در مجموعه تربیت کننده. در روابط (۲) و(۳)، Cn عدد اسکالری است که این برچسب‌ها را نشان می‌دهد. همچنین wT به معنای ترانسپوز w است. در این حالت پهنای نوار جدایش دو رده را می توان به‌صورت شکل ۱ نمایش داد.

گرادیان $g(z_n)$ نسبت به z_n عبارتست از w و بنابراین مربع پهنای نوار جدایش متناسب با عکس توان دوم w یعنی $(\| w \|^2 = w^T w)$ است. دلیل وجود این تناسب بحث مفصلی است ولی به همین نکته اکتفا می‌شود که وقتی گرادیان $g(z_n)$ نسبت به z_n برابر w شود در این صورت داریم: w=(تغییرات z_n / تغییرات $g(z_n)$) در نتیجه تغییرات z_n با عکس w (و همچنین با عکس توان دوم w) متناسب می‌شود. در این حال باید گفت که پهنای نوار جدایش که فاصله‌ای در فضای مشخصه‌هاست که با تغییرات z_n (مقادیر مشخصه‌های شئ n ام) متناسب است باید متناسب با عکس توان دوم w باشد. بنابراین برای بیشینه کردن پهنای نوار جدایش، رابطه متغیر L شامل ضرایب لاگرانژ (α_n) زیر نوشته می‌شود:

$$L = \frac{1}{2}\| w \|^2 + \sum_{n=1}^{Ns} \propto_n \left(C_n \left[w^T z_n + b\right] - 1\right), \propto_n \geq 0 \qquad (۴$$

(Heijden et al., 2004)

مشتق L نسبت به w و b مساوی صفر قرار داده می‌شود و با این کار دو رابطه زیر به دست می‌آید که شرط بیشینه شدن پهنای نوار جدایش هستند:

$$w = \sum_{n=1}^{Ns} \propto_n C_n z_n \qquad (۵$$

(Heijden et al., 2004)

$$w = \sum_{n=1}^{Ns} C_n \propto_n = 0 \qquad (۶$$

(Heijden et al., 2004)

بر پایه دو رابطه بالا رابطه متغیر L به‌صورت زیر در می‌آید:

$$L = \sum_{n=1}^{Ns} \propto_n - \frac{1}{2} \sum_{n=1}^{Ns} \sum_{n=1}^{Ns} C_n C_n \propto_n \propto_m Z_n^T Z_m, \quad \propto_n \geq 0 \qquad (۷$$

(Heijden et al., 2004)

L باید نسبت به ضرایب لاگرانژ بیشینه شود. این یک مسئله بهینه‌سازی است. پس از بهینه‌سازی با استفاده از ضرایب لاگرانژ مقدار w مناسب و با داشتن w، رده‌بندی کننده بهینه مناسب به‌دست می‌آید. بسیاری از ضرایب لاگرانژ صفر می‌شوند و بنابراین تنها z_nهای مربوط به ضرایب لاگرانژ غیر صفر در به‌دست آوردن رده‌بندی کننده مفید هستند که به آنها "بردارهای حمایت شده" گفته می‌شود. تا به اینجا SVC خطی بیان شد.

برای وضوح بیشتر مطلب باید گفت که فرض می‌شود یک رده نیمرخ‌های گرانی با منابع گرانی به شکل تاقدیس با ژرفای بالایی ۱۰ متر (نقاط توخالی در شکل ۲) و یک رده دیگر نیمرخ‌های گرانی با منابع گرانی به شکل تاقدیس با ژرفای بالایی ۲۰۰ متر (نقاط توپر در شکل ۲) باشد. اگر از جدایشگر خطی برای جدایش این ۲ رده بهره برده شود، در این صورت مرز جدایش میان دو رده را نمی‌توان به‌طور یکتا تعیین کرد و هر یک از خطوط مرزی نشان داده شده در شکل ۲ در میان ۲ رده می‌تواند مرز جدایش رده‌ها بر پایه الگوریتم جدایشگر خطی باشد.

ولی SVC برای جدایش همین ۲ رده بالا به یک نوار جدایش یکتا مانند شکل ۳ می‌انجامد.

برای وضوح بیشتر مطلب بهتر است بیان شود که بیشینه پهنای نوار جدایش مفهوما دارای دو مزیت است، اول اینکه وقتی میان ۲ رده متفاوت به دنبال مرزی باشیم که در مرکز نوار جدایش دارای بیشینه پهنا بین ۲ رده باشد، در این صورت چنین مرزی تنها می‌تواند یک مرز باشد و نیز چنین مرزی بهترین مرز جدایش است چرا که از هر ۲ رده بیشترین فاصله ممکن را دارد. این مزیت‌ها، مزیت‌های هوشمندی است و برای جدایش رده‌ها در ژئو فیزیک و یا دیگر شاخه‌های علم یکسان است.

در SVC مرز میان رده‌ها می‌تواند نواری غیر خطی باشد که روندی مشابه حالت بالا دارد. در این حالت رده‌بندی کننده خطی مورد استفاده که یک تابع جدایش خطی بود به یک تابع جدایش غیر خطی تبدیل می‌شود؛ به این صورت که به‌جای z_n از $y(z_n)$ استفاده می‌شود. در این حالت تابع جدایش به‌صورت زیر در می‌آید:

$$g(z)= w^T y(z_n)= \sum_{n=1}^{Ns} K\left(z_n, z_m\right) \qquad (۸$$

(Heijden et al., 2004)

$K(z_n,z_m)$ تابع کرنل نامیده می‌شود. کرنل غیر خطی که کرنل مناسبی برای جدایش ژرفای منابع تاقدیسی مولد نیمرخ‌های گرانی است «کرنل مینکوسکی» است. رابطه این کرنل به قرار زیر است:

$$K\left(z_n, z_m\right) = \left(\sum_{n=1}^{Ns} z_n - z_m |^P\right)^{\frac{1}{P}} \qquad (۹)$$

(Duin et al., 2007)

در رابطه بالا p توان اختلاف مقدار مشخصه‌های ۲ رده است. در این کرنل مرز جدایش دو رده بر پایه اختلاف مقدار مشخصه‌های آن ۲ رده مشخص می‌شود. و بنابراین این کرنل در زمانی که دو رده مورد نظر بر پایه بزرگی وکوچکی از هم جدا می‌شوند بسیار مناسب است چرا که مرز جدایش دو مقدار بزرگ وکوچک بر پایه اختلاف آنهاست. بنابراین این کرنل برای جدایش ژرفای زیاد و ژرفای کم مربوط به منابع تاقدیسی مولد گرانی مفید به نظر می‌رسد و همین گونه هم بود. مقدار به کار رفته برای P برابر یک در نظر گرفته شد.

۲– ۲. مشخصه‌ها و الگوریتم انتخاب مشخصه‌ها(FS)

مشخصه نوعی از اندازه‌گیری است که از یک شی به دست می‌آید و همان‌گونه که پیش‌تر گفته شد در این پژوهش شی به معنی هر نیمرخ گرانی است. مقادیر مشخصه‌های استخراج شده از "نیمرخ‌های گرانی اصلی" می‌توانند برای برآورد شکل منبع گرانی مولد آن نیمرخ گرانی مورد استفاده قرار گیرد. بر پایه Gret & Klingele (1998) نیمرخ گرانی اصلی عبارت از نیمرخ گرانی است که شامل مقدار گرانی بیشینه تولیدشده به‌وسیله منبع گرانی باشد و خطوط بی‌هنجاری را به‌صورت عمودی قطع کند. چنین نیمرخی در کامل‌ترین شکل، مشخصات ایجادشده به‌وسیله منبع گرانی در نیمرخ گرانی را از خود نشان می‌دهد. بنابراین استفاده از نیمرخ‌های دیگر خطای برآورد را بالا می‌برد. از این پس هر کجا نیمرخ گرانی مطرح شود منظور نیمرخ گرانی اصلی است. برخی از انواع مشخصه‌ها که می‌توانند دربرآورد شکل منبع گرانی مورد استفاده قرار گیرند در زیر آمده است :

$$F1= X_{g75} \quad (۱۰$$

(Gret & Klingele, 1998)

$$F2= X_{g25}$$

درمعادلات بالا، g_y مقدار گرانی است که برابر y درصد از مقدار بیشینه گرانی (g_{max}) است و X_{gy} مقدار X در g_y است. برای درک بهتر به شکل ۴ نگاه شود.

"انتخاب مشخصه‌های (مناسب)" یا FS امر بسیار مهمی است که بر پایه اطلاعات، تاکنون در امر برآورد ژرفای تاقدیس توسط داده‌های گرانی بلکه در همه علوم زمین استفاده نشده است. برای نمونه در Gret & Klingele (1998) که از شبکه عصبی در تفسیر گرانی بهره برده‌اند کاملاً با صراحت بیان شده که انتخاب مشخصه‌ها در پژوهش آنها کاملاً بر پایه نظر شخصی بوده است.

اگر مشخصه‌های مناسبی در برآورد ژرفای منابع تاقدیسی استفاده نشود، شاید سبب خطای زیادی در برآورد شود. در این پژوهش مشخصه‌های مناسب نه با نظر شخصی بلکه با الگوریتمی مناسب و با طراحی نرم‌افزار مناسب مربوط به‌دست آمده است. یکی از الگوریتم‌های مناسبی که در این زمینه قابل استفاده است، الگوریتم " نسبت فاصله میان‌رده‌ای به فاصله درون‌رده‌ای" یا IICD است (Heijden et al., 2004) که از همین الگوریتم در انتخاب مشخصه‌های مناسب در برآورد ژرفای منابع تاقدیسی بهره برده شده است.

این الگوریتم برای انتخاب مشخصه‌های مناسب، مقدار یک متغیر را استفاده می‌کند. به این معنی که در انتخاب مشخصه‌های مناسب، آن مشخصه‌هایی را انتخاب می‌کند که مقدار متغیر مزبور برای آنها بزرگ‌تر باشد. این متغیر، "اندازه انجام" (Performance Measure) یا PM نامیده می‌شود. PM مربوط به الگوریتم IICD بر پایه فاصله اقلیدسی بین هر دو شی در مجموعه آموزش دهنده در فضای مشخصه‌هاست. به‌طور کلی در این الگوریتم هر چه نسبت فاصله متوسط میان اشیا متعلق به رده‌های متفاوت به فاصله متوسط میان اشیا متعلق به یک رده در یک فضای ویژه مشخصه‌ها بیشتر باشد، مشخصه‌های آن فضای ویژه برای جدا کردن رده‌ها مناسب‌تر تشخیص داده می‌شود و PM بزرگ‌تری به آنها اختصاص می‌یابد (Heijden et al., 2004).

در اینجا لازم است رابطه نشان‌دهنده تعداد زیر مجموعه‌های شامل D مشخصه که از میان N مشخصه قابل انتخاب شدن هستند بیان شود:

$$q(D)=\frac{N!}{(N-D)!D!} \quad (۱۱$$

(Heijden et al., 2004)

این رابطه نشان می‌دهد که برای نمونه در میان رده مشخصه، ۴۵ زیر مجموعه ۲ مشخصه‌ای را می‌توان به‌دست آورد. در این صورت اگر الگوریتم IICD اعمال شود، مشخص خواهد شد که در میان ۴۵ حالت ۲ مشخصه‌ای کدامیک مناسب‌ترین ۲ مشخصه برای رده‌بندی است.

۳– روش کار

۳– ۱. تولید نیمرخ‌های مصنوعی گرانی

در اجرای این پژوهش، ۲۰ نیمرخ مصنوعی گرانی برای استفاده در تربیت کدهای نرم‌افزاری SVC مربوطه تولید شد (مجموعه تربیت کننده). همچنین برای امتحان کدهای نرم‌افزاری SVC یادشده، ۲۰ نیمرخ گرانی مصنوعی دیگر تولید شد (مجموعه امتحان کننده). پیش از توضیح بیشتر درباره نیمرخ‌های یادشده باید یادآوری شود که کدهای نرم‌افزاری SVC در محیط MATLAB و PRTOOLS (Duin et al., 2007) تهیه شده است و الگوریتم مورد استفاده برای تولید نیمرخ مصنوعی گرانی با منابع گرانی به شکل تاقدیس از Blakely (1996) استخراج شده است. همچنین منابع گرانی استفاده شده در این پژوهش با فرض گسترش بی‌نهایت (بی‌نهایت فیزیکی) از دو سو با مقطع یکسان در نظر گرفته شده‌اند و بنابراین به قولی منابع دو و نیم بعدی به شمار می‌روند. در واقع از روش‌های مستقیم و از روش Hubbert (1948) و Talwani et al. (1959) مربوط به فصل نهم مرجع یادشده به عنوان الگوریتم بهره گرفته شد.

در ارتباط با ژرفا و ابعاد منابع تاقدیسی مولد نیمرخ‌های گرانی مصنوعی یادشده باید گفت که با توجه به شکل ۵، منابع تاقدیسی یادشده دارای ژرفای بالایی یا کمینه D، گسترش عمودی V و گسترش افقی H هستند. باید گفت که همه نیمرخ‌های گرانی مصنوعی به‌گونه‌ای تولید شده اند که مرکز همه آنها در محل تصویر مرکز تاقدیس‌ها در سطح زمین (نقطه A در شکل ۵) است. تفاوت چگالی منابع با زمینه برابر ۱ gr/cm^3 انتخاب شد.

۳ متغیر یادشده در منابع مربوط به نیمرخ‌های گرانی مصنوعی مجموعه تربیت کننده کدهای نرم‌افزاری SVC در جدول ۱ و در مورد منابع نیمرخ‌های گرانی مصنوعی مجموعه امتحان کننده کدهای یادشده در جدول ۲ آمده است.

۳– ۲. استخراج مشخصه‌ها

از هر یک از ۲۰ نیمرخ گرانی مصنوعی مجموعه تربیت کننده، ۱۰ مشخصه استخراج شد که این مشخصه‌ها در جدول ۳ آمده‌اند.

۱۰ مشخصه یادشده، مشخصه‌های مرجع مناسبی برای برای انتخاب مشخصه برای برآورد ژرفای تاقدیس هستند، چرا که این مشخصه‌ها مربوط به ژرفای منبع تاقدیسی و شامل همه X_{gy} از $X_{g۵۰}$ تا $X_{g۹۰}$ به‌صورت سیستماتیک هستند. مشخصه‌های جدول ۳ شامل X_{gy} از $X_{g۱۰}$ تا $X_{g۹۰}$ به این دلیل از $X_{g۱۰}$ شروع شده است که احتمال اینکه نیمرخ گرانی واقعی برداشت‌شده در صحرا شامل مقدار ۵۰ درصدی بیشینه خود باشد زیاد است و در نتیجه برای X_{g50} اکثراً مقدار وجود دارد ولی معمولاً آنقدر نیمرخ را طولانی برداشت نمی‌کنند که شامل مقدار ۱۰ درصدی بیشینه خود باشد.

همچنین این مشخصه‌ها شامل ۲ مشخصه مهم g_{max} و انتگرال صد متر مرکزی نیمرخ گرانی هم هستند. هر نیمرخ گرانی از نقطه ابتدا تا نقطه انتهایی آن دارای یک طول است (مثلاً ۲۰۰ متر). در این حال انتگرال صد متر مرکزی نیمرخ گرانی به معنی انتگرال نیمرخ است ولی نه از ابتدا تا انتهای آن، بلکه انتگرال مربوط به صد متر مرکزی از طول نیمرخ.

۳– ۳. انجام عملیات انتخاب مشخصه و تربیت کدهای نرم‌افزاری SVC برآورد کننده ژرفا

از میان ۱۰ مشخصه یادشده در بخش ۳–۲، بر پایه الگوریتم IICD (مراجعه به بخش ۲–۲)، مناسب‌ترین ۲ مشخصه برای برآورد ژرفای کمینه منابع تاقدیسی به‌دست آمد. نرم‌افزار Select Features برای انتخاب بهترین دو مشخصه از میان ۱۰ مشخصه در محیط Visual Basic 6 طراحی شده است. با اجرای این نرم‌افزار اولین فرمی که ظاهر می‌شود، فرم معرفی نرم‌افزار و شامل دکمه "Select the most proper 2 features" است (شکل ۶) که پس از زدن دکمه یادشده فایل شامل مقادیر مشخصه‌های اشیا مجموعه تربیتی (10features.txt) به نرم‌افزار معرفی می‌شود.

فایل 10features.txt یا همان فایل شامل مقادیر مشخصه‌های اشیا مجموعه تربیتی، فایل ورودی به نرم‌افزار است و در هر رده خود ۱۰ شی دارد. برای وضوح بیشتر باید گفت همان‌گونه که در جدول ۱ دیده می‌شود، مجموعه تربیتی شامل ۲۰ نیمرخ گرانی مصنوعی است که ۲ رده دارد. یکی از رده‌ها، رده ژرفای کمینه ۱۰ متر و رده دیگر ژرفای کمینه ۲۰۰ متر است. به بیان بهتر ۱۰ عدد از نیمرخ‌ها که دارای منابع تاقدیسی با ژرفای کمینه ۱۰ متر هستند یک رده را تشکیل می‌دهند و ۱۰ عدد دیگر آنها دارای منابع تاقدیسی با ژرفای کمینه ۲۰۰ متر هستند و رده دیگر را به وجود می‌آورند. در فایل 10features.txt در واقع ۲۰ ردیف از اعداد وجود دارد که مقادیر مشخصه‌های اشیا یا همان نیمرخ‌های مجموعه تربیتی هستند (مشخصه‌های یادشده در جدول ۳). در ۱۰ ریف اول مقادیر مشخصه‌های نیمرخ‌های رده ژرفای کمینه ۱۰ متر و در ۱۰ ریف دوم مقادیر مشخصه‌های نیمرخ‌های رده ژرفای کمینه ۲۰۰ متر قرار گرفته است. و به‌طور کلی هر ردیف شامل مقادیر مشخصه‌های مربوط به یکی از نیمرخ‌هاست. فایل یادشده در شکل ۷ نشان داده شده است.

پس از معرفی فایل شامل مقادیر مشخصه‌های اشیا مجموعه تربیتی، نرم‌افزار محاسبه را انجام و بر پایه اندازه متغیر به نام "اندازه انجام" یا PM (مراجعه به بخش (۱-۲))، همه حالات ۲ مشخصه‌ای ممکن از میان ۱۰ مشخصه را تفکیک می‌کند. در اینجا، آن حالت ۲ مشخصه‌ای که PM بزرگ‌تری داشته باشد از نظر آلگویتم انتخاب مشخصه، بهترین حالت ۲ مشخصه‌ای برای رده‌بندی اشیاست. در شکل ۸، PM همه حالات ۲ مشخصه‌ای ممکن از میان ۱۰ مشخصه به توسط نرم‌افزار در یک نمودار نشان داده شده است (همه حالات ۲ مشخصه‌ای ممکن از میان ۱۰ مشخصه بر پایه رابطه ۱۱، ۴۵ حالت است).

همچنین فایل "Performance_measure.txt" که PM همه حالات ۲ مشخصه‌ای ممکن از میان ۱۰ مشخصه مربوط به مجموعه تربیتی ورودی به نرم‌افزار را درخود ذخیره می‌کند، توسط نرم‌افزار یادشده تولید می‌شود. مقادیر این فایل البته به شکل مرتب‌شده از حالات ۲ مشخصه‌ای با PM زیاد به حالات ۲ مشخصه‌ای با PM کم در جدول ۴ نشان داده شده است (بخش اعشاری مقادیر PM گرد شده است).

۳- ۴. امتحان SVC تربیت‌شده برآورد کننده ژرفا

ابتدا کدنرم‌افزاری SVC به‌وسیله مناسب‌ترین ۲ مشخصه استخراج‌شده از ۲۰ نیمرخ گرانی مجموعه تربیت کننده (یعنی مشخصه‌های ۸ و ۹ که بزرگ‌ترین PM را دارند) مورد تربیت قرار گرفت. سپس از ۲۰ نیمرخ گرانی مجموعه امتحان کننده، مناسب‌ترین ۲ مشخصه یادشده استخراج و کد نرم‌افزاری SVC تربیت‌شده نامبرده به‌وسیله آنها مورد امتحان قرار گرفت. نتایج این امتحان مطلوب بوده است و درباره آنها در ادامه توضیح خواهیم داد. همچنین برای اینکه اهمیت انتخاب مشخصه مناسب بیشتر روشن شود، کد نرم‌افزاری SVC به‌وسیله نامناسب‌ترین ۲ مشخصه استخراج‌شده از ۲۰ نیمرخ گرانی مجموعه تربیت کننده (یعنی مشخصه‌های ۹ و ۱۰ که کوچک‌ترین PM را دارند) هم مورد تربیت قرار گرفت. پس از آن، از ۲۰ نیمرخ گرانی مجموعه امتحان کننده، نامناسب‌ترین ۲ مشخصه یادشده را هم استخراج و کد نرم‌افزاری SVC تربیت‌شده دوم به‌وسیله آنها مورد امتحان قرار گرفت که نتایج این امتحان هم در بخش ۳ آمده است و نشان می‌دهد که در صورت استفاده از مقادیر مشخصه‌ها در حالت ۲ مشخصه‌ای نامناسب خطای برآورد ژرفا به مقدار قابل توجهی افزایش می‌یابد.

۳- ۵. انجام رویه کار روی داده‌های واقعی

داده‌های واقعی گرانی که در ادامه نمایش داده می‌شوند، از دید موقعیت جغرافیایی در جنوب ایران و در پیرامون چاهبار برداشت شده‌اند. ناحیه عملیاتی در محدوده ۲۵/۴۰-۲۵/۲۰ عرض جغرافیایی شمالی و ۶۰/۵۵-۵۹/۵۰ طول جغرافیایی خاوری قرار گرفته است. وسعت منطقه عملیاتی ۲۲۰۰ کیلومتر مربع است. در بخش شمالی این ناحیه رشته‌کوه‌های مکران قرار دارد که از مجموعه‌ای از سنگ‌های افیولیتی تشکیل شده است. از دید ساختمانی مشخصه چشمگیر حوزه رسوبی مکران وجود زنجیره‌هایی از ناودیس‌های بیضوی‌شکل خاوری- باختری است که شامل تاقدیس‌هایی هم هست. در این عملیات ۵۶ خط به تعداد ۱۰۳۴ ایستگاه خوانش به فواصل ۲ کیلومتر در ۱ کیلومتر به‌صورت شبکه‌بندی منظم پیاده شده است. برای نقشه‌برداری نقاط از دوربین T2 ویلد و طول یاب DI-20 استفاده شده و برای برداشت گرانی از دستگاه لاکوست رومبرگ مدل G440 با دقت ۰/۰۱ میلی‌گال استفاده شده است. در پایان گرانی بوگه وسپس مقادیر باقیمانده (Residual) آنها محاسبه و به‌صورت نقشه مربوط به شکل ۹ نمایش داده شده است. در این شکل بی‌هنجاری که نیمرخ AB از آن استخراج شده، ظاهراً مربوط به یک تاقدیس زیرزمینی است.

همان‌گونه که در شکل ۱۰ دیده می‌شود، ژرفای احتمالی یال‌های تاقدیس بر پایه روش اویلر بیشتر در حدود ۱۰۰۰ تا ۱۵۰۰ متر برآورد شده است.

با استفاده از نیمرخ AB، به برآورد ژرفای بالایی (ژرفای رأس) تاقدیس با استفاده از SVC و با استفاده از مناسب‌ترین ۲ مشخصه و همچنین با استفاده از نامناسب‌ترین ۲ مشخصه (که در بخش ۳- ۴ بیان شد) پرداخته شد. نتایج یادشده در بخش بحث و نتیجه‌گیری آمده است. البته در اینجا لازم است گفته شود که برای برآورد ژرفای بالایی تاقدیسی با ابعاد چند هزار متری (مانند آنچه در شکل ۹ و یا ۱۰ دیده می‌شود،) لازم است تربیت SVC با کمک منابع مصنوعی با همین ابعاد صورت گیرد که همین کار هم انجام شد. همچنین تفاوت چگالی چنین تاقدیسی با سنگ‌های پیرامون آن باید کوچک باشد که برابر یک دهم گرم بر سانتی‌متر مکعب در نظر گرفته شد.

۴- بحث و نتیجه‌گیری

پیش از ارائه نتایج، باید گفته شود که ژرفای کمینه منابع تاقدیسی به توسط SVC همان طور که بیان شد به کمک کرنل مینکوسکی (رابطه ۹) که متغیر آن مساوی ۱ قرار داده می‌شود، برآورد می‌شود که باید توضیح داده شود که در این راستا، ژرفای کمینه منابع تاقدیسی (با توجه به مجموعه تربیتی استفاده شده) بر پایه رابطه زیر به دست آمده است:

((احتمال داشتن ژرفای کمینه مساوی ۲۰۰ متر که به کمک SVC تربیت شده محاسبه شده است) ضرب در ۲۰۰) + ((احتمال داشتن ژرفای کمینه مساوی ۱۰ متر که به کمک SVC تربیت شده محاسبه شده است) ضرب در ۱۰).

همان طور که در بخش ۳- ۴ عنوان شد، مقادیر مشخصه‌های مربوط به مجموعه امتحان کننده هم در مناسب‌ترین حالت ۲ مشخصه‌ای (حالت ۱ در جدول ۵) و هم در نامناسب‌ترین حالت ۲ مشخصه‌ای (حالت ۲ در جدول ۵) به عنوان ورودی به کد نرم‌افزاری SVC برآورد کننده ژرفا ارائه شد. نتایج برآورد ژرفا در دو حالت ۱ و ۲ در جدول ۵ آمده است (بخش اعشاری ژرفاهای برآوردی و خطاها حذف شده است).

در جدول ۵ دیده شد که برآورد ژرفا در حالت استفاده از مقادیر مشخصه‌های مربوط به مجموعه امتحان کننده در مناسب‌ترین حالت ۲ مشخصه‌ای (حالت ۱ در جدول ۵) خیلی دقیق‌تر از برآورد ژرفا در نامناسب‌ترین حالت ۲ مشخصه‌ای (حالت ۲ در جدول ۵) است. برای نمونه در حالت اول حداکثر خطای برآورد ژرفا ۵۸ درصد ولی در حالت دوم ۸۸۲ درصد است. همچنین میانگین خطای برآورد ژرفا در حالت اول ۱۵ درصد ولی در حالت دوم ۱۰۸ درصد است. بنابراین دیده می‌شود که اهمیت انتخاب مشخصه مناسب چقدر زیاد است.

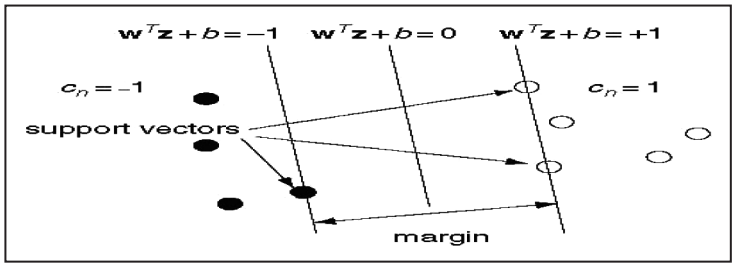
در ارتباط با داده های واقعی باید گفت: نتایج برآورد ژرفای بالایی (ژرفای رأس) تاقدیس با استفاده از SVC و با استفاده از مناسب‌ترین ۲ مشخصه و همچنین با استفاده از نامناسب‌ترین ۲ مشخصه به ترتیب برابر ۷۲۴ و ۳۵۶۶ متر به دست آمد. در ارتباط با تاقدیس یادشده، داده حفاری در دست نیست که بتوان به کمک آن دقت برآورد ژرفا را مشخص کرد ولی در مقایسه با برآورد ژرفای احتمالی یال‌های تاقدیس بر پایه روش اویلر که بیشتر در حدود ۱۰۰۰ تا ۱۵۰۰ متر برآورد شده است، باید گفت

برآورد ژرفای ۷۲۴ متری با استفاده از SVC و با استفاده از مناسب‌ترین ۲ مشخصه منطقی به‌نظر می‌رسد چرا که ژرفای یال‌های تاقدیس اگر ۱۰۰۰ تا ۱۵۰۰ متر باشد، ژرفای رأس آن می‌تواند ۷۲۴ متر باشد. اینجاست که باز انتخاب مشخصه در استفاده از روش‌های هوش مصنوعی، اهمیت خود را نشان می‌دهد.

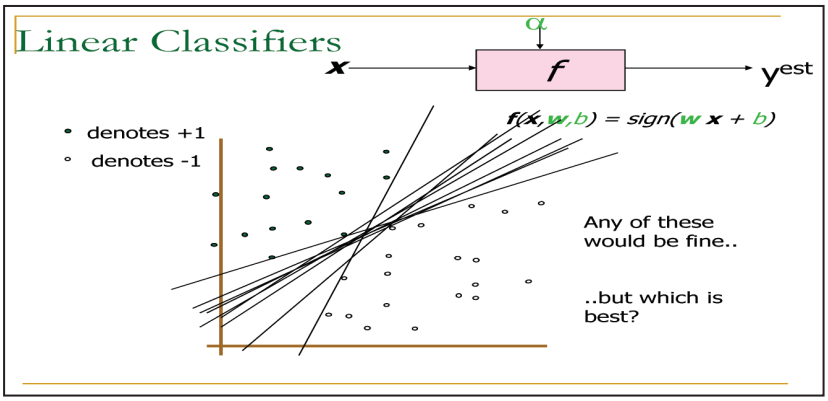
نکته دیگری که بهتر است به آن اشاره شود این است که خود SVC مزایایی در میان روش‌های شناسایی الگو دارد که در بخش ۲- ۱ و در رابطه با شکل‌های ۲ و ۳ مطرح شد. همچنین در مقایسه با روش اویلر باید گفت SVC می‌تواند ژرفای مکان ویژه‌ای از یک منبع زیرزمینی را برآورد کند ولی در روش اویلر اگر ژرفای یک منبع برآورد شود، مشخص نیست که ژرفای سطح بالایی منبع گرانی است یا ژرفایی بیشتر (مثلاً ژرفای میانی). البته برآورد ژرفا با SVC هم دارای خطاست ولی می‌توان آن را برای برآورد ژرفای مکان ویژه‌ای از منبع گرانی تربیت کرد. در هر صورت SVC هم می‌تواند به عنوان یک روش برای برآورد ژرفا استفاده شود.

سپاسگزاری

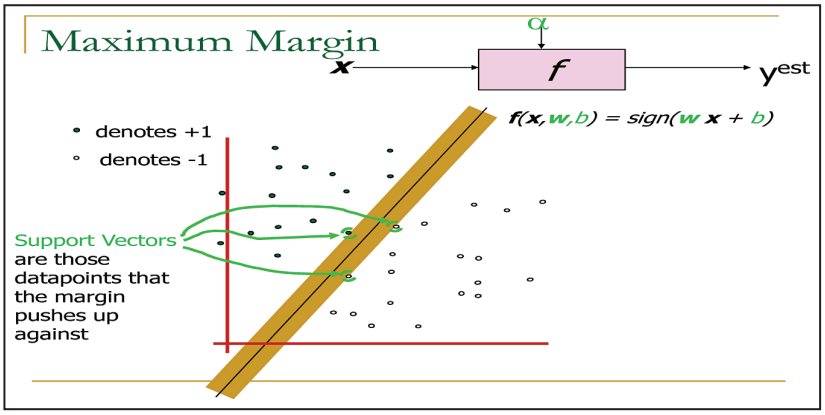
بدین وسیله از اعضای محترم هیئت علمی گروه ژئوفیزیک دانشکده علوم پایه دانشگاه علوم و تحقیقات، به‌ویژه جناب آقای دکتر زمردیان که زمینه لازم برای انجام این پژوهش را فراهم کردند و همچنین از جناب آقای دکتر هاشمی که در ابتدا نگارنده اول را با این زمینه پژوهشی آشنا کردند و مطالعه ۲ مقاله از ایشان انگیزه انجام این پژوهش بوده است (Hashemi, 2010; Hashemi et al., 2008)، سپاسگزاری می‌شود.



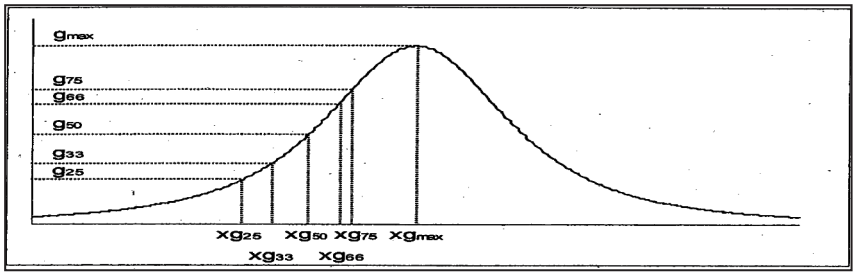
شکل ۱- نوار جدایش دو رده و بردارهای حمایت‌شده (Heijden et al., 2004).



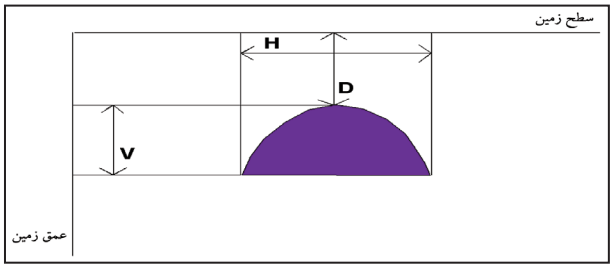
شکل ۲- خطوط جدایش متفاوت میان ۲ رده که هر یک از خطوط جدایش یادشده می‌تواند پاسخ رده‌بندی جدایشگر خطی باشد (Tan, 2004).



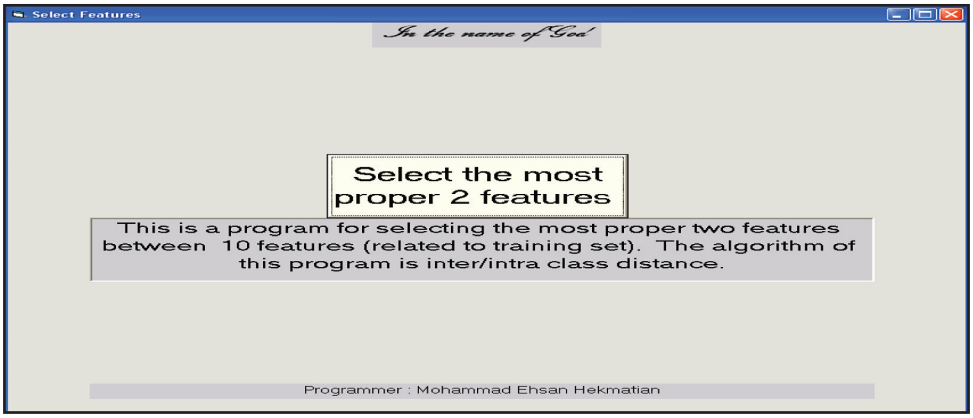
شکل ۳- نوار جدایش میان ۲ رده که به‌طور یکتا میان ۲ رده به‌وسیله SVC مشخص شده است (Tan, 2004).



شکل ۴- در این شکل مقدار گرانی g_y که برابر y درصد از مقدار بیشینه گرانی (g_{max}) است و X_{gy} که مقدار X در g_y است، نشان داده شده‌اند (Gret & Klingele, 1998).



شکل ۵- نمایش ژرفای بالایی یا کمینه D، گسترش عمودی V و گسترش افقی H مربوط به منابع تاقدیسی.

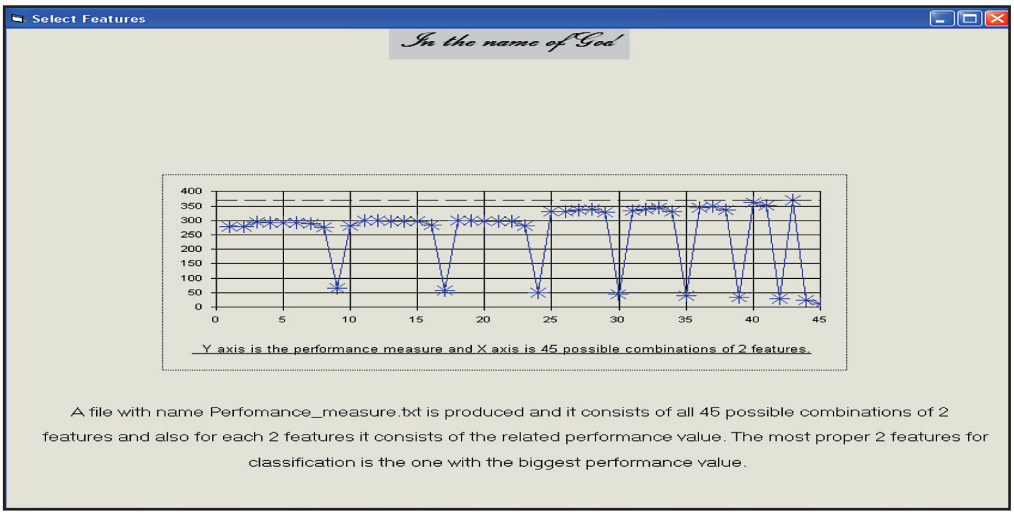


شکل ۶- فرم معرفی نرم‌افزار و شامل دکمه با نام "Select the most proper 2 features".

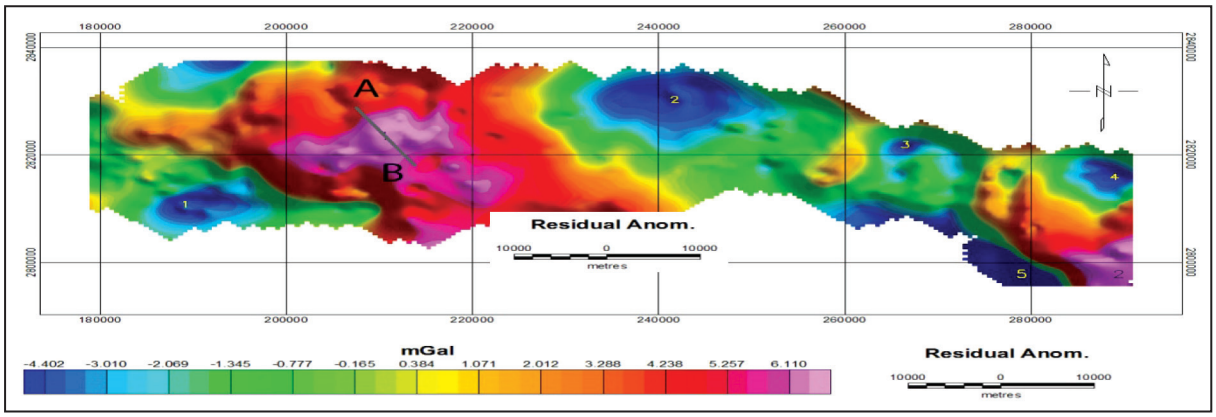
10features

	A	B	C	D	E	F	G	H	I	J	K
1	63	56	49	43	38	32	27	22	1.52508	124.5693	
2	62	55	49	43	37	32	27	22	1.49142	121.299	
3	61	54	48	42	36	31	26	21	1.45362	117.6369	
4	59	53	47	41	36	31	26	21	1.41148	113.5672	
5	58	51	45	40	35	30	25	21	1.364	108.9985	
6	56	50	44	39	34	29	25	20	1.31065	103.8866	
7	54	48	43	38	33	28	24	20	1.2509	98.19123	
8	52	46	41	36	32	27	23	19	1.18386	91.84075	
9	49	44	39	35	30	26	22	18	1.10919	84.82161	
10	47	42	37	33	29	25	21	18	1.02576	77.05409	
11	258	233	211	189	169	149	129	109	0.38153	37.68653	
12	256	232	210	188	168	148	128	108	0.36772	36.31783	
13	255	231	208	187	167	147	128	107	0.35256	34.81465	
14	253	229	207	186	166	146	127	106	0.33605	33.17844	
15	251	227	205	184	165	145	126	106	0.31792	31.38229	
16	249	225	203	183	163	144	125	105	0.29812	29.42124	
17	247	223	201	181	162	143	123	104	0.27664	27.29456	
18	244	221	199	179	160	141	122	103	0.25339	24.9935	
19	241	218	197	177	158	139	121	101	0.22853	22.53471	
20	238	215	194	175	156	137	119	100	0.20205	19.91629	
21											

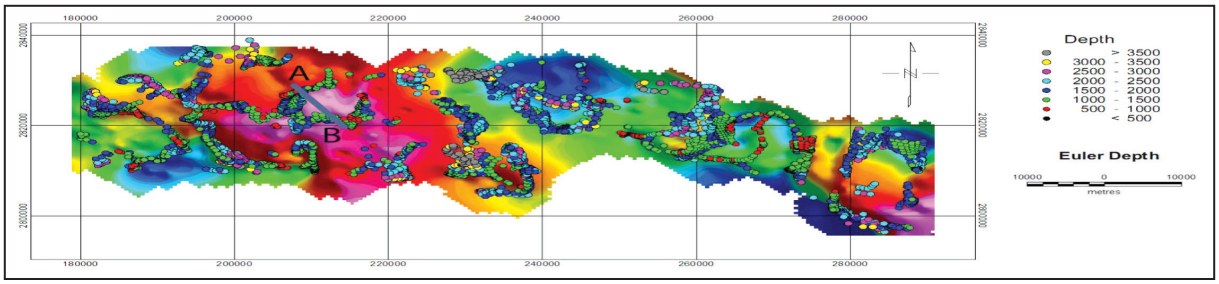
شکل ۷- فایل "10features.txt" یا همان فایل شامل مقادیر مشخصه‌های اشیای مجموعه تربیتی.



شکل ۸- نمودار PM همه حالات ۲ مشخصه‌ای ممکن از میان ۱۰ مشخصه موجود در مجموعه تربیتی استفاده‌شده بر پایه محاسبه نرم‌افزار Select Features.



شکل ۹- نقشه گرانی باقیمانده مربوط به ناحیه‌ای در پیرامون چابهار.



شکل ۱۰- نقشه برآورد ژرفای احتمالی بر پایه روش اویلر با استفاده از داده‌های نقشه شکل ۹.

جدول ۱- ژرفای بالایی یا کمینه D، گسترش عمودی V و گسترش افقی H بر حسب متر مربوط به منابع تاقدیسی مجموعه تربیت کننده.

شماره نیمرخ	D	H	V
۱	۱۰	۱۱۰	۵۸
۲	۱۰	۱۲۰	۶۲
۳	۱۰	۱۳۰	۶۶
۴	۱۰	۱۴۰	۷۰
۵	۱۰	۱۵۰	۷۳
۶	۱۰	۱۶۰	۷۶
۷	۱۰	۱۷۰	۷۸
۸	۱۰	۱۸۰	۸۰
۹	۱۰	۱۹۰	۸۲
۱۰	۱۰	۲۰۰	۸۳
۱۱	۲۰۰	۱۱۰	۵۸
۱۲	۲۰۰	۱۲۰	۶۲
۱۳	۲۰۰	۱۳۰	۶۶
۱۴	۲۰۰	۱۴۰	۷۰
۱۵	۲۰۰	۱۵۰	۷۳
۱۶	۲۰۰	۱۶۰	۷۶
۱۷	۲۰۰	۱۷۰	۷۸
۱۸	۲۰۰	۱۸۰	۸۰
۱۹	۲۰۰	۱۹۰	۸۲
۲۰	۲۰۰	۲۰۰	۸۳

جدول ۲- ژرفای بالایی یا کمینه D، گسترش عمودی V و گسترش افقی H بر حسب متر مربوط به منابع تاقدیسی مجموعه امتحان کننده.

شماره نیمرخ	D	H	V
۲۱	۱۰	۱۰۶	۵۶
۲۲	۲۰	۱۱۶	۶۱
۲۳	۳۰	۱۲۶	۶۵
۲۴	۴۰	۱۳۶	۶۹
۲۵	۵۰	۱۴۶	۷۲
۲۶	۶۰	۱۵۶	۷۵
۲۷	۷۰	۱۶۶	۷۸
۲۸	۸۰	۱۷۶	۸۰
۲۹	۹۰	۱۸۶	۸۲
۳۰	۱۰۰	۱۹۶	۸۳
۳۱	۱۱۰	۱۰۶	۵۶
۳۲	۱۲۰	۱۱۶	۶۱
۳۳	۱۳۰	۱۲۶	۶۵
۳۴	۱۴۰	۱۳۶	۶۹
۳۵	۱۵۰	۱۴۶	۷۲
۳۶	۱۶۰	۱۵۶	۷۵
۳۷	۱۷۰	۱۶۶	۷۸
۳۸	۱۸۰	۱۷۶	۸۰
۳۹	۱۹۰	۱۸۶	۸۲
۴۰	۲۰۰	۱۹۶	۸۳

جدول ۳- مشخصه‌های استخراج‌شده از ۲۰ نیمرخ گرانی مصنوعی مجموعه تربیت کننده و نام آنها.

شماره مشخصه‌ها	مشخصه‌های استخراج شده
۱	X_{g50}
۲	X_{g55}
۳	X_{g60}
۴	X_{g65}
۵	X_{g70}
۶	X_{g75}
۷	X_{g80}
۸	X_{g85}
۹	g_{max}
۱۰	انتگرال صد متر مرکزی نیمرخ گرانی

جدول ۴- فهرست ۴۵ حالت ۲ مشخصه‌ای ممکن از میان ۱۰ مشخصه استخراج‌شده (جدول ۳) از مجموعه تریتی و PM مربوط به آنها.

PM مربوطه	شماره مشخصه دیگر	شماره یک مشخصه	PM مربوطه	شماره مشخصه دیگر	شماره یک مشخصه
۳۷۱	۹	۸	۲۹۷	۸	۲
۳۶۱	۸	۷	۲۹۷	۸	۳
۳۵۳	۹	۷	۲۹۴	۴	۱
۳۴۸	۸	۶	۲۹۲	۵	۱
۳۴۴	۷	۶	۲۹۱	۶	۱
۳۴۳	۸	۵	۲۹۱	۷	۱
۳۴۰	۷	۵	۲۸۹	۸	۱
۳۳۹	۸	۴	۲۸۳	۹	۲
۳۳۶	۷	۴	۲۸۲	۳	۲
۳۳۶	۹	۶	۲۸۰	۹	۳
۳۳۴	۶	۵	۲۸۰	۲	۱
۳۳۲	۶	۴	۲۷۹	۳	۱
۳۳۱	۹	۵	۲۷۷	۹	۱
۳۳۰	۵	۴	۶۵	۱۰	۱
۳۲۹	۹	۴	۵۷	۱۰	۲
۳۰۰	۴	۳	۵۰	۱۰	۳
۳۰۰	۴	۲	۴۵	۱۰	۴
۲۹۹	۵	۳	۳۹	۱۰	۵
۲۹۹	۵	۲	۳۳	۱۰	۶
۲۹۸	۷	۳	۲۸	۱۰	۷
۲۹۸	۷	۲	۲۳	۱۰	۸
۲۹۸	۶	۳	۱۱	۱۰	۹
۲۹۸	۶	۲			

جدول ۵- ژرفای کمینه واقعی و ژرفاهای کمینه برآورد شده بر حسب متر و خطاهای مربوط به مجموعه امتحان کننده در دو حالت مناسب و نامناسب.

خطای حالت ۲ (درصد)	ژرفای برآوردی حالت ۲	خطای حالت ۱ (درصد)	ژرفای برآوردی حالت ۱	ژرفای واقعی	شماره نیمرخ
۸۸۲	۹۸	۵۸	۱۶	۱۰	۲۱
۳۹۶	۹۹	۱۰	۱۸	۲۰	۲۲
۲۳۴	۱۰۰	۲۷	۲۲	۳۰	۲۳
۱۵۴	۱۰۱	۲۹	۲۹	۴۰	۲۴
۱۰۵	۱۰۳	۲۷	۳۷	۵۰	۲۵
۷۳	۱۰۴	۱۶	۵۰	۶۰	۲۶
۵۰	۱۰۵	۸	۶۵	۷۰	۲۷
۳۳	۱۰۶	۳	۸۲	۸۰	۲۸
۱۹	۱۰۷	۱۲	۱۰۱	۹۰	۲۹
۸	۱۰۸	۲۰	۱۲۰	۱۰۰	۳۰
۱۶	۱۲۷	۴	۱۰۶	۱۱۰	۳۱
۶	۱۲۷	۴	۱۲۵	۱۲۰	۳۲
۳	۱۲۶	۱۲	۱۴۶	۱۳۰	۳۳
۱۰	۱۲۶	۱۵	۱۶۰	۱۴۰	۳۴
۱۷	۱۲۵	۱۴	۱۷۲	۱۵۰	۳۵
۲۲	۱۲۵	۱۳	۱۸۲	۱۶۰	۳۶
۲۷	۱۲۴	۱۰	۱۸۷	۱۷۰	۳۷
۳۱	۱۲۴	۶	۱۹۱	۱۸۰	۳۸
۳۵	۱۲۴	۲	۱۹۳	۱۹۰	۳۹
۳۸	۱۲۴	۳	۱۹۵	۲۰۰	۴۰

References

Blakely, J. R., 1996- Potential Theory in Gravity & Magnetic Applications, Cambridge University Press, 441 pages.

Duin, R. P. W., Juszczak, P., Paclik, P., Pekalska, E., Ridder, D. de, Tax, D. M. J. & Verzakov, S., 2007- PRTools4.1, A Matlab Toolbox for Pattern Recognition, Delft University of Technology, Netherlands, 61 pages. http://www.prtools.org/.

Gret, A. A. & Klingele, E. E., 1998- Application of Artificial Neural Networks for Gravity Interpretation in Two Dimensions, Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology, 88 pages.

Hashemi, H., Tax, D. M. J., Duin, R. P. W., Javaherian, A. & Groot, P., 2008- Gas chimney detection based on improving the performance of combined multilayer perceptron and support vector classifier, Nonlinear Processes in Geophysics, 15, 863-871.

Hashemi, H., 2010- Logical considerations in applying pattern recognition techniques on seismic data: Precise ruling, realistic solutions, Cseg Recorder, 35, 47-50.

Heijden, F., Van der, Duin, R. P. W., Ridder, D. de & Tax, D. M. J., 2004- Classification, Parameter Estimation and State Estimation, John Wiley & Sons, Ltd, 423 pages.

Hubbert, M. K., 1948- A line-integral method of computing the gravimetric effects of two-dimensional masses, Geophysics, 13, 215-225.

Talwani, M., Worzel, J. L. & Landisman, M., 1959- Rapid gravity computations for two dimensional bodies with application to the Mendocino submarine fracture zone, Journal of Geophysical Research, 64, 49-59.

Tan, M., 2004- Support Vector Machine & Its Applications. The University of British Columbia, Canada. http://www.iro.umontreal.ca/~pift6080/H09/documents/papers/svm_tutorial.ppt.

Importance of Features Selection in Application of Pattern Recognition in Estimating the Depth of Anticline Using Gravity Data

M. E. Hekmatian [1*], V. Ebrahimzadeh Ardestani [2], M. A. Riahi [2], A. Memar Koucheh Bagh [3] & J. Amini [4]

[1] Ph.D. Student, Faculty of Basic Sciences, Islamic Azad University, Science and Research Branch, Tehran, Iran

[2] Associate Professor, Institute of Geophysics of University of Tehran, Tehran, Iran

[3] Assistant Professor, Nuclear Fuel Cycle Research School, Nuclear Science & Technology Research Institute (NSTRI) , Tehran, Iran

[4] Associate Professor, Faculty of Engineering, University of Tehran, Tehran, Iran

Received: 2012 May 09 Accepted: 2013 March 09

Abstract

Pattern recognition algorithms especially neural network in geophysical interpretations and other Earth sciences have been used since some years ago. In neural network and other pattern recognition algorithms like support vector classifier (SVC) that the latter method is used in this research, by using the values of the features, which has been extracted from the objects (in our work gravity profiles are objects), classification of the objects can be done. Usually the features are selected subjectively. In this paper, we have presented a homemade software that can select proper features objectively. By using SVC and the mentioned features selection (FS) software, depth estimations of anticlines have been done in this research. We have shown the difference of using proper features and improper ones in the mentioned depth estimation (a kind of classification). In this paper, twenty synthetic gravity profiles with anticline shape sources are created for training SVC and the same amount of synthetic profiles are created for testing. It has shown that depth estimation with proper features is more precise than depth estimation with improper features. Also it should be emphasized that FS is important not only in depth estimation of anticlines, but also in all kinds of classifications in Earth sciences and the mentioned homemade software code is applicable in all of them.

Keywords: Feature selection, Pattern recognition, Depth estimation of anticline, Gravity, Support vector classifier.

For Persian Version see pages 141 to 148

*Corresponding author: M. E. Hekmatian; E-mail: mhekmatian@aeoi.org.ir

Created with nitroPDF professional
download the free trial online at nitropdf.com/professional