

האוניברסיטה העברית
המחלקה לסטטיסטיקה

מבחן סיום בקורס: (52609) שיטות חישוביות בסטטיסטיקה

תאריך הבחינה: 25.3.10, שעה 10:00. (מועד ב)

משך המבחן: שעתיים

חומר מותר בשימוש: 3 דפים כתובים בכתב ידו של הסטודנט.

הוראות: יש לענות על שתי השאלות ולנמק (בקיצור) כל אמירה. בהצלחה!

שאלה 1

הגרסיה בהטלה (*projection pursuit*) היא שיטת רגרסיה פרמטרית למחצה. בהתאם לשיטה זו תוחלת המשתנה המוסבר הוא קומבינציה לינארית של פונקציות של המשתנים המסבירים. מקדמי הקומבינציה הלינארית מסומנים כ- β . הפונקציה ppr מתאימה לנתונים את המודל של רגרסיה בהטלה.

מסד הנתונים oil.rock מכיל תוצאות אנליזה של סלעים שעשויים להכיל נפט. המשתנה המוסבר הוא (לוג) מדד החדירות של הסלעים. המשתנים המסבירים הם שטח הנקבוביות, היקפם, ומדד שמאפיין את צורתם. אנו מעוניינים להעריך את הקשר בין המשתנים המסבירים למשתנה המוסבר. לצורך כך הופעל הקוד הנ"ל:

```
> rock.ppr <- ppr(lperm ~ area + peri + shape,  
+ data=oil.rock, nterms = 2, max.terms = 5)  
> summary(rock.ppr)
```

Call:

```
ppr(formula = lperm ~ area + peri + shape, data = oil.rock, nterms = 2,  
max.terms = 5)
```

Goodness of fit:

```
2 terms 3 terms 4 terms 5 terms  
8.737806 5.289517 4.745799 4.490378
```

Projection direction vectors:

```
term 1 term 2  
area 0.34357179 0.37071027  
peri -0.93781471 -0.61923542  
shape 0.04961846 0.69218595
```

Coefficients of ridge terms:

```
term 1 term 2  
1.6079271 0.5460971  
> rock.ppr$beta  
term 1 term 2  
1.6079271 0.5460971
```

```
> BETA <- matrix(NA,200,2)
```

```
> index <- 1:length(perm)
```

```

> for(j in 1:200)
+ {
+   i <- sample(index)
+   rock.ppr.j <- ppr(lperm[i] ~ area[i] + peri[i] + shape[i],
+     data = oil.rock, nterms = 2, max.terms = 5)
+   BETA[j,] <- rock.ppr.j$beta
+ }
> cov(BETA)
      [,1] [,2]
[1,] 0.033045073 0.007105344
[2,] 0.007105344 0.008643761

```

1. מבין האפשרויות הבאות, מי המתאימה ביותר לאפיון האלגוריתם בו משתמשים?

- i. Bootstrap
- ii. EM
- iii. HMM
- iv. MCMC

נמקו!

2. האם ניתן לקבוע בהתאם לתוצאות ההרצה שוקטור המקדמים β שונה באופן מובהק מוקטור האפס? נמקו.
3. נטען כי אם היינו מחליפים את ביטוי הקוד "j in 1:200" בביטוי "j in 1:2000" אזי המסקנה בסעיף 2 הייתה משתנה באופן משמעותי. האם הנכם מסכימים עם טענה זו אם לא?
4. הפרמטר "nterms" קובע את מספר האיברים בקומבינציה הלינארית (אורך הוקטור β). רשמו קוד המאפשר לבדוק את ההשערה כי איבר שלישי בקומבינציה הלינארית אינו מובהק ושני איברים, כבמודל הנתון, מספיקים.

שאלה 2

נתונים תצפיות בלתי תלויות Y_1, \dots, Y_n . מודל הסתברותי לתצפית Y_i מניח כי בהינתן משתנה רקע λ_i , שאינו נצפה, התצפית מפולגת פואסונית עם תוחלת λ_i . התפלגות משתנה הרקע הינה גמה עם מקדם צורה $r=5$ ומקדם סקלה θ בלתי ידוע ($\lambda_i \sim \text{Gamma}(5, \theta)$). מעוניינים לאמוד את הפרמטר

1. רשמו את (לוג) הצפיפות המשותפת של התצפיות ושל משתני הרקע.
2. רשמו את ההתפלגות המותנית של λ_i , בהינתן ערכם של התצפיות.
3. הציעו אלגוריתם, המבוסס על שיטת ה-EM, כדי לאמוד את הפרמטר המבוקש.