



LUNDS
UNIVERSITET

Ekonomihögskolan

STAN51, Statistik: Maskininlärning ur ett regressionsperspektiv, 7,5 högskolepoäng

Statistics: Machine Learning from a Regression Perspective, 7.5 credits

Avancerad nivå / Second Cycle

Fastställande

Kursplanen är fastställd av Institutionsstyrelsen vid Statistiska institutionen 2020-12-10 att gälla från och med 2020-12-10, höstterminen 2020.

Allmänna uppgifter

Kurs på avancerad nivå i statistik. Kursen kan ingå i en magister- eller masterexamen i statistik. Kursen kan även läsas om fristående kurs eller inom andra magister- eller masterprogram vid Lunds universitet.

Undervisningsspråk: Engelska

(Undervisningen kan ske på svenska om alla registrerade studenter behärskarsvenska.)

Huvudområde

Statistik

Fördjupning

A1N, Avancerad nivå, har endast kurs/er på grundnivå som förkunskapskrav

Kursens mål

Kunskap och förståelse

Studenten ska tillägna sig en förståelse för:

- grundläggande regression och klassificering,
- hur modeller tränas och valideras,
- hur modeller väljs,
- metoder för regularisering och dimensionsreducering,
- grundläggande icke-linjära maskininlärningsmetoder såsom neuronät,
- hur man tolkar resultaten från de genomgångna metoderna.

Färdighet och förmåga

Studenten ska ha förmåga att självständigt:

- tillämpa de genomgångna metoderna i mjukvaruprogrammet R,
- välja en metod som är relevant för ett givet empiriskt problem,
- utvärdera lämpligheten av den valda metoden,
- generalisera sin kunskap till empiriska problem som inte behandlats under kursens gång,
- dra slutsatser från empiriska resultat och ge råd.

Värderingsförmåga och förhållningsätt

Studenten ska utveckla förmåga till vidare studier inom ämnet samt förmåga att med hög grad av självständighet söka och utvärdera ämnesrelevant information. Studenten ska även utveckla förmåga att på egen hand skriva ett empiriskt inriktat arbete där maskininlärning utgör ett väsentligt inslag.

Kursens innehåll

Maskininlärning handlar om statistiska prediktioner som förbättras genom erfarenhet; modellen lär och anpassar sig allt eftersom nya data blir tillgängliga. Priset som en matvarubutik kan ta ut av en leverantör för en annons beror på hur bra den är på att hitta de kunder som är benägna att köpa leverantörens produkter. På samma sätt är det pris som Google kan ta ut för en annonslänk direkt kopplat till dess förmåga att identifiera de individer som med hög sannolikhet kommer att klicka på länken. Det är här maskininlärning kommer in. Den här kursen lär ut grunderna i maskininlärning och det gör den genom att fokusera på de metoder som på ett eller annat sätt bygger på vanlig regressionsanalys. Några av delmomenten i kursen är klassificering med hjälp av logistisk regression, modellval genom informationskriterium och korsvalidering, krympningsmetoder såsom lasso och ridgeregression, dimensionsreduceringsmetoder såsom principalkomponentanalys och partiell regression, och neuronät. Teoristudier varvas med empiriska tillämpningar inom företags- och nationalekonomi. Vid de empiriska tillämpningarna används mjukvaruprogrammet R.

Kursens genomförande

Undervisningsform: Undervisningen består av föreläsningar och övningar.

Kursens examination

1. Examinationsuppgifter: Examinationen består av en skriftlig tentamen och inlämningsuppgifter. Den skriftliga tentamen äger rum vid slutet av kursen. Ytterligare examinationstillfällen anordnas i nära anslutning härtill. Poäng från inlämningsuppgifterna får tillgodoräknas vid examinationstillfällena under innevarande termin. Andra examinationsformer kan användas i begränsad utsträckning.
 2. Begränsning av antal examinationstillfällen: –
- Lunds universitet ser mycket allvarigt på fusk och kommer att vidta disciplinåtgärder mot alla slags försök till fusk i samband med tentamina eller andra examinationsformer. Plagiering betraktas som ett mycket allvarligt akademiskt

brott. Det straff som universitetet kan utdela för detta, och för andra slags fusk i samband med olika former av examination, inkluderar avstängning från universitetet under en viss tidsperiod.

Om så krävs för att en student med varaktig funktionsnedsättning ska ges ett likvärdigt examinationsalternativ jämfört med en student utan funktionsnedsättning, så kan examinator efter samråd med universitetets avdelning för pedagogiskt stöd fatta beslut om alternativ examinationsform för berörd student.

Betyg

Betygsskalan omfattar betygsgraderna Underkänt, E, D, C, B, A.

A (Utmärkt) 85-100 poäng/procent. Ett framstående resultat som är utmärkt vad gäller teoretiskt djup, praktisk relevans, analytisk förmåga och självständighet.

B (Mycket bra) 75-84 poäng/procent. Ett mycket bra resultat som karakteriseras av mycket bra teoretiskt djup, praktisk relevans, analytisk förmåga samt självständighet.

C (Bra) 65-74 poäng/procent. Ett bra resultat som karakteriseras av bra teoretiskt djup, praktisk relevans, analytisk förmåga samt självständighet.

D (Tillfredsställande) 55-64 poäng/procent. Ett resultat som är tillfredsställande vad gäller teoretiskt djup, praktisk relevans, analytisk förmåga och självständighet.

E (Tillräckligt) 50-54 poäng/procent. Ett resultat som möter minimikraven vad gäller teoretiskt djup, praktisk relevans, analytisk förmåga och självständighet, men inte mer.

U (Otillräckligt/Underkänt) 0-49 poäng/procent. Ett resultat som är otillräckligt vad gäller teoretiskt djup, praktisk relevans, analytisk förmåga och självständighet.

För att få godkänt på en kurs måste studenten få betyg E eller högre.

Förkunskapskrav

90 hp i statistik varav minst 7,5 högskolepoäng regressionsanalys eller ekonometri, eller motsvarande.

Övrigt

Denna kurs ersätter tillsammans med STAN52 kursen STAN45. Kursen kan inte ingå tillsammans med STAN45 i en examen.