

بسمه تعالی

تکلیف سری دوم درس داده کاوی

- 1- خوشه بندی سلسله مراتبی را توضیح دهید.
- 2- شبکه عصبی پرسپترون برای چه کاربردهایی مناسب است؟ توضیح دهید.
- 3- فرض کنید یک مجموعه داده به صورت زیر دارید که شامل مختصات دو بعدی از نقاط مختلف است، الگوریتم KMeans را برای خوشه بندی این داده ها با $K=2$ (دو خوشه) اجرا کنید.

نقطه	x	y
1	1	2
2	2	3
3	3	3
4	6	5
5	7	8
6	8	8

- 4- فرض کنید مجموعه داده زیر برای پیش بینی اینکه آیا یک فرد در یک روز بارانی به پیاده روی می رود یا نه، موجود است، با استفاده از درخت تصمیم با استفاده از الگوریتم ID3، پیش بینی کنید که اگر دمای روز 27 درجه، رطوبت 80 درصد و باد زیاد باشد، آیا فرد به پیاده روی می رود یا خیر؟

به پیاده روی می رود؟	باد	رطوبت	دما
بله	کم	80	30
نه	زیاد	70	25
بله	کم	75	28
نه	کم	85	20
بله	زیاد	80	22

5- فرض کنید مجموعه داده‌ی زیر برای پیش‌بینی تأیید وام بانکی (Approve / Reject) در اختیار شما قرار دارد. بر اساس شاخص جینی، مشخص کنید کدام ویژگی باید به عنوان ریشه درخت تصمیم انتخاب شود.

تایید وام	درآمد	اعتبار مشتری	مشتری
Approve	High	Good	1
Approve	High	Good	2
Approve	Low	Good	3
Approve	Low	Good	4
Approve	High	Bad	5
Reject	Low	Bad	6
Reject	Low	Bad	7
Reject	Low	Bad	8

6- گیت OR را با شبکه عصبی پرسپترون پیاده سازی کنید. تابع فعال سازی را پله در نظر بگیرید.

$w_1 = 0.1$, $w_2 = 0.2$, $w_b = -0.3$ و نرخ یادگیری را 0.1 در نظر بگیرید.

7- یک شبکه عصبی پیش‌خور (Feedforward Neural Network) با ساختار زیر در نظر بگیرید:

2 نرون ورودی x_1 و x_2 ، 1 لایه مخفی با 2 نرون h_1, h_2 ، 1 نرون خروجی y ، تابع فعال‌سازی تمام نرون‌ها سیگموید است.

وزن‌ها و بایاس ورودی به لایه مخفی $w_{21}=0.4, w_{11}=0.1, w_{12}=-0.2, w_{22}=0.5, b_{h1}=0.1, b_{h2}=-0.1$

وزن‌ها و بایاس لایه مخفی به خروجی $v_1=0.3, v_2=-0.4, b_y=0.2$

نرخ یادگیری 0.1 و مقدار هدف 1 است. مقدار ورودی $x=(x_1=1, x_2=0)$ است.

الف) خروجی هر نرون مخفی h_1 و h_2 و خروجی شبکه را بدست آورید.

ب) مقدار خطای مربعی را محاسبه کرده و مقدار خطای نرون خروجی y را بدست آورید. $E=0.5(t-y)^2$

ج) مقدار خطای نرون مخفی h_1 را محاسبه کرده و وزن v_1 را پس از یک مرحله یادگیری به روزرسانی کنید.

- 8- فرض کنید یک مدل دسته‌بندی دودویی با نظارت برای تشخیص بیماری (Positive) و عدم بیماری (Negative) طراحی شده است. نتایج پیش‌بینی مدل برای 20 نمونه از داده‌های آزمون در جدول زیر آمده است:
- الف) ماتریس سردرگمی را برای آن رسم کنید.
- ب) معیارهای ارزیابی precision، recall، F1 score و Accuracy را بدست آورید.

نمونه	برچسب واقعی	برچسب پیش‌بینی شده
1	Positive	Positive
2	Negative	Positive
3	Positive	Positive
4	Negative	Negative
5	Positive	Negative
6	Positive	Positive
7	Negative	Negative
8	Negative	Positive
9	Positive	Positive
10	Negative	Negative
11	Positive	Negative
12	Positive	Positive
13	Negative	Negative
14	Negative	Negative
15	Positive	Positive
16	Negative	Positive
17	Positive	Positive
18	Negative	Negative
19	Positive	Negative
20	Negative	Negative