**حل تمرين داده کاوي**

**(بخش دوم)**

**استاد:جناب آقاي دكتر فراهاني**

**استاد يار: آقاي شريفي**

**دانشجو: ياشار موسي پور**

**8-** **Bootstrapping چیست و چه تفاوتی با Validation Cross دارد؟ در کجا ها از Bootstrappinاستفاده میشود ؟**

**پاسخ:  
Bootstrapping يك روش باز نمونه گيري براي تخمين پارامترهاي جامعه.**

**Validation Cross بررسی در توانایی الگوریتم برای فراگیری و تعمیم.** 

**زماني كه بررسي بر مواردي از جامعه ميباشد كه اعضاي جامعه به راحتي قابل شناسايي يا در دسترس نيستند.**

**به عنوان مثال  
- بررسي در مورد يك گونه جانوري يا گياهي كمياب  
- بيمارهاي خاص كه اعضاي واقعي قابل شناسايي نيستند**

**10- در خصوص الگوریتم های مختلف ساخت درخت تصمیم (همانند ID3 ، CART و... ) تحقیق کنید و به صورت کلی مشخص نماييد تفاوت الگوریتم های مختلف ساخت درخت تصمیم در چیست ؟**

**پاسخ :**

## زمانی که درخت برای کارهای طبقه‌­بندی استفاده می‌­شود، به عنوان درخت طبقه­‌بندی (Classification Tree)شناخته می‌­شود .

## و هنگامی که برای فعالیت‌­های رگرسیونی به کار می‌­رود درخت رگرسیون (RegressionDecisionTree) نامیده می­شود .

**انواع الگوریتم‌­های شایع برای برقراری درخت تصمیم**

* **ID3**

**یکی از الگوریتم‌­های بسیار ساده درخت تصمیم که در سال 1986 توسط Quinlan مطرح شده است. اطلاعات به دست آمده به عنوان معیار تفکیک به کار می­‌رود. این الگوریتم هیچ فرایند هرس کردن را به کار نمی­‌برد و مقادیر اسمی و مفقوده را مورد توجه قرار نمی­‌دهد.**

* **C4.5**

**این الگوریتم درخت تصمیم، تکامل یافته ID3 است که در سال 1993 توسط Quinlan  مطرح شده است.**

**-Gain Ratio**

**به عنوان معیار تفکیک در نظر گرفته می­‌شود. عمل تفکیک زمانی که تمامی نمونه‌­ها پایین آستانه مشخصی واقع می­‌شوند، متوقف می­‌شود. پس از فاز رشد درخت عمل هرس کردن بر اساس خطا اعمال می­‌شود. این الگوریتم مشخصه­‌های اسمی را نیز در نظر می­‌گیرد.**

* **CART**

**برای برقراری درخت­‌های رگرسیون و دسته‌­بندی از این الگوریتم استفاده می­‌شود. در سال 1984توسط Breiman و همکارانش ارائه شده است. نکته حائز اهمیت این است که این الگوریتم درخت­‌های باینری ایجاد می­‌کند به طوری که از هر گره داخلی دو لبه از آن خارج می­‌شود و درخت­‌های بدست آمده توسط روش اثربخشی هزینه، هرس می­‌شوند.**

**یکی از ویژگی‌­های این الگوریتم، توانایی در تولید درخت­‌های رگرسیون است. در این نوع از درخت­‌ها برگ‌ها به جای کلاس مقدار واقعی را پیش­بینی می‌­کنند. الگوریتم برای تفکیک کننده‌­ها، میزان مینیمم مربع خطا را جستجو می­‌کند. در هر برگ، مقدار پیش­بینی بر اساس میانگین خطای گره‌­ها می­‌باشد.**

* **CHID**

**این الگوریتم درخت تصمیم به جهت در نظرگرفتن مشخصه­‌های اسمی در سال 1981 توسط Kass طراحی شده است. الگوریتم برای هر مشخصه ورودی یک جفت مقدار که حداقل تفاوت را با مشخصه هدف داشته باشد، پیدا می­‌کند.**

**.....................................................................................................................................................**

**١٣ . در خصوص هرس کردن Pruning درخت تصمیم تحقیق کنید . چرا ما به بحث هرس کردن درخت تصمیم نیاز دارد و چه کمکی به ما میکند؟**

**پاسخ: هرس کردن (Pruning) یک الگوریتم ساده و شهودی است. انواع مختلفی وجود دارد، اما ایده اصلی بر روی هر شبکه عصبی کار می‌کند. ایده این است. درون یک شبکه عصبی بزرگ آموزش‌دیده، مقداری وزن با اندازه بزرگ و مقداری با اندازه کوچک وجود خواهد داشت. به طور طبیعی، وزن‌های با اندازه بزرگ بیشتر به خروجی شبکه کمک می‌کنند. بنابراین، برای کاهش اندازه شبکه، ما از وزن‌های کوچک (هرس) خلاص می‌شویم.**

**یک مزیت هرس این است که واقعا خوب عمل می‌کند. شواهد تجربی قابل‌توجهی برای حمایت از آن وجود دارد. به طور خاص‌تر، هرس برای حفظ دقت و در عین حال کاهش اندازه شبکه (حافظه) نشان‌داده شده‌است.**