**سوال اول:**

به عنوان مثال مسیله حدس زدن اعداد دست نوشته، یک مسیله دسته­ بندی است.

مسایل رگرسیون:

* برای پیش­بینی میزان ثروت افراد
* در صنعت دارویی برای تست خون فوری
* در صنعت هتل­داری برای پیش­بینی ظرفیت مورد نیاز
* در صنعت بازی و پیدا کردن تبلیغ مناسب
* در ورزش برای پیش­بینی تاثیر تمرینات مختلف

مسایل دسته بندی:

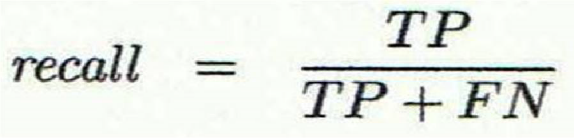
* Spam filtering
* Image classification
* Malware classification
* fraud detection
* Document classification

**سوال دوم:**

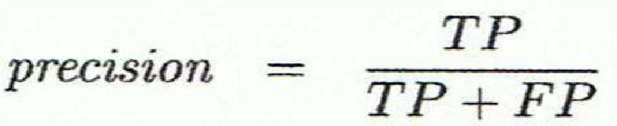
* Accuracy: درصد تاپل­های مجموعه تست که به درستی طبقه بندی شده­اند.



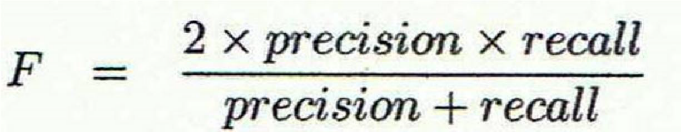
* Recall: کامل بودن - درصدی از تاپل­های مثبت که طبقه­بندی کننده به عنوان مثبت برچسب گذاری کرده است.



* Precision: دقت - درصدی از تاپل­هایی که طبقه­بندی کننده آنها را به عنوان مثبت برچسب گذاری کرده است و در واقع مثبت هستند.



* F1-Score: میانگین هارمونیک دو پارامتر بالا را می­گویند.



**سوال سوم:**

فرض می­کنیم مسیله پیدا کردن داشتن بیماری عرقی است.

حال باید برای ویژگی­ بیماری قلبی انتروپی را محاسبه کنیم:

در ادامه Gain شاخه عروق خونی بسته را بررسی می­کنیم:

از آنجایی که Gain ویژگی عروق بسته برابر 1 شده است پس می­توان این ویژگی را در ریشه درخت قرار داد و دسته­بندی را انجام داد.

**سوال چهارم:**

فرض می­کنیم مسیله پیدا کردن دوست داشتن سریال کلاه قرمزی است.

حال باید برای ویژگی­ دوست داشتن سریال کلاه قرمزی انتروپی را محاسبه کنیم:

شاخه سن را به سه دسته تقسیم می­کنیم:

* کمتر از 18 سال
* بین 18 تا 38 سال
* بالای 38 سال

حال Gain را به دست می­آوریم:

از آنجایی که Gain ویژگی سن برابر 0.862 شده است پس می­توان این ویژگی را در ریشه درخت قرار داد و دسته­بندی را انجام داد.

**سوال پنجم:**

این ویژگی میزان عدم خلوص را در یک نود نشان می­دهد. روش محاسبه آن در فرمول زیر آمده است:

زمانی این ویژگی ماکسیمم یا 1 می­شود که از هرکدام از کلاس­ها به تعداد برابر داشته باشیم که نشان دهنده ارزنده بودن آن اطلاعات می­شود.

زمانی که در یک نود فقط یک کلاس داشته باشیم آنگاه این ویژگی مینیمم می­شود که نشان دنده کم ارزش بودن این اطلاعات می­باشد.

آنتروپی نیز شبیه به این ویژگی کار می­کند. البته جینی اطلاعات بهتری نسبت انتروپی می­دهد.

**سوال ششم:**

بیش برازش زمانی اتفاق می­فتد که مدل بجای یادگیری از مجموعه یادگیری شروع به حفظ کردن آن کند. این باعث می­شود که زمانی مجموعه تست یا جدید به مدل داده می­شود، نتواند نتیجه مطلوبی بگیرد.

در روش Regularization پیچیدگی مدل را کنترل می­کنیم و اجازه نمی­دهیم مدل بیش برازش کند.

روش بعدی Drop Out نام دارد که به صورت رندوم وزن­های شبکه را از پروسه یادگیری خارج می­کند. این عمل باعث می­شود که بیش برازش اتفاق نیافتد.