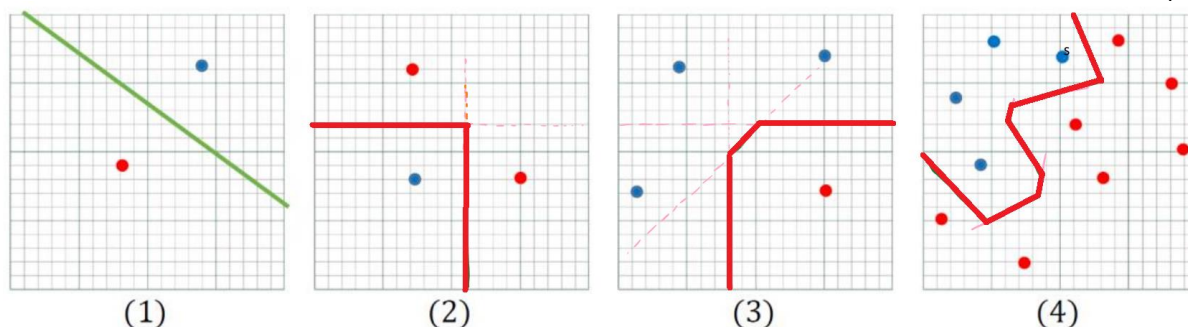


پاسخ سوال ۱

(الف)



(ب) خیر، همواره داده های ما به صورت مثال مطرح شده از یکدیگر جدا نیستند بعضا پیچیدگی در جداسازی داده ها وجود دارد در بدترین حالت ممکن برای جداسازی یک داده (نقطه) از سایر داده های کلاس دیگر نیاز به چند خط (مرز تصمیم) داریم که تعداد پارامترهای مورد نیاز برای ترسیم خط بسیار بیشتر از نقاط خواهند بود. بنابراین ذخیره ی آن ها از نظر زمانی و حافظای به صرفه نیست.

(ج) با توجه به اینکه الگوریتم NN مدلی از پیش نمی سازد و تنها در زمان تست داده ها را بررسی می کند لذا مشکلی با داده هایی که در زمان های دیگر وارد عرصه می شوند ندارد.

پاسخ سوال ۲

درخت تصمیمی که شاخه ها و عمق زیادی داشته باشد، به سمت **overfit** شدن مدل پیش می رود که نتیجه ی آن عدم دقت مناسب روی داده های تست است . هرس کردن به حل این مشکل کمک می کند در در حالت کلی اصطلاحا درختی جنرال تر خواهیم داشت . در پیش هرس قبل از اینکه درخت عمیق شود، الگوریتم را با تعیین یک آستانه مناسب متوقف می کنیم. در پس هرس کردن الگوریتم درخت تصمیم را به صورت کامل ایجاد می کند و سپس به جهت جنرال سازی درخت، بعضی از شاخه ها حذف می شوند .

ابتدا درخت **T** را روی همه ی داده های **D** اعمال می کنیم . ساختار درخت به این صورت خواهد بود که هر داده در یک برگ قرار می گیرد که ممکن است درست یا غلط طبقه بندی شده باشد . هرکدام از برگ های **T** را که حداقل یک داده ی جدید در آن قرار گرفته حذف کرده و در آن نقطه الگوریتم تصمیم را روی داده ها (برگ ها) اجرا می کنیم و درخت را در آن شاخه گسترش می دهیم . برای این کار باید توجه داشته باشیم در هر کدام از برگ های درخت **T** چه تعداد داده و از چه کلاسی قرار داشته است .