



SF2930 Regressionsanalys 7,5 hp

Regression Analysis

Fastställande

Skolchef vid SCI-skolan har 2020-04-15 beslutat att fastställa denna kursplan att gälla från och med HT 2020, diarienummer: S-2020-0376.

Betygsskala

A, B, C, D, E, FX, F

Utbildningsnivå

Avancerad nivå

Huvudområden

Matematik

Särskild behörighet

- Slutförd grundkurs i numerisk analys (SF1544, SF1545 eller motsvarande)
- Slutförd grundkurs i sannolikhets teori och statistik (SF1922, SF1914 eller motsvarande)

Undervisningsspråk

Undervisningsspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

Lärandemål

För att bli godkänd på kursen skastudenten kunna:

- Formulera och tillämpa statistisk regressionsteori
- Formulera och tillämpa avancerade metoder inom statistisk regressionsmodellering
- Designa och implementera avancerade metoder inom regressionsanalys för tillämpningar

Kursinnehåll

Kursen börjar med modellanpassning, inferens och prediktion för enkla och multipla linjära regressionsmodeller. Särskild uppmärksamhet ägnas åt diagnostiska strategier som är viktiga komponenter för bra modellpassning. Ytterligare ämnen inkluderar transformer och viktningar för att korrigera modellbrister, multikollinearitet -problemet, variabelselektion och modellbyggnadsteknik. Senare i kursen presenteras några generella strategier för regressionsmodellering med särskild inriktning på de generaliserade linjära modellerna (GLM) med hjälp av exemplen med binära och andra kategoriska responsvariabler.

Eftersom insamling av högdimensionella data, med storleksordningen mycket större än vad som krävs för den klassiska regressionsteorin, är numera en regel snarare än undantag (t.ex. inom informationsteknologi, finans, genetik och astrofysik, för att bara nämna några), presenteras regression metoder som tillgodoser sådan data. Tyngdpunkten ligger på regulariseringsmetoder (Ridge, Lasso- och Elastic-Net regression), liksom metoder som använder härledda inmatningsriktningar (principalkomponentsregression och partiell minsta-kvadrat) som gör det möjligt att dämpa statistisk variation i högdimensionella skattnings- och prediktions problem.

Ett antal statistiska inlärningsmetoder med fokus på datorbaserade/datorintensiva algoritmer presenteras från regressionsperspektivet.

Datorbaserade projekt reella dataproblem utgör en viktig lärandeaktivitet.

Examination

- TENA - Skriftlig tentamen, 4,5 hp, betygsskala: A, B, C, D, E, FX, F
- OVN1 - Inlämningsuppgifter, 3,0 hp, betygsskala: P, F

Examinator beslutar, baserat på rekommendation från KTH:s handläggare av stöd till studenter med funktionsnedsättning, om eventuell anpassad examination för studenter med dokumenterad, varaktig funktionsnedsättning.

Examinator får medge annan examinationsform vid omexamination av enstaka studenter.

När kurs inte längre ges har student möjlighet att examineras under ytterligare två läsår.

Etiskt förhållningssätt

- Vid grupparbete har alla i gruppen ansvar för gruppens arbete.

- Vid examination ska varje student ärligt redovisa hjälp som erhållits och källor som använts.
- Vid muntlig examination ska varje student kunna redogöra för hela uppgiften och hela lösningen.