

Kurs PM

Statistiska metoder - 30 p

Beskrivning av kursen/innehåll:

Kursen behandlar sannolikhetslära och grunderna inom statistik som krävs för maskininläring. Kursen kommer täcka vanliga fördelningar, korrelation, kovarians, väntevärde, estimering och skattning. Kursen kommer även gå igenom tekniker som hypotestestning och regressionsanalys, samt ge en kortfattad introduktion till bayesiansk statistik. Genom kursens gång kommer praktiska uppgifter lösas med hjälp av Python och beräkningspaketen NumPy, SciPy och Statsmodels. Kursens mål är att den studerande ska förstå och kunna tillämpa metoder inom grundläggande statistik samt ta fram diverse statistiska mått givet en datamängd eller fördelning. Studerande ska även förstå och kunna genomföra ett hypotestest samt kritiskt granska resultaten, samt kunna tillämpa enkel regressionsanalys. Kursens syfte är att ge den studerande den kunskap som krävs för att själv tolka och dra slutsatser kring data och statistik samt förståelse för statistik och sannolikhetslära i den mån som krävs för att utveckla lösningar inom maskininläring och AI.

Teknologier/verktyg som ska användas:

- Jupyter Notebooks
- Python:
 - Pandas
 - NumPy
 - SciPy
 - Statsmodels
 - Seaborn
 - Matplotlib

Utbildare:

David Bernson - david.bernson@iths.se / david.bernson@infotiv.se

Datum:

2023-01-03 - 2023-02-10 - v. 1-6

Kurslitteratur:

Bok: Grundläggande Statistik Analys, Björn Lantz, Studentlitteratur. Tredje upplagan.

Bok: Övningar i Grundläggande Statistisk Analys, Björn Lantz, Studentlitteratur. Första upplagan.

Handouts, jupyter notebooks, föreläsningsanteckningar och videoklipp som utbildaren tillhandahåller.

Studieplan:



Vecka	Datum	Tid	Moment/ämne	Kapitel i Lantz	Rekommenderade uppgifter i Lantz
1	Tis 3/1	9-12, 13-16	Kursintroduktion, 1. Introduktion till statistisk analys,	1.1-1.5; 1.8	1.1-1.5
1	Ons 4/1	9-12, 13-16	2. Sannolikhetslära, Bayes sats	2.1-2.6	2.1-2.3, 2.5-2.8, 2.10-2.11, 2.13-2.17, 2.21-2.32
2	Mån 9/1	9-12, 13-16	Repetition/Räknestuga 1, 2		
2	Tis 10/1	9-12, 13-16	3. Diskreta fördelningar	3.1-3.9	3.1-3.9, 3.11-3.15, 3.20-3.28
2	Tors 12/1	9-12, 13-16	4. Kontinuerliga fördelningar	4.1-4.6	4.2-4.23
2	Fre 13/1	9-12, 13-16	Repetition/Räknestuga 3, 4 Introduktion inlämningsuppgift		
3	Tis 17/1	9-12, 13-16	5. Stickprovsfördelningar/Centrala gränsvärdessatsen	5.1-5.5	5.1-5.6, 5.7-5.11, 5.14-5.17, 5.20
3	Tors 19/1	9-12, 13-16	6. Konfidensintervall	6.1-6.7	6.1-6.7, 6.10, 6.16-6.32
3	Fre 20/1	9-12, 13-16	Repetition/Räknestuga 5, 6		
4	Tis 24/1	9-12, 13-16	7. Hypotestest	7.1-7.6	7.1, 7.6-7.9, 7.11-7.24
4	Tor 26/1	9-12, 13-16	9-12 8. Jmf två populationer 13-16 Introduktion till bayesiansk statistik	8.1-8.6	8.1-8.2, 8.6-8.8, 8.10-8.14, 8.16-21
4	Fre 27/1	9-12, 13-16	Repetition/Räknestuga 7, 8		



5	Tis 31/1	9-12, 13-16	12. Regression och korrelation	12	
5	Tor 2/2	9-12, 13-16	13. Mer om regression	13	
5	Fre 3/2	9-12, 13-16	Repetition/Räknestuga Deadline inlämningsuppgift		
6	Tis 7/2	9-12, 13-16	Repetition/Räknestuga		
6	Tor 9/2	9-12, 13-16	Repetition/Räknestuga		
6	Fre 10/2	9-12, 13-16	Tenta		

Examinationsformer:

Kunskapskontroller görs under kursen genom inlämningsuppgift och tenta.

Betygssättning:

Studerande ska få sitt kursbetyg senast 3 veckor efter det att kursen avslutats

Betygskriterier:

Kursen bedöms med betygen Icke Godkänd, Godkänd eller Väl godkänd. Grunden för betyg avgörs av hur väl studenten har uppfyllt betygskriterierna.

Studenten ska få sitt kursbetyg senast 3 veckor efter det att kursen avslutats.

Icke godkänt (IG) Den studerande har fullföljt kursen men inte nått alla mål för kursen

Godkänt (G) Den studerande har nått samtliga mål för kursen

Kunskaper

1. Grundläggande statistik och kvantitativa statistiska mått
2. Sannolikhetslära, beroende och oberoende händelser
3. Vanliga fördelningar och dess egenskaper samt användningsområden
4. Estimering och skattning
5. De delar av paketen NumPy, SciPy och Statsmodels som kan appliceras på övriga kunskapsmål

Färdigheter

6. Beräkna informativa statistiska mått, så som kovarians, korrelation och väntevärde
7. Tolka sannolikheter, diagram och statistik
8. Utföra interpolation och extrapolation samt regressionsanalys
9. Utföra hypotestestning och beräkna konfidensintervall

Kompetenser

10. Självständigt samla in data, utföra statistiska beräkningar och identifiera felkällor
11. Kunna kritiskt granska statistik och resultat samt dra egna slutsatser och tolkningar
12. Svara på statistiska frågeställningar genom nyttjande av paketen NumPy, SciPy och Statsmodels

Väl godkänt (VG) Den studerande har nått samtliga mål för kursen. Den studerande kan dessutom:

- Med säkerhet redogöra för regressionsanalys
- Med säkerhet presentera och visualisera statistisk data
- Med säkerhet redogöra för statistisk inferens

Examinerande moment:

Innehåll	Introduktion	Deadline	Täcker upp betygskriterier
Inlämningsuppgift	24/1	3/2	1,4,5,6,9,10,12
Tenta	10/2	10/2	1,2,3,4,6,7,8,9,10,11