State Final Examination (Sample Questions)

2023-06-27

1 Mergesort (3 points)

1. Formulate the algorithm for sorting a sequence of n integers by merging (Mergesort). Write it in pseudo-code.
2. Analyze time and space complexity of this algorithm.
3. Design an efficient algorithm for merging k sorted sequences with n elements total. Change Mergesort to use this kind of merging. How does this affect the time complexity? Express the complexity as a function of variables n and k.

2 Converting a context-free grammar to an automaton (3 points)

1. Give a definition of a *context-free grammar*.
2. Construct a context-free grammar G generating the following language:

$$L = \{a^i b^j \mid i, j \geq 0 \text{ and } 2i = 3j\}$$

3. Convert the grammar G from the previous part to a pushdown automaton accepting by empty stack. (If you were unable to construct the grammar, construct any pushdown automaton accepting the language L by empty stack.)

3 Linear algebra (3 points)

Consider matrices

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}.$$

1. Find $x \in \mathbb{R}^2$ such that two vectors Ax and Bx are perpendicular (in the standard inner product).
2. Find the matrix representation of the linear transformation that maps vector Ax to vector Bx for any $x \in \mathbb{R}^2$.
3. Decide whether the linear transformation $x \mapsto B^{-1}A^3B^{-1}x$ increases or decreases the areas of geometric objects in $\mathbb{R}^2$.

4 Probability (3 points)

1. Write the theorem about linearity of expectation (do not prove it).
2. In a group of n people, everyone rolls a (fair) six-sided die. Each pair of people who roll the same number gets one point. Let Y denote the number of points scored by one pre-selected pair. Determine the expectation of Y.
3. Denote by X the total number of points scored by all pairs together (one person can score in more than one pair). Determine the expectation of X.

5 Hidden Markov Model (specialization question – 3 points)

Define Hidden Markov Model (HMM). What is the Markov assumption? What is the difference between filtering, prediction, and smoothing? Using the formula

$$P(X_t|E_{1:t}) = \alpha P(E_t|X_t) \sum_{X_{t-1}} P(X_t|X_{t-1}) P(X_{t-1}|E_{1:t-1})$$

solve the following example:

Let's assume that the variable X_i represents the weather on day i. It can either be raining ($X_i = R$) or sunny ($X_i = S$). We cannot directly observe this variable; we only have access to the observations E_i of whether our colleague brought an umbrella today ($E_i = U$) or not ($E_i = N$). The conditional probabilities of weather change from the previous day to today and the dependence of umbrella observation on the actual weather are given in the following tables.

X_{i-1}	$P(X_i = R \mid X_{i-1})$	$P(X_i = S \mid X_{i-1})$
R	0.7	0.3
S	0.3	0.7

X_i	$P(E_i = U \mid X_i)$	$P(E_i = N \mid X_i)$
R	0.9	0.1
S	0.2	0.8

Let's assume that on day 0, the probability of rain is $P(X_0 = R) = 0.5$, and the probability of sunshine is $P(X_0 = S) = 0.5$. What is the probability of rain on the second day ($P(X_2 = P) =?$) if we observe that our colleague has brought an umbrella on both the first and second days ($E_1 = U$ and $E_2 = U$)?

6 Binary classifier evaluation (specialization question – 3 points)

To test a binary classifier, a data set consisting of 100 positive and 400 negative examples was used.

a) It turned out that the ROC curve goes through the point TPR = FPR = 0.2. Calculate the precision at this point.

b) Another classifier shows the constant precision equal to 0.5. Draw the ROC curve on that assumption.

Remark on abbreviations: ROC — Receiver Operating Characteristic curve; TPR — true-positive rate (also: sensitivity, or recall); FPR — false-positive rate (also: fall-out).

7 Decision trees (specialization question – 3 points)

We want to train a decision tree on the data given in the table below that contain two categorical attributes x_1 and x_2 and a categorical target value y.

Hint: While solving this problem, you may need the values of some logarithms. You can use the following approximations: $\log_2 \frac{1}{3} = -1.5$, $\log_2 \frac{2}{3} = -0.6$

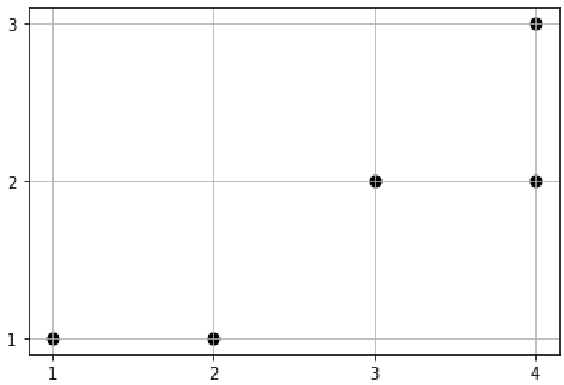
x_1	x_2	y
0	0	0
1	0	1
0	1	1
2	1	0
0	0	1
1	1	1
2	0	0
0	1	0

1. Describe briefly the algorithm for training decision trees and the branching criteria.
2. Which attribute will be used for branching in the root of the tree? Why?
3. Is it possible to classify the training data with 100 % accuracy? Explain.

8 Clustering (specialization question – 3 points)

Consider the data with two numerical attributes in the table below that are also plotted in the figure. We want to apply clustering algorithms to them.

x_1	x_2
1	1
2	1
3	2
4	3
4	2



1. First, we want to apply the k-means algorithm with $k = 2$. Initially, we randomly selected $(1, 1)$ and $(3, 2)$ as the cluster centers. How will the cluster centers change after one iteration of the k-means algorithm?
2. How will the cluster centers change in additional iterations of the algorithm?
3. Describe a hierarchical clustering technique. What would its result be with the given data?

9 Smoothing for language models (specialization question – 3 points)

Explain the notion of smoothing in language modelling, and give a concrete example of a situation in which using an unsmoothed language model would lead to a problem in an NLP application. Give examples of at least two techniques used for language model smoothing. For one of the chosen smoothing techniques, illustrate how it would change an estimated probability of the bigram “the cat” (more exactly an estimate of the conditional probability $P(w_i = cat \mid w_{i-1} = the)$), assuming that the training corpus contains this single sentence: “There is a cat on the table.”

10 Binary classifier evaluation (specialization question – 3 points)

To test a binary classifier, a data set consisting of 100 positive and 400 negative examples was used.

a) It turned out that the ROC curve goes through the point TPR = FPR = 0.2. Calculate the precision at this point.

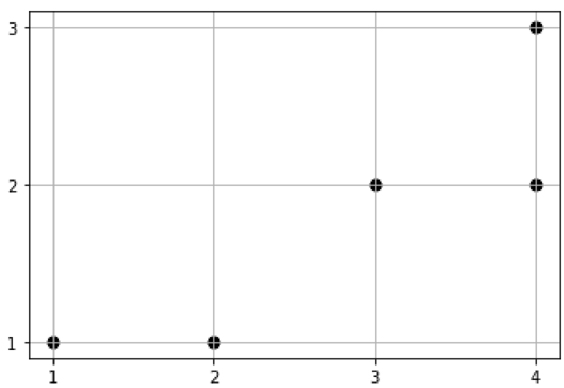
b) Another classifier shows the constant precision equal to 0.5. Draw the ROC curve on that assumption.

Remark on abbreviations: ROC — Receiver Operating Characteristic curve; TPR — true-positive rate (also: sensitivity, or recall); FPR — false-positive rate (also: fall-out).

11 Clustering (specialization question – 3 points)

Consider the data with two numerical attributes in the table below that are also plotted in the figure. We want to apply clustering algorithms to them.

x_1	x_2
1	1
2	1
3	2
4	3
4	2



1. First, we want to apply the k-means algorithm with $k = 2$. Initially, we randomly selected $(1, 1)$ and $(3, 2)$ as the cluster centers. How will the cluster centers change after one iteration of the k-means algorithm?
2. How will the cluster centers change in additional iterations of the algorithm?
3. Describe a hierarchical clustering technique. What would its result be with the given data?