

1- الف) در مسئله رگرسیون با در نظر گرفتن مربع خطا باقیمانده، به مسئله بهینه سازی $\min \|X\beta - y\|_2^2$ می رسیدیم. اگر به جای مربع خطا از قدرمطلق استفاده شود، مسئله بهینه سازی حاصل به چه صورتی خواهد بود و انتظار دارید در نظر گرفتن قدرمطلق سبب چه تغییری روی نتیجه به دست آمده شود. ب) در نظر گرفتن قدرمطلق سبب چه تغییری در روند حل با رویکرد بهینه سازی gradient descent خواهد شد. ج) 3 کلاس داده تحت عناوین کلاس های A, B و C داریم. از هر کلاس، 5 داده در اختیار داریم و هر داده شامل k ویژگی است. هدف مشخص نمودن کلاس داده جدید y است (که این داده هم شامل همان k ویژگی است). برای این منظور می خواهیم با استفاده از رگرسیون این کلاس بندی را انجام دهیم. روش پیشنهادی شما چیست. مراحل را شرح دهید.

داده های کلاس A: $x_{A1}, x_{A2}, x_{A3}, x_{A4}, x_{A5}$

داده های کلاس B: $x_{B1}, x_{B2}, x_{B3}, x_{B4}, x_{B5}$

داده های کلاس C: $x_{C1}, x_{C2}, x_{C3}, x_{C4}, x_{C5}$

2- قارچ ها به دو دسته سمی و غیرسمی تقسیم می شوند. اطلاعات زیر را در اختیار داریم. هدف استفاده از درخت های تصمیم برای تعیین سمی یا غیرسمی بودن داده شماره 9 است و داده های 1 تا 8 داده های آموزشی هستند. از معیار gini برای ساخت درخت تصمیم استفاده نمایید.

	Heavy	Spotted ¹	Smooth	Poisonous
1	No	No	No	No
2	No	Yes	No	No
3	Yes	No	Yes	No
4	Yes	No	Yes	Yes
5	No	Yes	No	Yes
6	No	Yes	Yes	Yes
7	No	No	Yes	Yes
8	Yes	No	No	Yes
9	No	No	Yes	?

الف) کدام ویژگی را به عنوان ریشه درخت انتخاب می کنید ب) بهره به دست آمده بر اثر انتخاب ویژگی قسمت الف را محاسبه نمایید ج) درخت تصمیم را تا دو سطح سوال از ویژگیها به دست آورید. د) پیش بینی درخت به دست آمده در بخش ج برای داده آخر (داده شماره 9) چیست. در تمامی مراحل تمامی محاسبات ذکر شود. نکته : اگر در انتخاب ویژگی ها برای تصمیم گیری در ساخت درخت، چندین ویژگی از نظر معیار ارزیابی مشابه بودند، به طور تصادفی یکی را انتخاب نمایید