



FIK3221 Nätverkssystem för maskininlärning 7,5 hp

Networked Systems for Machine Learning

Fastställande

Kursplanen gäller från och med VT 2025 enligt fakultetsnämndsbeslut: J-2025-3294.
Beslutsdatum: 2025-01-21.

Betygsskala

P, F

Utbildningsnivå

Forskarnivå

Särskild behörighet

Kunskaper i avancerad internetteknik, 7,5 hp, motsvarande slutförd kurs IK2215. Kunskaper och färdigheter i programmering i C++, Java eller Python, 6 hp, motsvarande slutförd kurs DD1310-DD1319/DD1331/DD1337/DD100N/ID1018.

Undervisningsspråk

Undervisningsspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

Lärandemål

Efter godkänd kurs ska studenten kunna:

- beskriva och analysera ett exempel på en tjänst som använder virtualisering av nätverksfunktioner(NFV)
- lista och analysera ett exempel på en tjänst som använder virtualisering av nätverksfunktioner (NFV)
- förklara och särskilja de viktiga fördelarna med RDMA (Remote Direct Memory Access) och hur det fungerar
- analysera metoder för att utföra nätverks-I/O direkt till/från grafikprocessorer (GPU)
- förklara metoder för att förbättra slutledningslatens i detalj
- beskriva och analysera lastbalanseringens roll för servrar
- beskriva och analysera exempel på aktuella forskningsproblem med att betjäna arbetsbelastningar för maskininlärning i datacenter
- tillämpa kunskapen från kursen för att analysera din forskningsdomän och visa dess praktiska användning och inverkan
- analysera sambanden mellan kursmaterialet och din egen forskning, och betona deras betydelse
- argumentera för giltigheten av dessa samband, genom att ge tydliga och evidensbaserade resonemang

Kursinnehåll

Nätverksfunktioner. Virtualisering. Kernel bypass-tekniker (t.ex. DPDK) för nätverk med över 100 gigabit per sekund. Avlastning till Smart Network Interface Cards (SmartNIC). Snabbt nätverk med liten eller ingen CPU-inblandning med hjälp av direkt fjärrminnesåtkomst (RDMA).

Nätverksaspekter av maskininlärningsslutledning med hjälp av grafiska processorer (GPU). Lastuppskattning och belastningsutjämning. Begäran om utskick och schemaläggning. Effektiv, storskalig maskininlärningsslutledning. Slutledning med stora språkmodeller (LLM).

Examination

- PRO1 - Projektuppgifter, 2,5 hp, betygsskala: P, F
- SEM1 - Rapportsammanfattningar, 2,5 hp, betygsskala: P, F
- TEN1 - Skriftlig tentamen, 2,5 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.