

DANMARKS TEKNISKE UNIVERSITET

KURSUSNAVN	INTRODUKTION TIL INTELLIGENTE SYSTEMER
KURSUSNUMMER	02461
HJÆLPEMIDLER	ALLE HJÆLPEMIDLER ER TILLADT
VARIGHED	2 TIMER
VÆGTNING	ALLE DELOPGAVER VÆGTES ENS

INDHOLD

OPGAVE A: CLAIRVOYANCE	2
OPGAVE B: AFSTAND MELLEM TO SÆT AF TAL	3
OPGAVE C: REGRESSION MED EN PERIODISK OG STIGENDE FUNKTION	4
OPGAVE D: PROGRAMMERINGS-PRØVE	5
OPGAVE E: REINFORCEMENT LEARNING	6

BESVARELSE AF OPGAVERNE

Alle spørgsmål skal besvares med et entydigt resultat, som skal angives med understregning sidst i hver besvarelse. Besvarelsen skal altid underbygges med relevante betragtninger og/eller beregninger. Det skal klart fremgå hvilke teorier og formler der tages udgangspunkt i, og alle relevante mellemregninger skal medtages.

En ven hævder at være synsk. For at teste hans evner, foreslår du at lave et videnskabeligt forsøg hvor du tænker på en af jeres fem venner: Anna, Bettina, Christian, Dennis, eller Elinor. Din ven skal gætte hvem du tænker på, hvorefter du kan afsløre om det er korrekt, og forsøget gentages n gange.

Din ven påstår at han i hvert fald kan gætte 30% korrekt. Hvis man gætter tilfældigt vil man kun gætte 20% korrekt i gennemsnit. Efter at have udført forsøget vil I beregne et 95% konfidensinterval for proportionen af korrekte gæt, og hvis intervallet ikke indeholder 20% vil I konkludere at han er synsk.

Spørgsmål A.1: Beregn et estimat af den nødvendige stikprøve-størrelse n .

I ender med at benytte en stikprøvestørrelse på $n = 100$, og din ven gætter rigtigt i $n_x = 29$ tilfælde.

Spørgsmål A.2: Beregn et 95% konfidensinterval for proportionen, og angiv svaret som $0.29 \pm e$ hvor e er bredden af intervallet.

Spørgsmål A.3: Hvis din ven havde gættet korrekt i færre end $n_x = 29$ tilfælde, ville det beregnede konfidensinterval da blive bredere, smallere, eller forblive uændret?

Det beregnede interval indeholder ikke 20%, så derfor erklærer din ven, at det nu er endegyldigt bevist, at han er synsk. Men du mener dog at en så ekstra-ordinær påstand kræver ekstra-ordinær bevisførelse.

Spørgsmål A.4: Angiv mindst to måder hvorpå forsøgs-metoden kunne ændres, således at vi med større statistisk sikkerhed vil kunne fastslå om din ven er synsk.

Opgave B Afstand mellem to sæt af tal

Et *talsæt* er en liste af tal, hvor rækkefølgen ikke har nogen betydning. En måde at måle hvor “ens” to talsæt X og Y er, kan være at beregne summen af den kvadratiske afstand fra hvert tal i sættet X til det nærmeste tal i sættet Y . Matematisk kan det skrives som

$$d(X, Y) = \sum_i \min_j ((X_i - Y_j)^2),$$

hvilket kan implementeres i Python med følgende funktion:

```
def set_distance(X, Y):  
    d = 0  
    for x in X:  
        d += min([(x-y)**2 for y in Y])  
    return d
```

Spørgsmål B.1: Hvis de to sæt er givet ved

$$X = \{5, 8, 6, 9, 6\}, \quad Y = \{3, 5, 5, 4, 7\}$$

hvad er da afstanden $d(X, Y)$?

Vi antager i det følgende at sættene X og Y begge indeholder n tal, og at n kan antage en vilkårlig værdi.

Spørgsmål B.2: Hvis vi tæller anvendelsen af plus (+), minus (-) eller opløftet i anden potens (**2) som én operation hver, hvad er da tidskompleksiteten, $T(n)$, af algoritmen i Python-koden?

Spørgsmål B.3: Hvad er køretidskompleksiteten af Python-algoritmen i store-O notation?

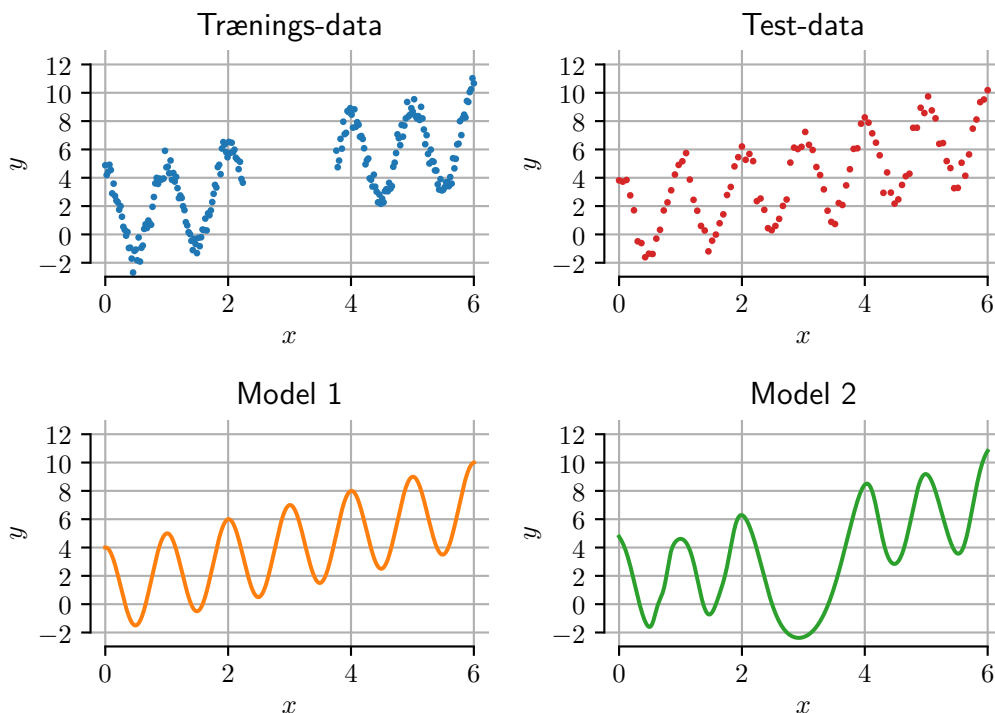
En kollega foreslår at omskrive Python-koden ved hjælp af en list-comprehension, og når frem til følgende funktion, som er mere kort og kompakt, men måske også lidt sværere at læse og forstå:

```
def set_distance_new(X, Y):  
    return sum([min([(x-y)**2 for y in Y]) for x in X])
```

Spørgsmål B.4: Sammenlignet med funktionen `set_distance`, er køretidskompleksiteten $O(n)$ af den nye funktion `set_distance_new` da større, mindre, eller uændret?

Opgave C Regression med en periodisk og stigende funktion

Vi ønsker at finde en model $f(x)$ der kan forudsige værdien af y for et givet punkt x . I plottet øverst til venstre ses trænings-data som vi vil benytte til at fitte modellen. Plottet øverst til højre viser test data som vi efterfølgende vil benytte til at estimere modellens generaliseringsfejl.



Da data ser ud til at være vel beskrevet ved et bias, en lineær komponent, samt en cosinus-komponent med perioden 2π , vil vi fitte følgende funktion, som vi kalder *Model 1*:

$$f(x) = w_0 + w_1 \cdot x + w_2 \cdot \cos(2\pi x)$$

Spørgsmål C.1: Hvilken konkret metode kan passende benyttes til effektivt at estimere parametrene w_0 , w_1 og w_2 ?

Et fit af Model 1 er vist i plottet nederst til venstre, og det oplyses at amplituden af cosinus-komponenten er fundet til $w_2 = 3$.

Spørgsmål C.2: Giv ud fra figuren et estimat af parametren w_0 .

Som et alternativ har vi også fittet et neuralt netværk bestående af ét skjult lag med 50 neuroner og en hyperbolsk tangens aktiveringsfunktion, defineret med følgende Pytorch kode:

```
model2 = torch.nn.Sequential(  
    torch.nn.Linear(1, 50),  
    torch.nn.Tanh(),  
    torch.nn.Linear(50, 1))
```

Denne model, som vi kalder *Model 2*, er fittet til data og vist i plottet nederst til højre. Model 2 er væsentligt mere fleksibel end Model 1, da den har mange flere parametre.

Spørgsmål C.3: Hvor mange parametre har Model 2?

Spørgsmål C.4: Hvilken model har den laveste generaliseringsfejl (dvs. fejl på test-data)?

Vi forestiller os en undersøgelse foretaget blandt studerende på et stort universitet: I alt 100 tilfældigt udvalgte studerende medbringer deres egen computer og tager en svær prøve i programmering. Det noteres hvilken computer-type de har medbragt og om de består prøven. Resultatet er følgende:

	Ikke bestået	Bestået	Total
Mac	40	10	50
Windows	24	24	48
Linux	0	2	2
Total	64	36	100

Resultatet viser at $\frac{10}{50} = 20\%$ af Mac-brugerne, $\frac{24}{48} = 50\%$ af Windows-brugerne og $\frac{2}{2} = 100\%$ af Linux-brugerne bestod prøven.

Spørgsmål D.1: Selv om undersøgelsen viser, at 100% af Linux-brugerne bestod, kan vi *ikke konkludere* at alle ville bestå, hvis universitetet pålagde alle at benytte Linux. Giv *to* argumenter for hvorfor vi ikke kan konkludere dette.

Spørgsmål D.2: Er der statistisk belæg for at konkludere, at mindre end 50% af Mac-brugere blandt hele universitetes population vil kunne bestå prøven?

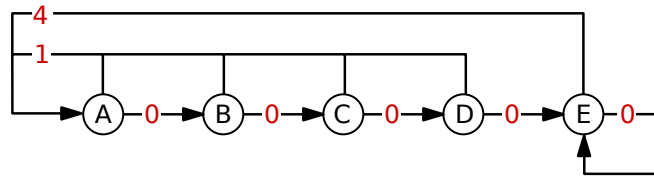
Universitetes administration laver en ny undersøgelse hvor 1000 studerende bliver bedt om at tage prøven på en computer som stilles til rådighed af universitetet. Halvdelen af de studerende udvælges tilfældigt til at tage prøven med en Mac, og den anden halvdel med en Windows computer. Resultatet af den nye undersøgelse er følgende:

	Ikke bestået	Bestået	Total
Mac	324	176	500
Windows	329	171	500
Total	650	350	1000

Spørgsmål D.3: Hvad kan du, ud fra de to undersøgelser, konkludere om den kausale sammenhæng mellem computer type og beståelse af prøven?

Spørgsmål D.4: Giv et eksempel på en mulig confounder, som kan forklare forskellen mellem de to undersøgelser.

Vi betragter følgende miljø som indeholder fem tilstande (states), A–E. I hver tilstand er der to mulige actions: at gå **op** eller til **højre**. Pilene angiver de mulige transitioner mellem tilstandene, og tallene på pilene angiver de tilhørende belønninger (rewards). At gå **op** fører altid til tilstand A.



At gå **op** fra tilstand A–D giver en belønning på 1, og at gå **op** fra tilstand E giver en belønning på 4. Alle andre belønninger er nul.

Spørgsmål E.1: Hvad er den optimale action, når man står i tilstand E?

I de følgende tre delopgaver anvender vi en discount factor på $\gamma = 0.9$. Det oplyses at værdien af tilstandene A og E er henholdsvis $v(A) = 10$ og $v(E) = 13$. Det oplyses endvidere at den optimale action, når man står i tilstand B, er at gå **op**.

Spørgsmål E.2: Hvad er værdien af tilstand B, $v(B)$?

Spørgsmål E.3: Hvad er værdien af tilstand D, $v(D)$?

Spørgsmål E.4: Beskriv den optimale policy, hvis man i stedet anvender en discount faktor på $\gamma = 0$.