



دانشگاه صنعتی اصفهان
دانشکده مهندسی برق و کامپیوتر

درس مبانی یادگیری ماشین

تکلیف تئوری اول

مهلت تحویل: ۲۰ آبان ۱۴۰۲

سؤال ۱ (۷ نمره)

صحیح یا غلط بودن گزاره‌های زیر را مشخص کرده و دلیل آن را کامل شرح دهید.

الف) استفاده از روش‌های منتظم سازی به فرار از بهینه‌های محلی در طول آموزش منجر می‌شود. (۱ نمره)

ب) اگر y بردار مقادیر پیش‌بینی شده در یک مدل رگرسیون خطی^۱ و t بردار مقادیر هدف باشند، آنگاه کمینه خطای جمع مربعات، در واقع متناظر با بردار y ای در فضای ویژگی‌ها است که نزدیکترین به t باشد. در این حالت y تصویر t در فضای ویژگی‌ها است. (۱ نمره)

ج) در رگرسیون خطی، یک مدل خیلی ساده معمولاً مشکل زیربرازش^۲ و واریانس کم وجود دارد. (۱ نمره)

د) در تقسیم داده‌ها زیر مجموعه‌های ارزیابی^۳، آموزش و آزمایش داده‌های زیر مجموعه‌های ارزیابی و آزمایش می‌توانند از توزیع‌های مختلفی باشند. (۱ نمره)

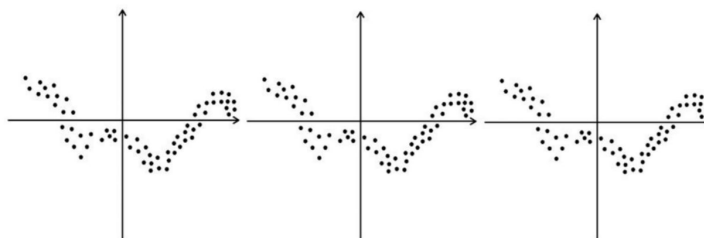
ه) برای تقسیم کردن دیتاست موجود به زیر مجموعه‌های ارزیابی، آموزش و آزمایش نمونه برداری تصادفی از دیتاست به اندازه لازم برای هر زیر مجموعه همواره کفایت می‌کند. (۱ نمره)

و) در یک مسئله رگرسیون خطی برای داده‌هایی که در آنها مقادیر پیش‌بینی شده و واقعی با یکدیگر برابر می‌شوند استفاده از تابع هزینه MAE بهتر از MSE است. (۱ نمره)

ز) در روش اعتبارسنجی متقابل K برابر^۴ پس از مشخص شدن پیچیدگی مدل نتایج مدل با بهترین دقت برای گزارش نهایی استفاده می‌شود. (۱ نمره)

سؤال ۲ (۵/۰ نمره)

در شکل سمت چپ یک نمونه‌ی بیش‌برازش، در شکل وسط یک نمونه‌ی مدل مناسب و در شکل راست یک نمونه‌ی زیربرازش رسم کنید.



¹Linear Regression

²Underfitting

³Validation

⁴K-Fold Cross Validation

سوال ۳ (۱/۵ نمره)

در صورتی که برای مسئله‌ی رگرسیون خطی، رابطه‌ی مدل به صورت

$$y(x, w) = w^T x, \quad w \in \mathbf{R}^{10} \quad (۱)$$

باشد و تابع هزینه برای مسئله‌ی رگرسیون خطی به صورت

$$\tilde{E} = \frac{1}{2} \sum_{n=1}^N \{y(x_n, w) - t_n\}^2 + \frac{\lambda}{2} \|w\|^2 \quad (۲)$$

در نظر گرفته شده باشد. مشخص کنید که مقادیر وزن‌های هرستون متعلق به کدام مقادیر 0 ، -۱۸ و $-\infty$ برای $\ln \lambda$ است. دلیل خود را کامل توضیح دهید.

	a	b	c
w_0^*	0.35	0.35	0.13
w_1^*	232.37	4.74	-0.05
w_2^*	-5321.83	-0.77	-0.06
w_3^*	48568.31	-31.97	-0.05
w_4^*	-231639.30	-3.89	-0.03
w_5^*	640042.26	55.28	-0.02
w_6^*	-1061800.52	41.32	-0.01
w_7^*	1042400.18	-45.95	-0.00
w_8^*	-557682.99	-91.53	0.00
w_9^*	125201.43	72.68	0.01

سؤال ۴ (۱/۵ نمره)

فرض کنید نمونه‌های x_1, x_2, \dots, x_n از یک توزیع یکنواخت با بازه $[0, \theta]$ نمونه‌برداری شده اند و $\theta > 0$ باشد. بیشینه درست‌نمایی پارامتر θ را بیابید.

سؤال ۵ (۲ نمره)

توزیع برنولی یک توزیع گسسته است که احتمال موفقیت آزمایش برابر p و احتمال عدم موفقیت برابر $1 - p$ در زیر نشان داده شده است.

$$p(x) = p^x (1 - p)^{1-x} \quad (۳)$$

با فرض توزیع پیشین به صورت زیر، تخمین p را با MAP به دست آورید.

$$f(p|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1} \quad (۴)$$

$$B(x, y) = \int_0^1 t^{x-1} (1+t)^{-x-y} dt \quad (۵)$$

سؤال ۶ (۲ نمره)

فرض کنید x یک متغیر تصادفی گسسته با احتمال زیر باشد

x	۰	۱	۲	۳
$p(x)$	$\frac{\theta}{2}$	$\frac{\theta}{4}$	$\frac{3(1-\theta)}{4}$	$\frac{(1-\theta)}{2}$

اگر مشاهدات حاصل از نمونه‌برداری از این متغیر به صورت $(0, 3, 1, 2, 1, 0, 2, 1, 3, 2)$ باشد، تخمین حداکثر درست‌نمایی θ را به دست آورید.

سؤال ۷ (۲ نمره)

برای هریک از مجموعه وزن‌های ستون جدول زیر، تعیین کنید کدام ستون متعلق به تابع هزینه $Ridge$ و $Lasso$ می‌باشد. دلیل آن را به همراه رسم نمودار در ۲ بعد شرح دهید. برتری هریک بر دیگر را توضیح دهید.

	Column A	Column B
w_1	0.38	0.50
w_2	0.23	0.20
w_3	-0.02	0.00
w_4	0.15	0.09
w_5	0.21	0.00
w_6	0.03	0.00
w_7	0.04	0.00
w_8	0.12	0.05

سؤال ۸ (۳/۵ نمره)

برای مسئله‌ی رگرسیون خطی، داده‌های زیر موجود هستند. در صورتی که مقادیر وزن‌ها به صورت زیر باشد، به سوال‌های زیر جواب دهید.

$$w_0 + w_1X + w_2Y = Z, \quad w_0 = -59.5, \quad w_1 = -0.15, \quad w_2 = 0.6 \quad (۶)$$

$$LearningRate = 0.1$$

X	Y	Z
41	138	37.99
42	153	47.34
37	151	44.38
46	133	28.17

الف) خطای MSE و خطای MAE را در این مرحله محاسبه کنید. (۰/۵ نمره)

ب) به روش SGD وزن‌ها را با MSE تا ۴ مرحله به‌روزرسانی کنید. (۱/۵ نمره)

ج) وزن‌ها را با MSE به روش GD یک مرحله به‌روزرسانی کنید. (۱ نمره)

د) خطای MSE و وزن‌های حاصل شده برای هریک از روش‌ها مقایسه کنید. (۰/۵ نمره)

نکات تکمیلی

۱. لزومی به تایپ کردن سوالات تئوری نیست؛ ولی در صورتیکه پاسخ آنها به صورت تایپ شده تحویل داده شود، ۵ درصد نمره اضافه به شما تعلق میگیرد. در صورتیکه پاسخهای شما تایپ شده نیست، باید پاسخها خوانا و باکیفیت در قالب فایل pdf ارسال شوند.

۲. فرمت نامگذاری تکلیف ارسالی باید به صورت زیر باشد: HWX_Theory_LastName_StudentID که X شماره تکلیف LastName نام خانوادگی شما و StudentID شماره دانشجویی شما است.

۳. انجام این تکلیف به صورت تک نفره است. در صورت مشاهده تقلب، نمرات هم مبدا کپی و هم مقصد آن صفر لحاظ می شود.

۴. برای تکالیف تئوری امکان ارسال با تاخیر وجود ندارد.

۵. در صورت وجود هر گونه ابهام و یا سوال می توانید سوالات خود را در گروه تلگرام بپرسید. هم چنین می توانید برای رفع ابهامات با دستیاران آموزشی از طریق تلگرام در تماس باشید.
آیدی ها:

@AlirezaT

@Yasinhmv