

Kommenterat facit

Hur fungerar intelligenta maskiner?

Artificiell intelligens nivå 1 · Viktor Arohlén

För eleven

Svarsexemplen är inte texter att memorera. Andra formuleringar fungerar om de visar samma förståelse. De angivna svarslängderna hjälper dig att se uppgiftens omfattning; tydliga svar behöver inte vara långa.

Flervalsfrågornas svar

1a	2c	2d	3b	6b	6c	7b	8c
B	D	A	C	B	D	A	C

1. Regler eller träning?

a) B. A:s sorteringsregel är skriven av en människa. B har utvecklats genom träning på märkta exempel.

- A blandar ihop de två sätten att utveckla systemet.
- C blandar ihop märkning under träning med användning av ett färdigt system.
- D blandar ihop att lagra data med att faktiskt använda den för att träna modellen.

b) Exempel på ett tillräckligt svar:

Regeln reagerar på ordet "vinst", så mejlet hamnar i skräppost trots att det är en vanlig hälsning från en vän. Regeln tar inte hänsyn till sammanhanget som ordet används i.

Det viktiga är att koppla felet till just regeln och mejlets innehåll. Det tränade systemet är inte automatiskt felfritt.

2. Regression och modellens förutsägelser

a)

Uppgift	Typ av uppgift	Typ av utdata
Antal personer som kommer till en match	Regression	Ett uppskattat antal
Positiv eller negativ spelrecension	Klassificering	En klass
Vinst eller förlust, kodade som 1 och 0	Klassificering	En klass

En klass kan representeras med ett tal. Det är vad svaret **betyder** som skiljer klassificering från regression, inte bara om ett tal syns på skärmen. En framtida vinst/förlust är också en förutsägelse, men uppgiften är klassificering.

b) Indata: **antal personer i kön**. Utdata: **förutsagd väntetid i minuter**.

$2 \times 4 + 3 = 11$ minuter. Det är modellens förutsägelse, inte ett löfte om exakt väntetid.

c) **D**. Medelfelet beskriver den genomsnittliga storleken på skillnaden mellan förutsagd och verklig väntetid. Det anger varken den största avvikelsen eller om modellen främst gissar för högt eller för lågt.

- A beskriver ett genomsnitt av modellens svar, inte dess fel.
- B beskriver det största felet, inte genomsnittet.
- C säger att felet har en viss riktning. Ett medelfel på två minuter säger inte om modellen främst gissar för högt eller för lågt.

Ingen beräkning av medelfelet krävs i denna fråga.

d) **A**. 25 personer ligger utanför intervallet 2–10 personer i träningsdatan. Modellen extrapolerar, och mönstret från kortare köer behöver inte gälla där.

- B: en ny indata är inte i sig överanpassning.
- C: modellen uppskattar fortfarande en tid, inte en kökategori.
- D: att använda modellen på en ny kö innebär inte ny träning.

3. Träning och användning

a) Parametrar och övervakad inlärning.

Även **vikter** för den första luckan och **övervakat lärande** för den andra fungerar. Parametrar är det bredare begreppet; vikter är exempel på parametrar.

b) **C.** Den tränade modellen använder sina inlärdade parametrar för att beräkna ett svar på ny indata. Användning och träning är olika steg.

- A blandar ihop användning med att ändra parametrarna.
- B förväxlar omdömena om träningslåtarna med ett facit för den nya låten.
- D antar att modellens eget svar automatiskt blir ett korrekt facit och leder till ny träning.

En tjänst kan vara byggd för fortsatt träning, men detta följer inte enbart av att den besvarar ny indata.

4. Vikter i ett vardagligt exempel

a) $30 \times 2 + 15 \times 1 = 75$ kronor.

b) Extra godispåse: **30 kronor**. Extra läskflaska: **15 kronor**.

c) **Exempel på ett tillräckligt svar:**

En vikt bestämmer hur mycket varje extra enhet av en indata bidrar till summan. I ett neuronnät justeras vikterna med hjälp av träningsexempel för att minska modellens fel.

I exemplet ökar en extra godispåse summan med 30 kronor och en extra läskflaska med 15 kronor. Här är vikterna givna priser; prisformeln är inte en tränad AI-modell bara för att den använder vikter. Jämförelsen gäller en extra enhet när övrigt är oförändrat; det totala bidraget beror också på antalet köp. Ingen beräkning av förhållandet mellan priserna krävs.

5. En enkel perceptron

a)

Foto	Viktad summa	Utdata	Modellens svar
A	4	1	Mugg
B	2	0	Glas
C	-1	0	Glas

Summan måste vara **större än 2** för att ge 1. En summa som är exakt 2 ger därför 0. Den negativa summan kräver ingen beräkning; även den ligger under tröskeln.

b) **Tröskelvärde.** Det används som gräns när modellen väljer utdata och klass. Den viktade summan är en mellanberäkning, inte svaret som modellen lämnar här.

6. Neuronnät, generativ AI och utvärdering

a) Orden i ordning är **indata** → **lager** → **lager** → **utdata**. Resultat från ett lager blir indata till nästa lager. Det sista steget ger modellens svar. Samma princip kan visas med en egen tydlig skiss.

b) **B**. Att skapa text och att skapa bilder är två tillämpningar av generativ AI. En stor språkmodell kan generera text; diffusionsmodeller är exempel på modeller som kan generera bilder.

- A blandar ihop typen av instruktion med vilken sorts modell som används för att skapa utdatan.
- C blandar ihop tjänstens gränssnitt med modellernas uppbyggnad.
- D beskriver uppslagning av sparad innehåll, inte generering av nytt innehåll.

Frågan säger inte att alla bildmodeller är diffusionsmodeller eller att en viss tjänst använder diffusion. Ingen beskrivning av tokens eller diffusionsprocessen behövs.

c) **D**. B:s medelfel på testdatan är 2 minuter, jämfört med A:s 4 minuter. B ger alltså bättre förutsägelser på dessa nya exempel. A passar träningsdatan perfekt men får större fel på testdatan, vilket kan tyda på överanpassning. Fler parametrar ger inte automatiskt bättre förutsägelser på nya exempel.

- A bedömer användbarheten på nya köer enbart från träningsresultatet.
- B gör det mindre antalet parametrar till en förklaring som jämförelsen inte bevisar.
- C blandar ihop att mäta modellens fel på testdata med att ändra parametrarna under träning.

Testet visar inte att B kommer att vara bäst för alla tänkbara köer, och det bevisar inte orsaken till A:s sämre resultat. Uppgiften kräver bara tolkning av de givna medelfelen, inte nya beräkningar.

7. Fungerar bildmodellen i ett annat sammanhang?

a) Exempel på ett tillräckligt svar:

Modellen kan ha använt bakgrunden som ledtråd, eftersom muggarna hade vit vägg och glasen hade bokhylla. När den kopplingen ändras kan den välja fel klass.

En relevant annan möjlig förklaring kan godtas om den kopplas till insamlingen och de nya bilderna. Formuleringen "kan ha" skiljer en möjlig förklaring från något vi redan vet säkert.

b) **A.** Båda föremålen visas mot båda bakgrunderna. Då kan vi undersöka om modellens svar följer föremålet eller bakgrunden, medan andra förhållanden hålls ungefär lika.

- B återanvänder träningsexempel och prövar inte den ändrade kopplingen mellan bakgrund och klass.
- C ändrar flera saker samtidigt, så det blir svårare att koppla skillnaden till just bakgrunden.
- D förändrar träningen och mäter träningsresultat i stället för att pröva den aktuella modellens användning av bakgrunden.

c) Exempel på ett tillräckligt svar:

Modellen svarar "mugg" mot den vita väggen och "glas" mot bokhyllan, även när föremålet byts. Det stöder förklaringen att bakgrunden påverkar svaret. Jag skulle samla bilder av både muggar och glas mot flera gemensamma bakgrunder, så att bakgrunden inte följer en viss klass.

Svaret kopplar tabellens mönster till förklaringen och föreslår en relevant förändring. Ett mer utvecklat svar kan nämna att fyra foton är ett begränsat underlag och att modellen efter ny träning behöver jämföras på nya testbilder. Det är sätt att nyansera resonemanget, inte extra obligatoriska deluppgifter i denna fråga.

Att bara samla fler muggbilder mot vit vägg och fler glasbilder mot bokhyllan bryter inte kopplingen mellan bakgrund och klass.

8. Dagsljusmodellen och människan

a) Exempel på ett tillräckligt svar:

Modellen har anpassat en rät linje till data från den del av året då dagarna blir längre. Människan kan använda kunskap om årstiderna och vet att dagsljuset senare minskar i Sverige. Jag skulle använda data från hela året och en modell som kan beskriva både ökningen och minskningen.

Förklaringen kopplar felet till modellens begränsade underlag eller form och jämför det med människans kunskap om sammanhanget. En relevant förbättring kan vara mer varierad data från hela året, en modellform som kan följa årstidernas förändring eller att begränsa användningen till den period modellen fungerar för. En rät linje kan inte följa hela årets ökning och minskning enbart genom fler träningspunkter.

b) Exempel på ett tillräckligt svar:

Modellen kan snabbt göra många uppskattningar under våren om den har testats och fungerar för den perioden. För november skulle jag vilja att en människa kontrollerar om modellen är lämplig och använder kunskap om årstiderna eller söker rätt dagsljusdata. Människans kunskap hjälper till att upptäcka problemet, men ger inte automatiskt en exakt dagsljuslängd.

Många beräkningar kan göras snabbt med modellen, men det gör inte en olämplig modell korrekt. Människan är inte heller garanterat rätt. Det räcker att motivera när modellen kan vara användbar och när dess begränsning behöver upptäckas av en människa.

c) **C.** De gav samma utdata för just dessa datum. Det visar inte att de använder samma arbetssätt eller fungerar lika bra under andra delar av året.

- A drar en slutsats om arbetssätt från enbart överensstämmande svar.
- B drar en generell slutsats om andra delar av året från ett begränsat test.
- D förväxlar att två bedömningar stämmer överens med att de har jämförts med rätt svar. Båda kan ha fel.