

KURS 2 · HVORDAN LÆRER MASKINER? · ØKT 2

Hva lærte treet egentlig?

Ulv eller hund? Treningsdata og testdata

HVA TROR DU?

**Hvis en maskin får se mange bilder av
ulver og hunder, hva tror du den legger
merke til?**

Skriv svaret på en lapp.

3 til 5 minutter

Tren et beslutningstre

1. Dere har tolv dyrekort med merkelapp: ulv eller hund.
2. Lag et tre med høyst tre ja/nei-spørsmål som sorterer alle tolv riktig.
3. Skriv hvor mange treningskort treet sorterer riktig.

6 til 8 minutter

Test treet

1. Sorter de åtte testkortene. Bruk bare treet.
2. Skriv treets svar på arket.
3. Læreren leser opp fasiten. Hvor mange ble riktige?

4 til 5 minutter

TRINN 3 · SAMMENLIGN

Treningssett 1

Riktige på testen:

Treningssett 2

Riktige på testen:

Hva var spørsmål 1 i treet deres?

3 til 4 minutter

Treningsdata og testdata

Treningsdata

Eksemplene systemet lærer av.

Testdata

Nye eksempler systemet ikke har sett.
Bare testen viser om det har lært noe
som virker.

Det enkleste mønsteret vinner

I treningssett 1 var det snøen som skilte kortene, ikke dyret.

Treet var perfekt på treningen og dårlig på testen.

Det har skjedd på ordentlig

Forskere trente med vilje en maskin på 20 bilder der alle ulvene sto i snø.

Den kalte bilder «ulv» når det var snø i bakgrunnen, uansett hvilket dyr som var på bildet.

Snarveilæring

Systemet scorer høyt ved å bruke tilfeldige trekk ved dataene i stedet for det det skulle lære.

Forskere skriver fortsatt om dette. Snarveiene er ofte vanskelige å oppdage.

Fakta: 6 til 10 minutter

DISKUSJON

Hva lærte treet egentlig?

Hvordan kunne dere oppdaget feilen før testen?

Var treningssett 2 perfekt? Hva gikk galt med kort T5?

6 til 10 minutter

Dagens spørsmål

Hva ville du gjort med treningssett 1 for å rette det?

- Nå tenker jeg ...
- Jeg lurer på ...

3 til 5 minutter · kan flyttes til neste økt

Om opplegget

Ribeiro, Singh og Guestrin (2016), «Why Should I Trust You?». Pathak m.fl. (2025), Frontiers in Artificial Intelligence.

Dyrekortene er oppdiktet og forenklet til øvelsen.