

# Instuderingsfrågor maskininlärning AI21

---

Facit finns inte till nedanstående frågor. Frågorna ska utgöra ett stöd för er att lära er teori för maskininlärning. Samarbeta gärna och diskutera gärna med varandra.

---

1. Vad är data leakage, varför är det dåligt och hur kan du undvika data leakage?

---

2. Beskriv vad poängen kan vara att dela upp datasetet till träningsdata och testdata för maskininlärning?

---

3. Beskriv vad poängen kan vara att dela upp datasetet till tränings-, validerings- och testdataset för maskininlärning?

---

4. Hur skiljer sig batch gradient descent och mini batch gradient descent?

---

5. Vad är en aktiveringsfunktion i en artificiell neuron? Ge två exempel på aktiveringsfunktioner.

---

6. Du har ett program som klassificerar email till spam eller hams (inte spam). Vilka evalueringsmetrics ska du kolla på och varför?

---

7. Ett brandlarm larmar ofta när det inte brinner. Tror du precision är högt eller lågt, förklara varför.

---

8. Förklara vad curse of dimensionality är och hur det kan påverka en maskininlärningsmodell.

---

9. Vad är skillnaden på feature standardization och normalization?

---

10. Du har ett dataset som ser ut som följande (1000 observationer):

temperatur i °C	vind (m/s)
22.5	2.0
8.3	10.3
13.2	4
...	...

En meteorolog vill gruppera dessa. Ge ett förslag på hur hen skulle kunna göra med en maskininlärningsalgoritm.

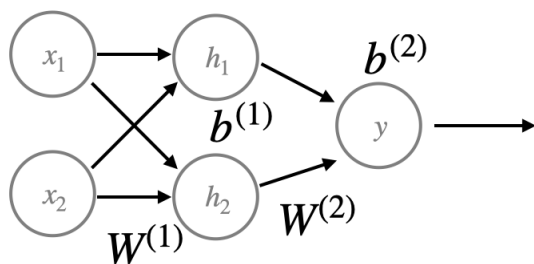
---

11. Många maskininlärningsmodeller har svårt att arbeta med textdata direkt. Ge förslag på hur man skulle kunna förbearbeta en textsträng.
- 
12. Du ska använda KNN för klassificering. Ge förslag på hur du skulle kunna gå tillväga för att bestämma ett lämpligt värde på antalet neighbors.
- 
13. Beskriv begreppen: perceptron, multilayered perceptron, feedforward network, fully connected layers.
- 
14. Aktiveringsfunktionen rectified linear unit (ReLU) är definierad som  $f(z) = \max(0, z)$ . En nod i ett MLP har denna aktiveringsfunktion och får inputs  $\vec{x} = (1, 2, 3)^T$  och har vikterna  $\vec{w} = (0, 2, 1)^T$  och bias  $b = -2$ . Beräkna output  $y$  för denna nod.
- 
15. XOR grind i digitalteknik är en exklusiv disjunktion vilket innebär:

A	B	A XOR B
1	1	0
1	0	1
0	1	1
0	0	0

Visa att följande MLP kan simulera XOR-grind. Varje nod är en perceptron, dvs att aktiveringsfunktionen är en stegfunktion.

$$W^{(1)} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, W^{(2)} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}, b^{(1)} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, b^{(2)} = 0$$



16. Formeln för Naive Bayes är:

$$y = \arg \max_{k \in 1, \dots, K} P(C_k) \prod_{i=1}^n P(x_i | C_k)$$

Du har ett dataset som består 100 reviews, varav 80 var positiva och 20 negativa. Nedan finns tabeller på hur frekvent olika ord förekommer för positiva respektive negativa reviews:

**Positiva reviews:**

	happy	good	book	author	bad
Antal	50	60	20	10	1

**Negativa reviews:**

	happy	good	book	author	bad
Antal	2	5	40	40	40

Vi får ett review med texten "happy bad", använd Naive Bayes och klassificera den till positiv eller negativ review.

17. Beskriv kort skillnaderna mellan decision tree och random forest.
18. Ge ett exempel på ett maskininlärningsproblem där man kan applicera logistisk regression.
19. Beskriv kort skillnader mellan supervised learning och unsupervised learning.
20. Beskriv kort skillnader mellan regression och klassificering.