

گزارش بخش دوم تمرین دوم

محمد ویس مصطفی پور 97222085

سوال دوم: دو راه وجود دارد. راه اول استفاده از cross validation است، که آلفاهای مختلفی را به ما میدهد. راه حل دوم آزمون و خطاست تا بهترین آلفا را بدست بیاوریم.

سوال سوم: با افزایش فولد ها دقت واقعی افزایش میابد و بایاس مدل کمتر میشود، اما در عوض زمان بیشتری برای اجرا لازم خواهد داشت و این عیب افزایش فولد هاست. همچنین تعداد داده ها نباید از حد خاصی پایین تر بیاید وگرنه برای اعتبارسنجی به مشکل میخوریم.

سوال چهارم: زیر مجموعه خاصی از cross-validation است. که در آن تعداد instance ها با تعداد فولد ها برابر میشود. جاهایی که نتیجه خیلی برای ما اهمیت دارد و درصد دقت بالایی میخواهیم از طرفی هزینه زمانی برایمان چندان مشکل آفرین نیست میتوانیم از این روش استفاده کنیم تا به دقت بالاتری دست پیدا کنیم

سوال پنجم: در این روش یادگیرنده های یکسان از مجموعه های آموزشی متفاوت استفاده میکنند. از یک مجموعه آموزشی به طول N ، N داده به صورت تصادفی و همراه با جایگزینی استخراج میشود. فرایند فوق L بار تکرار میشود. و از میانگین گیری برای ترکیب نتایج استفاده میشود.

یک مجموعه داده در Bootstrap ممکن است چندین مورد از موارد اصلی یکسان را داشته باشد و سایر موارد اصلی را کاملاً حذف کند. ولی نمونه های cross validation بدون جایگزینی و در نتیجه مجموعه داده های جایگزین کوچکتر از نسخه اصلی تولید می کند

سوال ششم: این عبارت به این معناست که fold-2 را برای پنج مرتبه تکرار میکنیم. به این دلیل از 2fold استفاده میکنیم تا اطمینان حاصل شود که هر مشاهده تنها یک بار در آزمون مدل استفاده میشود.