תאריך הבחינה:	19.4.2024
המרצה:	דר' יקיר ברצ'נקו
שם הקורס:	מודלים של רגרסיה ליניארית
מספר קורס:	364-1-1061
משך הבחינה:	שתי שעות
חומר עזר:	מחשבון, 5 דפי נוסחאות, טבלאות סטטיסטיות

מיועד לתלמידי הנדסת תעשיה וניהול.

שנה: 2024 סמסטר: א' מועד: ב'

הנחיות חשובות:

א. במבחן זה שלוש שאלות

ב. יש לפתור את כל השאלות, סך הנקודות עשוי להגיע ל108, אך הציון הסופי המקסימלי במבחן הינו 100.

ג. יש ללוות את דרך הפתרון בהסברים מילוליים קצרים

ד. ניתן להשתמש בכל הנוסחאות והתוצאות שפותחו או הוצגו בכיתה ואין צורך להוכיחן מחדש אלא אם כן זה נדרש במפורש

<u>שאלה 2 (45 נקודות)</u>

מונטגומרי המבולבל הכין את האיור שלמטה עבור ספר הקורס שלו ברגרסיה, אך הוא שכח לצרף את ההסבר שמלווה את האיור (שמיועד לתלמידי הנדסה באוניברסיטה). בנוסף, לרוע מזלו נשפך לו קפה על חלקים באיור.

א. מה האיור אמור לתאר? ומה ההסבר שכדאי שילווה אותו? כיתבו את התיאור וההסבר שאמור ללוות את האיור שלמטה. השתמשו בהסבר שלכם בכל אחד מהמושגים a-g.

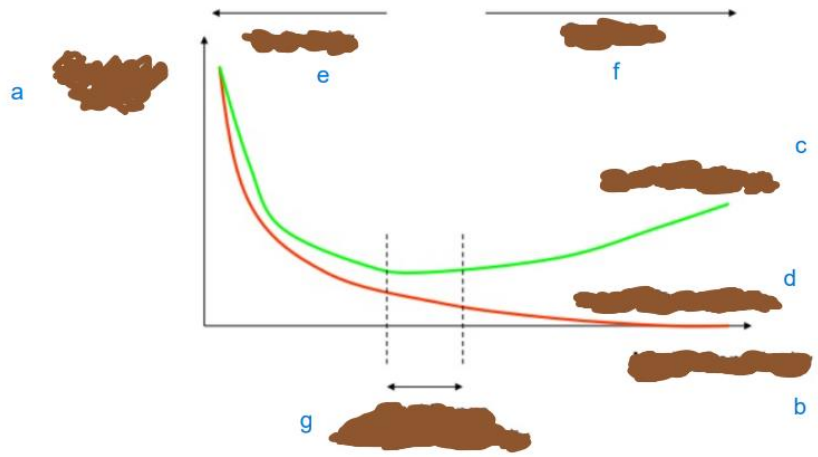
ב. בשלב בחירת מודלים, מתי תשתמשו ב-AIC ומתי ב-BIC? נמקו בקצרה.

ג. יותם מתלבט בין שני מודלים:

M_b, מודל גדול עם הרבה פרמטרים, שיכול לתאר במדויק את פונקציית הצפיפות (הבלתי-ידועה, שאותה מנסים לאמוד), ולעומתו

M_s, מודל קטן **שאיננו** יכול לתאר במדויק את פונקציית הצפיפות הבלתי-ידועה (אבל בעל מעט פרמטרים).

נניח שיש ליותם מספר גדול מאוד של תצפיות, מה יקרה אם ישתמש ב-AIC או BIC? האם/איזה הבדל יהיה בין שתי השיטות? נמקו.



פיתרון:

א. איור מלא ופרטים נוספים מופיעים בחוברת ההרצאות של הקורס (עמוד 66).

ב. נבחר ב-AIC כאשר יותר חשובה לנו שגיאת החיזוי העתידית, ואילו נבחר ב-BIC כאשר יותר חשוב לנו להכניס למודל הסופי אך ורק משתנים רלוונטיים. ניתן (ורצוי היה) לנמק זאת בעזרת העקומה הירוקה שבאיור הנ"ל, שעולה ימינה ולמעלה מנקודת המינימום במתינות רבה יותר מאשר יורדת אליה משמאל, ולכן בחירת מודל (אקראית, בגלל המידגם) שמושכת שמאלה לכיוון מודלים קטנים יותר, כמו BIC, יכולה ללכת רחוק מדי יותר בקלות ולספוג שגיאת-תחזית גדולה יותר.

ג. נשים לב שעבור מדגם גדול לא יהיה בעצם הבדל גדול בין AIC ל-BIC (כן, קצת בניגוד לעבודה הקשה של סעיף ב')!

כתלות ב-N, מספר התצפיות, לשני הקריטריונים יש את גורם טיב-ההתאמה שגדל בשניהם באופן לינארי ב-N. עבור המודל הגדול של יותם תהיה התאמה טובה יותר, ואילו עבור המודל הקטן של יותם תהיה התאמה קטנה יותר. הקנס בשני הקריטריונים, לעומת טיב-ההתאמה, גדל באופן הרבה יותר איטי כתלות ב-N: אצל BIC הוא גדל באופן לוגריתמי, ואילו אצל AIC הוא בכלל קבוע.

בהתאם, גורם הקנס (בשני הקריטריונים) הינו זניח לעומת גורם טיב-ההתמאה, שזהה בשני הקריטריונים (לוג הניראות) ולכן טיב ההתאמה יהיה הגורם המכריע בבחירת המודל הסופי – המודל הגדול במקרה זה.