COMP5318/COMP4318 Machine Learning and Data Mining
Semester 1, 2024

Information about the exam

- The exam is in-person, paper-based. This is the only option. There is no online exam option.
- Permitted materials:
    - Calculator - handheld, non-programmable
- No other materials or devices are allowed
- You need to have a photo identification - Student ID card or other accepted documents
- The exam paper is like a booklet. You write your answers on the exam paper, in the space provided. There is a question, then a space for you to write the answer; another question and space for the answer, etc. To write the answers you should use black or blue pen, not pencil.

- The exam paper is **confidential**. You must not discuss the exam questions with other people, post or distribute the exam questions in any way **after the exam**.
- The duration of the exam is standard: 2 hours + 10 minutes reading time.
- The exam is worth 100 marks (=60% of your final mark). To pass the course you need at least 40% on the exam (i.e. 40 marks), regardless of what your mark during the semester is. This is a school rule.
- There are 3 types of questions: 1) multiple-choice questions (about 20% of the exam), 2) questions requiring short answers, 3) problem-solving/calculation questions.

Sample exam questions

Question 1. (Multiple choice question)

Select the correct answer.

1. Leave-one-out cross validation is suitable for large data sets.

a) True
b) False

2. The regression line minimizes the sum of the residuals

a) True
b) False

3. A single perceptron can solve the XOR problem.

True False

Question 2. (Short answer question)

1. Why do we need to apply normalization when using distance-based algorithms such as k-Nearest Neighbor?

Because we need to make sure all attributes in the same scale, then some distance will not overwhelm other

2. In linear support vector machines, we use dot products both during training and during classification of a new example. What vectors are these products of?

During training: two training data
During classification of new example: support vectors and new data

3. List one disadvantage of applying a multi-layer perceptron neural network to perform handwritten digits image classification.

It can not deal with moving images. (CNN can solve it)

Such as $\boxed{1}$ and $\boxed{\quad 1}$ will be regard as different

Calculation (problem solving) questions

Question 3. Decision tree

Given is the following training data where ***location, weather*** and ***expensive*** are the features and ***holiday*** is the class.

location	weather	expensive	holiday	
nice	sunny	Y	good	+
nice	sunny	N	bad	-
boring	rainy	Y	good	+
boring	sunny	N	bad	-
nice	rainy	Y	good	+
boring	rainy	N	good	+
boring	rainy	N	good	+

7 datas in total
2 - 5 +

a) What is the entropy of this set of training examples with respect to the class?
b) We would like to build a decision tree using information gain. Which attribute will be selected as a root of the tree? Show your calculations.

You may use this table:

x	y	-(x/y)* log₂(x/y)	x	y	-(x/y)* log₂(x/y
1	2	0.50	1	6	0.43
1	3	0.53	5	6	0.22
2	3	0.39	1	7	0.40
1	4	0.5	2	7	0.52
3	4	0.31	3	7	0.52
1	5	0.46	4	7	0.46
2	5	0.53	5	7	0.35
3	5	0.44	6	7	0.19
4	5	0.26			

(a) $E = -\frac{2}{7}\log_2\frac{2}{7} - \frac{5}{7}\log_2\frac{5}{7}$

$= 0.52 + 0.35 = 0.87$

(b) weather

Question 4. Naïve Bayes

Given is the following training data where ***location, weather***, ***companion*** and ***expensive*** are the features and ***holiday*** is the class.

location	weather	companion	expensive	holiday
nice	sunny +	annoying +	Y +	good
nice	sunny +	annoying +	N	bad
boring +	rainy	great	Y +	good
boring +	sunny +	great	Y +	bad
nice	rainy	great	Y +	good
boring +	rainy	annoying +	N	good
boring +	rainy	great	N	good

$P(good|E) = \frac{\frac{5}{7} \times \frac{3}{5} \times \frac{1}{5} \times \frac{2}{5} \times \frac{3}{5}}{P(E)}$ $\frac{18}{875}$

$P(bad|E) = \frac{\frac{2}{7} \times \frac{1}{2} \times \cancel{\frac{2}{2}} \times \frac{1}{2} \times \frac{1}{2}}{P(E)} = \frac{1}{28}$ $\cancel{\frac{1}{56}}$

Use Naïve Bayes to predict the value of ***holiday*** for the following new example, showing your calculations: ***location=boring, weather=sunny, companion=annoying, expensive=Y***.

$P(bad|E)$

3. (b) Location:

Location
nice / boring
(S_1) ggb   gggb (S_2)

$$T_2 = \frac{3}{7}H(S_1) + \frac{4}{7}H(S_2) = \left[-\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}\right]\frac{3}{7} + \frac{4}{7}\left[-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}\right]$$

$\frac{69}{175}$ $\frac{81}{175}$

$$= (0.39 + 0.53) \times \frac{3}{7} + \frac{4}{7} \times [0.31 + 0.5]$$

$$= \frac{6}{7}$$

$$\text{Gain} = T_1 - T_2 = 0.87 - \frac{6}{7} = 0.01286$$

Weather:

weather
Sunny / Rainy
(S_1) gbb   gggg (S_2)

$$T_2 = \frac{3}{7}H(S_1) + \frac{4}{7}H(S_2) = \frac{3}{7} \times \left[-\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}\right] + 0$$

$$= \frac{69}{175}$$

$$\text{Gain} = 0.87 - \frac{69}{175} = 0.4757$$ biggest

Expensive:

E
Y / N
(S_1) ggg   bbgg (S_2)

$$T_2 = \frac{3}{7}H(S_1) + \frac{4}{7}H(S_2)$$

$$= 0 + \frac{4}{7} \times \left[-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right]$$

$$= \frac{4}{7}$$

$$\text{Gain} = 0.87 - \frac{4}{7} = 0.29857$$

choose weather

Question 5. 1R

Given the training data in the table below where ***credit history, debt, deposit*** and ***income*** are attributes and ***risk*** is the class, predict the class of the following new example using the 1R algorithm: *credit history=unknown, debt=low, deposit=none, income=average.* If needed, settle ties by random selection. Show your calculations.

? X

credit history	debt	deposit	income	risk
bad	high	none	low	high
unknown	high	none	average	high
unknown	low	none	average	moderate
unknown	low	none	low	high
unknown	low	none	high	low
unknown	low	adequate	high	low
bad	low	none	low	high
bad	low	adequate	high	moderate
good	low	none	high	low
good	high	adequate	high	low
good	high	none	low	high
good	high	none	average	moderate
good	high	none	high	low
bad	high	none	average	high

Question 6. Perceptron

Given is the following training set:

	input	output
ex. 1:	1 0 0	1
ex. 2:	0 1 1	0
ex. 3:	1 1 0	1
ex. 4:	1 1 1	0
ex. 5:	0 0 1	0

$W_{new} = W_{old} + e x^T$

$b_{new} = b_{old} + e$

a) Train a perceptron **with a bias** on this training set. Assume that all initial weights (including the bias of the neuron) are 0. Show the set of weights (including the bias) at the end of the first epoch. Apply the examples in the given order.

Recall that the perceptron uses a step function defined as:
step(n) = 1, if n >= 0
= 0, otherwise.

(a) $W = [0, 0, 0]$ $b = 0$

ex 1. $1 \times 0 + 0 \times 0 + 0 \times 0 = 0 \geq 0$ 1 correct

ex 2. . . . $= 0 \geq 0$ 1 incorrect

$0 - 1 = -1$

$e = t - a = \cancel{1 - 0 = 1}$

$w = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \cancel{+} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix} \quad b = 0 + 1 \cancel{= 1} \; -1$

ex3. $\begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 1 = \cancel{2}\ 0 \geq 0$ so it's 1 correct

ex4. $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 1 = 3 \geq 0$ so it's 1 incorrect

$e = t - a = 0 - 1 = -1$

$w = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + -1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \quad b = 1 - 1 = 0$

ex5. $\begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + 0 = 0$ so it's 1 incorrect

$e = t - a = 0 - 1 = -1$

$w = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix}}_{} \quad b = 0 \underbrace{-1 = -1}_{}$

final

ex_1 : $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$ $\quad t_1 = 1$

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 0 = 0 \quad ; \ step(0) = 1 = a \ , \text{ correct}$$

ex_2 $\begin{bmatrix} 0, 1, 1 \end{bmatrix}$ $\quad t_2 = 0$

$$\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + 0 = 0 \ , \quad step(0) = 1 = a \quad \text{incorrect}$$

$$e = t_2 - a = -1$$

$$w = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} - \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix} \quad b = 0 - 1 = -1$$

ex_3 $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ $\quad t_3 = 1$

$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix} - 1 = -2 \ , \quad step(-2) = 0 \quad \text{incorrect}$$

$$e = t_3 - a = 1$$

$$w = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad b = -1 + 1 = 0$$

ex_4 $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ $\quad t_4 = 0$

$$\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \times \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 0 = 0 \ , \ step(0) = 1 \ , \ \text{incorrect}$$

$$e = t_4 - a = -1$$

$$w = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ -2 \end{bmatrix} \quad b = 0 - 1 = -1$$

ex_5 correct

Question 7. K-means clustering

Suppose that we are given 7 examples to cluster: A, B, C, D, E, F and G. The distance between them is given by the following matrix:

	A	B	C	D	E	F	G
A	0	10	2	1	12	5	4
B	10	0	4	3	6	23	7
C	2	4	0	5	9	14	19
D	1	3	5	0	1	7	4
E	12	6	9	1	0	2	18
F	5	23	14	7	2	0	6
G	4	7	19	4	18	6	0

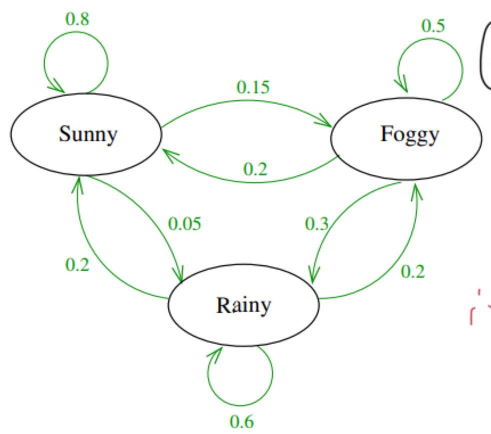
	A	B
C	2 ✓	4
D	1 ✓	3
E	12	6 ✓
F	5 ✓	23
G	4 ✓	7

Run the **k-means** algorithm to group these examples into 2 clusters for 1 epoch. The initial centroids are A and B. Show the resulting clusters.

{A, C, D, F, G} {B, E} ✓

Question 8. Markov models

Given is the following Markov model for the weather in Sydney:



(a) $P(\pi_3, \pi_2 \mid \pi_1) = P(\pi_3 \mid \pi_2, \pi_1) \cdot P(\pi_2 \mid \pi_1)$

$= 0.05 \times 0.8$

$= 0.040$

by markov it is equivalent to $P(\pi_3 \mid \pi_2)$

a) Given that today the weather is *Sunny*, what is the probability that it will be *Sunny* tomorrow and *Rainy* the day after tomorrow, i.e. what is the probability $P(\pi_3 = Rainy,\ \pi_2 = Sunny \mid \pi_1 = Sunny)$?

Hint: P(A,B|C) = P(A|B,C) P (B|C)

b) If the weather yesterday was *Rainy*, and today is *Foggy*, what is the probability that tomorrow it will be *Sunny*?

$1 \times 1 \times 0.2 = 0.2$

For both questions, briefly show your calculations.

✓✓

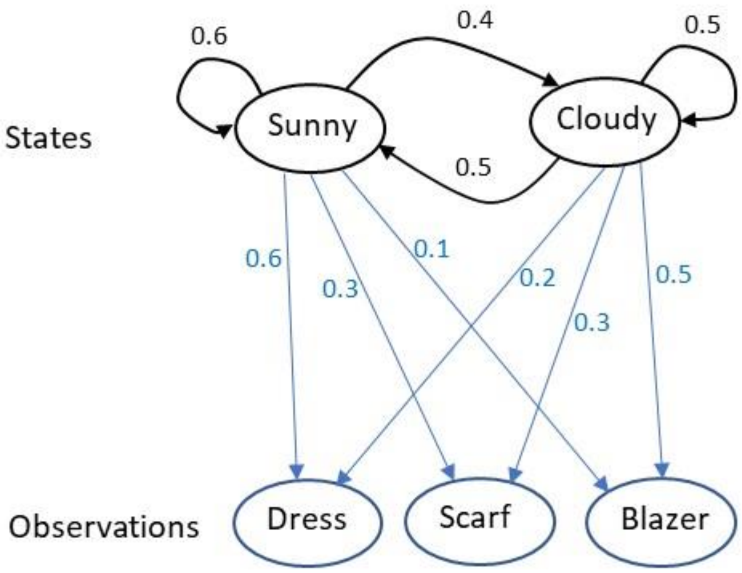
Since ~~they~~ are given, so the P should be 1

By markov only see previous one day

Question 9. Hidden Markov models

Julia tested positive to COVID and had to quarantine at home for several days. Her friend Nicole came to bring her food every day. We don't know what the weather was on the quarantine days but we know the type of clothing Nicole wore and it provides evidence about the weather.

The following Hidden Markov Model models the situation. The initial state probabilities are: A_0(Sunny)=0.5 and A_0(Cloudy)=0.5.



x_1 $\quad$ x_2

dress $\quad$ blazer

Suppose that on the first quarantine day Nicole wore a dress and on the second she wore a blazer.

a) What is the probability of the observation sequence?
b) What is the most likely sequence of hidden states?

Briefly show your calculations.

(a) HHM, ,

$$f_s(1) = A^0(s)\, e_s(x_1) = 0.5 \times 0.6 = 0.3$$

$$f_c(1) = A^0(c)\, e_c(x_1) = 0.5 \times 0.2 = 0.1$$

$$f_s(2) = e_s(x_2)\left[f_s(1) \times a_{s\to s} + f_c(1) \times a_{c\to s}\right]$$

$$= 0.1 \times [0.3 \times 0.6 + 0.1 \times 0.5] = 0.023$$

$$f_c(2) = e_c(x_2)\left[f_c(1) \times a_{c\to c} + f_s(1) \times a_{s\to c}\right.$$

$$= 0.5 \times [0.1 \times 0.5 + 0.3 \times 0.4] = 0.085$$

$$P(x) = f_s(2) + f_c(2) = 0.108$$

(b) $V_s(1) = 0.3$

$V_c(1) = 0.1$

$V_s(2) = 0.1 \times \max\{0.3 \times 0.6, 0.1 \times 0.5\}$

$= 0.1 \times 0.18$

$= 0.018$

$Ptr_s(2) = \text{Sunny}$

$V_c(2) = 0.5 \times \max\{0.1 \times 0.5, 0.3 \times 0.4\}$

$= 0.5 \times 0.12 = 0.06$

$Ptr_c(2) = \text{sunny}$

$\pi_2 = \text{argmax}\{V_c(2), V_s(2)\}$

$= \text{cloudy}$

$\pi_1 = Ptr_{\text{cloudy}}(2) = \text{sunny}$

sunny → cloudy