



Linnéuniversitetet

Kalmar Växjö

Kursplan

Fakulteten för teknik

Institutionen för matematik

4MA508 Statistisk dataanalys, 7,5 högskolepoäng

Statistical Data Analysis

Huvudområde

Matematik

Ämnesgrupp

Matematik

Nivå

Avancerad nivå

Fördjupning

A1N

Fastställande

Fastställd 2021-09-27

Senast reviderad 2022-06-13 av Fakulteten för teknik. Revidering av litteraturlistan. Kursplanen gäller från och med höstterminen 2022

Förkunskaper

Grundläggande behörighet för studier på avancerad nivå samt särskild behörighet:

- I matematik: Linjär algebra (1MA451), 7,5 hp, eller motsvarande.
- I matematisk statistik: Sannolikhetslära och statistik (1MA501), 7,5 hp eller motsvarande.

Mål

Kursen ger en introduktion till moderna statistiska metoder för modellering och förutsägelse av data.

Efter slutförd kurs skall studenten kunna:

Kunskap och förståelse:

- Förklara grundläggande koncept och principer för moderna metoder inom statistisk dataanalys gällande regression, klassificering, återsamlingsmetoder, linjärt modellval, regularisering, trädbaserade metoder, stödvektormaskiner, en- och flerskikts neurala nätverk och icke-väglett lärande.
- Härleda elementära samband inom statistisk dataanalys gällande regression, klassificering, återsamlingsmetoder, linjärt modellval, regularisering, trädbaserade metoder, stödvektormaskiner, en- och flerskikts neurala nätverk och

icke-väglett lärande.

- Förklara klassificering och regularisering utifrån bayesianska ansatser.
- Illustrera geometriska algebraiska uttryck gällande regression, klassificering, regularisering, stödvektormaskiner, och icke-väglett lärande.
- Redogöra för i vilka sammanhang vilka metoder inom regression, klassificering, återsamlingsmetoder, linjära modellval, regularisering, trädbaserade metoder, stödjande vektormaskiner, en- och flerskiktets neurala nätverk och icke-väglett lärande är lämpliga.

Färdighet och förmåga:

- Utföra för små stickprov beräkningar med papper och penna grundläggande regression, klassificering, återsamlingsmetoder, linjärt modellval, regularisering, trädbaserade metoder, stödvektormaskiner, en- och flerskiktets neurala nätverk och icke-väglett lärande.
- Översätta vagt formulerade frågeställningar till statistiskt formelspråk gällande regression, klassificering, återsamlingsmetoder, linjära modellval, regularisering, trädbaserade metoder, stödvektormaskiner, en- och flerskiktets neurala nätverk och icke-väglett lärande.
- Implementera regression, klassificering, återsamlingsmetoder, linjära modellval, regularisering, trädbaserade metoder, stödvektormaskiner, en- och flerskiktets neurala nätverk och icke-väglett lärande för ett givet datamaterial med lämpliga statistiska dataverktyg.
- Genomföra modellval med ensemblemetoder.
- Visa förmåga att planera och med adekvata metoder genomföra kvalificerade uppgifter inom givna ramar samt utvärdera detta arbete.
- Visa förmåga till lagarbete och samverkan i grupp.
- Visa förmåga att muntligt och skriftligt klart redogöra för och diskutera sina slutsatser och den kunskap och de argument som ligger till grund för dessa.

Värderingsförmåga och förhållningssätt:

- Identifiera lämpliga ansatser för regression, klassificering, återsamlingsmetoder, linjärt modellval, regularisering, trädbaserade metoder, stödvektormaskiner, en- och flerskiktets neurala nätverk och icke-väglett lärande för ett givet datamaterial.
- Utvärdera modellval med ensemblemetoden.
- Utvärdera resultat av datautskriften från ett lämpligt statistiskt dataverktyg gällande regression, klassificering, återsamlingsmetoder, linjärt modellval, regularisering, trädbaserade metoder, stödvektormaskiner, en- och flerskiktets neurala nätverk och icke-väglett lärande för ett givet datamaterial.
- Redogöra för varför man för ett givet datamaterial måste iaktta försiktighet gällande slutsatser vid analys av ett givet datamaterial gällande regression, klassificering, återsamlingsmetoder, linjärt modellval, regularisering, trädbaserade metoder, stödvektormaskiner, en- och flerskiktets neurala nätverk och icke-väglett lärande.

Innehåll

- Multipel linjär, polynomiell regression
- logistisk regression
- generaliserade additiva modeller
- linjär och kvadratisk diskriminantanalys

- k-närmaste granne-klassificering
- korsvalidering
- linjärt modellval
- regularisering
- trädbaserade metoder
- ensemblemetoder
- stödvektormaskiner
- en- och flerskiktade neurala nätverk
- principalkomponentanalys
- k-medelklustring

Undervisningsformer

Föreläsningar, lärarledda räkneövningar och lärarledda möten relaterade till inlämningsuppgifterna. Projektuppgifter genomförs i grupp.

Examination

Kursen bedöms med betygen A, B, C, D, E, Fx eller F.

Betyget A utgör det högsta betygssteget, resterande betyg följer i fallande ordning där betyget E utgör det lägsta betygssteget för att vara godkänd. Betyget F innebär att studentens prestationer bedömts som underkända.

Examinationen av kursen delas in i följande moment:

2201 - Inlämningsuppgifter - 2 hp - A, B, C, D, E, Fx eller F

2202 - Muntlig tentamen - 3 hp - A, B, C, D, E, Fx eller F

2203 - Projekt - 2 hp - A, B, C, D, E, Fx eller F

2204 - Opposition projekt - 0,5 hp - Underkänd eller Godkänd

Slutbetyget bestäms från 2201 (2/7), 2202 (3/7), och 2203 (2/7).

Förnyad examination ges i enlighet med Lokala regler för kurs och examination på grundnivå och avancerad nivå vid Linnéuniversitetet.

Om universitetet beslutat att en student har rätt till särskilt pedagogiskt stöd på grund av funktionsnedsättning, har examinator rätt att ge ett anpassat prov eller att studenten genomför provet på ett alternativt sätt.

Kursvärdering

Under kursens genomförande eller i nära anslutning till kursen genomförs kursvärdering. Resultat och analys av genomförd kursvärdering ska skyndsamt återkopplas till de studenter som genomfört kursen. Studenter som deltar vid nästa kurstillfälle erhåller återkoppling vid kursstart. Kursvärdering genomförs anonymt.

Övrigt

Betygskriterier för A-F-skalan kommuniceras till studenten via särskilt dokument. Studenten informeras om kursens betygskriterier senast i samband med kursstart.

Kurslitteratur och övriga läromedel

Obligatorisk litteratur

James, Gareth; Witten Daniela, Hastie, Trevor; Tibshirani Robert: An Introduction to Statistical Learning with applications to R, Springer, senaste upplagan. Antal sidor: 400 av 607.

Bredvidläsningslitteratur

Friedman, Jerome; Hastie, Trevor; Tibshirani Robert: The Elements of Statistical

Learning; Data Mining, Inference and Prediction, senaste upplagan.

Christopher M. Bishop, Pattern recognition and machine learning, Springer, senaste upplagan

Richard A. Johnson, Applied multivariate statistical analysis, Pearson, senaste upplagan