



SF2930 Regressionsanalys 7,5 hp

Regression Analysis

Fastställande

Kursplan för SF2930 gäller från och med VT18

Betygsskala

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

Utbildningsnivå

Avancerad nivå

Huvudområden

Matematik

Särskild behörighet

Avklarade grundkurser i Sannolikhetsteori och statistik, Linjär algebra, En- och Flervariabel analys, Numeriska metoder.

Undervisningsspråk

Undervisningsspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

Lärandemål

Efter fullgjord kurs förväntas studenten kunna:

- känna till statistiska egenskaperna hos punktskattningar som används inom linjär regression såväl som principer och antaganden bakom olika tillämpade skattnings tekniker
- lista och kunna redogöra för antagandena bakom standard parametrisk och modell interferens i linjära regressionsmodellerna
- bedöma passning för en regressionsmodell till data och veta hur man identifierar, diagnostiserar och åtgärder potentiella problem med en linjär regressionsmodell
- utforma och genomföra strategin för att korrigera modellbrister och rapportera om den förväntade noggrannheten som kan uppnås med den föreslagna modellen
- identifiera och utveckla regressionsmodelleringsstrategier som är lämpliga för stora och massiva datamängder samt för små datamängder
- förklara hur den multipla linjära regressionen kan generaliseras för att hantera en respons variabel som är binär eller kategorisk
- använd resampling algoritmer, i synnerhet bootstrap och cross-validering, för uppskattning av modellens prediktiva noggrannhet. Argumentera behoven och styrkor med resampling metoder i regressionsmodellering och validering
- kritiskt utvärdera regressionsmodeller i en verklig tillämpning, presentera analysen och avrapportera förväntad precision i en skriftlig rapport
- läsa nya forskningsuppsatser och redogöra för de frågor som behandlas av aktuell forskning

För att få det högsta betyget skall studenten därtill kunna

- kombinera olika modeller för att få förbättrade resultat

Kursinnehåll

Kursen börjar med modellenpassning, 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.

Kursupplägg

Föreläsningar, presentationer, projektarbete.

Kurslitteratur

Se kursens websida.

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 enskilda studenter.

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

Övriga krav för slutbetyg

Godkända inlämningsuppgifter och tentamen.

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.