

نمونه سوال درس یادگیری ماشین (دکتر ربیعی)

سوالات کلی و کوتاه: تعیین کنید هر کدام عبارات زیر صحیح یا غلط هستند. دلیل خود را با یک جمله شرح دهید

- خطای آموزش الگوریتم K-means با یک خوشه ($K=1$) همیشه برابر صفر است.
- فرض کنید n داده در اختیار است. این مجموعه بطور مساوی به دو دسته آموزش و تست تقسیم شده است. هر چه n افزایش یابد، بازدهی تست و آموزش به یکدیگر نزدیک می‌شوند.
- در یک شبکه عصبی مصنوعی، هر چه تعداد نرون‌های لایه مخفی افزایش یابد، بازدهی تست و آموزش به یکدیگر نزدیک می‌گردد.
- عملیات cross-validation، از overfitting در آموزش جلوگیری میکند.
- عمق یک درخت تصمیم می‌تواند از تعداد داده‌های آموزشی بیشتر باشد.
- عمق یک درخت تصمیم نمی‌تواند از تعداد attribute‌های داده‌ها بیشتر باشد.
- یک درخت تصمیم با عمق ۲ می‌تواند تابع XOR را مدل کند.
- یک شبکه عصبی تک لایه پرسپترون می‌تواند تابع XOR را مدل کند.
- انتخاب الگوریتم یادگیری online یا batch، در overfitting موثر است.
- الگوریتم خوشه بندی k-means، نیاز به یک پارامتر ورودی مربوط به تعداد خوشه‌ها دارد.
- الگوریتم خوشه بندی سلسله مراتبی، می‌تواند دادگان را به k کلاس مختلف قسمت بندی (پارتیشن بندی) کند.
- الگوریتم خوشه بندی k-means، راه حل بهینه مطلق برای تفکیک دادگان به K خوشه را تولید می‌کند.
- هر چه تعداد دادگان آموزشی بیشتر باشد، k-means با دقت بیشتری خوشه بندی را انجام می‌دهد.
- هر چه ابعاد هر نمونه در دادگان آموزشی بیشتر باشد، k-means با دقت بیشتری خوشه بندی را انجام می‌دهد.
- یادگیری تقویتی فقط برای مسائلی با تعداد حالات و اعمال محدود قابل اجراست.
- در شبکه‌های عصبی مصنوعی، هر چه تعداد دادگان آموزشی بیشتر باشد، شانس بیش آموزش (overfitting) افزایش می‌یابد.
- در شبکه‌های عصبی مصنوعی، افزایش تعداد لایه‌ها همیشه منجر به افزایش بازدهی در دادگان تست می‌گردد.
- در شبکه‌های عصبی مصنوعی، افزایش تعداد لایه‌ها همیشه منجر به افزایش بازدهی در دادگان آموزش می‌گردد.
- در مراحل اولیه یادگیری تقویتی، بیشتر exploration انجام می‌شود و هر چه به سمت انتهای فرایند یادگیری پیش می‌رویم، الگوریتم از exploitation استفاده می‌کند.

3.1. Give decision trees to represent the following boolean functions:

- (a) $A \wedge \neg B$
- (b) $A \vee [B \wedge C]$
- (c) $A \text{ XOR } B$
- (d) $[A \wedge B] \vee [C \wedge D]$

3.2. Consider the following set of training examples:

Instance	Classification	a_1	a_2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

- (a) What is the entropy of this collection of training examples with respect to the target function classification?
- (b) What is the information gain of a_2 relative to these training examples?

۳-۳- فرض کنید داده‌های یک مسئله با ۵ ویژگی مشخص می‌شوند. مقادیر قابل قبول برای هر ویژگی، طبق جدول زیر تعیین می‌گردد.

Attr1:	A, B
Attr2:	A, B, C
Attr3:	1, 2, 3
Attr4:	a, b
Attr5:	1, 2, 3, 4

مراحل اجرای الگوریتم ID3 روی داده‌های زیر را نشان دهید.

- + (A,B,2,a,3)
- (B,B,1,a,3)
- + (A,C,2,a,1)
- (B,A,1,a,3)
- + (A,B,2,a,4)
- (A,C,3,b,3)
- (A,C,2,a,2)

۳-۴- در استفاده از یادگیری درخت تصمیم برای دادگان آموزشی که مقادیر ویژگی‌های آن بازه‌ی پیوسته‌ای از مقادیر را در بر می‌گیرد، چه راه حلی را پیشنهاد می‌کنید؟

فصل ۴. شبکه‌های عصبی مصنوعی

- 4.1. What are the values of weights w_0 , w_1 , and w_2 for the perceptron whose decision surface is illustrated in Figure 4.3? Assume the surface crosses the x_1 axis at -1 , and the x_2 axis at 2 .
- 4.2. Design a two-input perceptron that implements the boolean function $A \wedge \neg B$. Design a two-layer network of perceptrons that implements $A XOR B$.
- 4.5. Derive a gradient descent training rule for a single unit with output o , where

$$o = w_0 + w_1x_1 + w_1x_1^2 + \dots + w_nx_n + w_nx_n^2$$

- 4.7. Consider a two-layer feedforward ANN with two inputs a and b , one hidden unit c , and one output unit d . This network has five weights (w_{ca} , w_{cb} , w_{c0} , w_{dc} , w_{d0}), where w_{x0} represents the threshold weight for unit x . Initialize these weights to the values $(.1, .1, .1, .1, .1)$, then give their values after each of the first two training iterations of the BACKPROPAGATION algorithm. Assume learning rate $\eta = .3$, momentum $\alpha = 0.9$, incremental weight updates, and the following training examples:

a	b	d
1	0	1
0	1	0

- 4.8. Revise the BACKPROPAGATION algorithm in Table 4.2 so that it operates on units using the squashing function \tanh in place of the sigmoid function. That is, assume the output of a single unit is $o = \tanh(\vec{w} \cdot \vec{x})$. Give the weight update rule for output layer weights and hidden layer weights. Hint: $\tanh'(x) = 1 - \tanh^2(x)$.
- 4.10. Consider the alternative error function described in Section 4.8.1

$$E(\vec{w}) \equiv \frac{1}{2} \sum_{d \in D} \sum_{k \in \text{outputs}} (t_{kd} - o_{kd})^2 + \gamma \sum_{i,j} w_{ji}^2$$

Derive the gradient descent update rule for this definition of E . Show that it can be implemented by multiplying each weight by some constant before performing the standard gradient descent update given in Table 4.2.

You have a dataset with one real-valued input x and one real-valued output y in which you believe

$$y_k = \exp(wx_k)$$

where (x_k, y_k) is the k th datapoint. This is thus a neural net with just one weight: w .

Give the update equation for a gradient descent approach to finding the value of w that minimizes the mean squared error.

۱۲-۴- نرخ یادگیری و مومنتم هر کدام چه نقش هایی در فرایند یادگیری شبکه های عصبی مصنوعی ایفا می کنند؟

۱۳-۴- کدامیک از توابع زیر را می توان با یک شبکه عصبی با یک لایه مخفی آموزش داد؟ پاسخ را به تفکیک برای شبکه ای که از تابع انتقال خطی (linear) یا حد آستانه (hardlim یا threshold) استفاده می کند شرح دهید.

• تابع چند جمله ای درجه ۱

• تابع چند جمله ای درجه ۲

• تابع جزء صحیح

• تابع $f(x) = \max(1-x, 0)$

۱۴-۴- توضیح دهید هر کدام از موارد زیر با چه درجه ای در بیش آموزش (overfitting) یک شبکه عصبی مصنوعی موثر است.

الف) وزن های اولیه شبکه

ب) نرخ آموزش

ج) تعداد گره های مخفی

د) انتخاب الگوریتم یادگیری دسته ای یا آنلاین

فصل ۹. الگوریتم‌های ژنتیک

۹-۱- شرح دهید چگونه می‌توان مسأله‌ی ۸ وزیر را با استفاده از الگوریتم‌های ژنتیک حل کرد؟ نحوه‌ی نمایش کروموزوم‌ها و تابع ارزشیابی را برای این مسئله چگونه می‌توان تعریف کرد؟

۹-۲-

Assume we have the following function

$$f(x) = x^3 - 60 * x^2 + 900 * x + 100$$

where x is constrained to $0..31$. We wish to maximize $f(x)$ (the optimal is $x=10$)
Using a binary representation we can represent x using five binary digits.

b) Given the following four chromosomes give the values for x and $f(x)$.

Chromosome	Binary String
P_1	11100
P_2	01111
P_3	10111
P_4	00100

c) If P_3 and P_2 are chosen as parents and we apply one point crossover show the resulting children, C_1 and C_2 . Use a crossover point of 1 (where 0 is to the very left of the chromosome)
Do the same using P_4 and P_2 with a crossover point of 2 and create C_3 and C_4

d) Calculate the value of x and $f(x)$ for $C_1..C_4$.

e) Assume the initial population was $x=\{17, 21, 4 \text{ and } 28\}$. Using one-point crossover, what is the probability of finding the optimal solution? Explain your reasons.

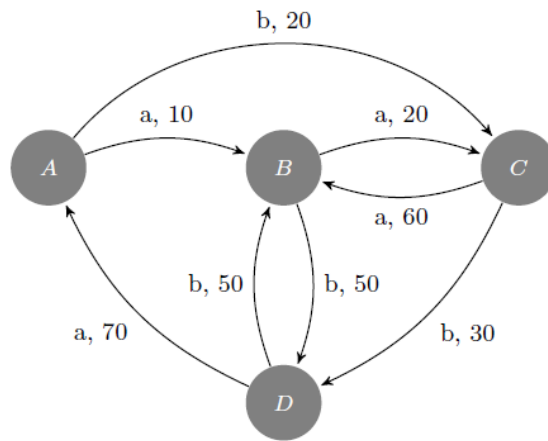
۹-۳- چگونه می‌توان با کمک الگوریتم ژنتیک، برنامه‌امتحانی دانشجویان یک دانشکده را تنظیم کرد؟ نحوه‌ی نمایش کروموزوم‌ها و تابع ارزشیابی را برای این مسئله چگونه می‌توان تعریف کرد؟

۹-۴- چگونه می‌توان با کمک الگوریتم ژنتیک، مسئله‌ی فروشنده دوره گرد را حل کرد؟ نحوه‌ی نمایش کروموزوم‌ها و تابع ارزشیابی را برای این مسئله چگونه می‌توان تعریف کرد؟

فصل ۱۳. یادگیری تقویتی

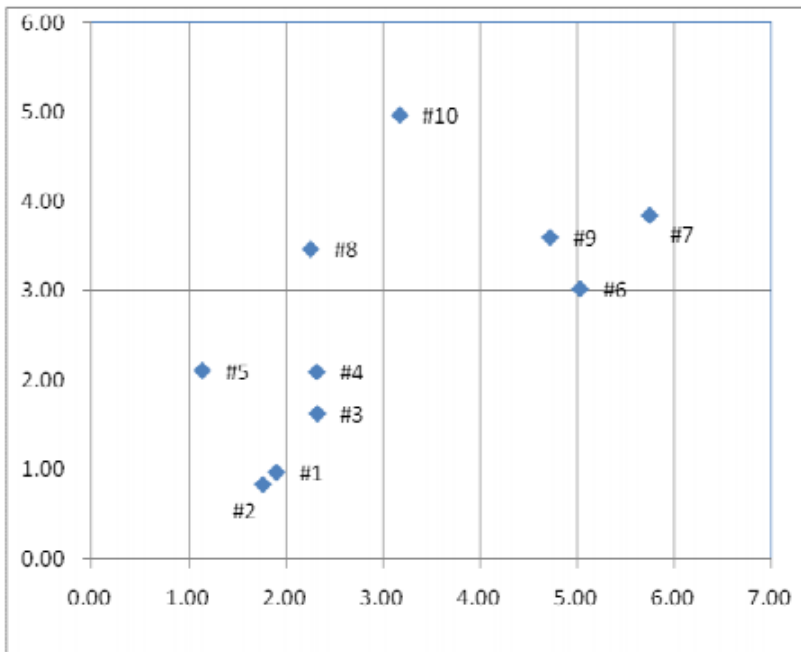
۱۳-۱- تفاوت‌های اصلی یادگیری با ناظر و یادگیری تقویتی چیست؟ (۳ مورد)

۱۳-۲- سیستمی برای کنترل یک فرایند با ۴ حالت $\{A, B, C, D\}$ ، بصورت زیر طراحی شده است. در هر حالت امکان انجام دو عمل a و b وجود دارد. پاداش بلادرنگ (r immediate reward) متناسب با هر فعالیت در شکل نشان داده شده است. با استفاده از الگوریتم یادگیری تقویتی و با فرض $\gamma = 0.9$ و انجام دنباله فعالیت $aaababa$ ، جدول Q چگونه تغییر خواهد کرد؟



خوشه‌بندی

۱- مراحل اجرای الگوریتم K -means با دو خوشه ($K=2$) روی داده‌های زیر را نشان دهید. الگوریتم را با داده‌های شماره ۱ و ۱۰ به عنوان مراکز خوشه‌های تصادفی اولیه شروع کنید.



Data #	x	y
1	1.90	0.97
2	1.76	0.84
3	2.32	1.63
4	2.31	2.09
5	1.14	2.11
6	5.02	3.02
7	5.74	3.84
8	2.25	3.47
9	4.71	3.60
10	3.17	4.96

۲- اگر در مورد تعداد خوشه‌ها در مثال فوق، مطمئن نباشیم، چه راه حلی برای انتخاب بهترین K پیشنهاد می‌کنید.

۳- نقاط زیر را در نظر بگیرید.

	x	y
1	4	0
2	3	2
3	2	2
4	2	4
5	2	1

(الف) ماتریس فاصله مربوط به این نقاط را محاسبه کنید (تکمیل جدول زیر). دقت کنید که ماتریس فاصله متقارن است.

x1	0				
x2	?	0			
x3	?	?	0		
x4	?	?	?	0	
x5	?	?	?	?	0
	x1	x2	x3	x4	x5

(ب) dendogram حاصل از اعمال الگوریتم خوشه‌بندی سلسله مراتبی را روی این ماتریس نشان دهید.

موفق و موید باشید

رپیدی